Gaia
DPAC

Data Processing & Analysis Consortium

Proposal for the Gaia Data Processing

April 2007

GAIA-CD-SP-DPAC-FM-030-2

Gaia Data Processing & Analysis Consortium

Response to ESA's Announcement of Opportunity

Proposal for the Gaia Data Processing

reference	GAIA-CD-SP-DPAC-FM-030-2
date of issue	05 April 2007
status	Submitted to ESA

Compiled for the DPAC by

F. De Angeli, C. Babusiaux, C.A.L. Bailer-Jones, U. Bastian, M. Beck,
M. Biermann, A.G.A. Brown, J. de Bruijne, C. Cacciari, A. Cellino,
J.F. Claeskens, M. Cropper, R. Drimmel, P. Dubath, D.W. Evans, L. Eyer,
C. Fabricius, M. Gai, C. Huc, S. Jordan, C. Jordi, D. Katz, S. Klioner,
M. Kontizas, C. Lagerkvist, U. Lammers, M. Lattanzi, F. van Leeuwen,
T. Levoir, L. Lindegren, X. Luri, E. Masana, F. Mignard, W. O'Mullane,
X. Passot, D. Pourbaix, A. Recio-Blanco, A. Robin, S. Rosen, C. Soubiran,
P. Tanga, F. Thevenin, J. Torra, A. Vallenari, A. Volpicelli, J. Warell

LaTeX environment: U. Lammers
Editors: F. Mignard, R. Drimmel

Cover illustration: Design of K. O'Flaherty, ESA, from an artist's concept of the Milky way based on observations of the NASA Spitzer Space Telescope.

DPAC logo: Design by Tiziana Franconeri and production by Effetti (`www.ineffetti.it`) for INAF-OATo

Contents

Applicable Documents

[Exe06]	Announcement of Opportunity for the Gaia DP
[Tea05]	Gaia Mission Requirement Document
[LJMD$^+$07]	DPAC Product Assurance Plan
[ONHL07]	DPAC Software Configuration Management Plan
[Hoa07]	Gaia MOC SOC Interface Requirements Document

Executive Summary

Following Gaia's launch, expected in 2011–12, data will be transmitted by the spacecraft from its vantage point around L2 at a few Mbps for at least five years, delivering to the ground station an uncompressed raw data set of approximately 100 TB. Scientifically valuable information will be encased in this continuous data stream resulting from the collection of photons in the approximately 100 on-board CCDs in the astrometric, photometric and spectroscopic fields of Gaia. However in its original telemetry format the data is totally unintelligible to scientists, not only because it is squeezed into packets by the numerical coding but, more significantly, because of the way Gaia scans the sky, picking up fragments of the astrometric parameters at each transit of the one billion plus observable sources in the sky. Therefore a large and complex data analysis must follow this raw harvest in order to decipher Gaia's signal and translate it into positional and physical parameters useful to scientists to probe the very nature of astronomical objects. This is precisely the purpose of the data analysis system we have designed and which is presented in the different sections of this document. As the reader will discover himself the whole system is extremely complex, not so much in view of the amount of data to be processed (although this should not be underestimated) but rather because of the intricate relationships between the different pieces of information gathered throughout the mission by the different instruments.

To cope with this challenge, the scientific community, together with ESA, has started a joint effort to set up a data processing ground segment comprising a single processing pipeline which will deliver the intermediate and final mission science products. Since mission selection, the underlying principles of the data processing have been developed by the Gaia scientific community and individual pieces were successfully tested on small or intermediate size simulations. During this phase we have attempted to identify the critical elements of this processing (size, iterative procedures, instrument calibration, data exchange, human and financial resources, computing power) and assess the risks inherent to an endeavour of that size, unprecedented in astronomy. Based on these preparatory activities we think that we are in position to propose a complete data processing system capable of handling the full size and complexity of the real data within the tight schedule of the mission. The details of this system, its expected performances, funding, organisation and management are described in this document. Formally, this is the proposal of the Gaia Data Processing and Analysis Consortium (DPAC) in response to the AO for the ground-segment data processing of the Gaia data.

The DPAC is a European collaboration including the ESA Gaia Science Operations Centre (SOC) and a broad, international science community of over 320 individuals, distributed on more than 15 countries, and including six large Data Processing Centres (DPCs). The Consortium has carefully estimated the effort required and has united in a single organisation the material, financial and human resources, plus appropriate expertise, needed to conduct this processing to its completion in around 2020.

The Consortium is structured around a set of eight Coordination Units (CUs) each in charge of a specific aspect of the data processing. The CUs are the building blocks of the Gaia Data Processing and Analysis Consortium. They are reasonably small in number, with clearly-defined responsibilities and interfaces, and their boundaries match naturally with the main relationships between tasks and the associated data flow. Responsibilities of the coordination units include: (a) defining data processing tasks and assigning responsibilities; (b) establishing development priorities; (c) optimizing, testing and implementing algorithms; (d) verifying the quality of the science products. Each coordination unit is headed by a scientific manager assisted by one or two deputies and, where appropriate, a technical manager. The management team of each CU is responsible for acquiring and managing the resources needed for their activities. While the CUs are primarily structured for software development, all of them are closely associated with at least one DPC where the algorithms will be implemented for the data processing in the operational phases.

The Consortium is coordinated by the 'Data Processing and Analysis Consortium Executive' (DPACE) committee. This top-level management structure deals with matters that are not specific to the internal management of a CU, defining standards and policies needed to ensure an efficient interaction between all the CUs. Consistent with the Science Management Plan, the DPACE and its chair will serve as the interface between the DPAC and the Project Scientist and the Gaia Science Team. They are ultimately responsible for the data processing carried out by the DPAC.

The Gaia data processing we propose is based on an iterative self-calibrating procedure where the astrometric core solution is the single largest treatment and the cornerstone of the whole processing. The overall processing starts with an initial data treatment aiming to obtain a first estimate of the star positions and magnitudes together with an initial attitude, primarily based on the on-board crude determination improved on the ground with more accurate (and more computer greedy) algorithms. The next step is the iterative astrometric core solution which provides calibration data and attitude solution needed for all the other treatments, in addition to the astrometric solution of about 100 million primary sources and the rigid reference system. Once these two steps are performed and their products stored in the central data base, the more specialised tasks are launched with the photometric processing and variability detection, the global analysis of the spectroscopic data and the data treatment of all the difficult sources, like planets and multiple stars, that do not fit in the general astrometric or photometric solution of single stars. Two more processing close the chain, dealing with the analysis of all the types of variable stars and the retrieval of the stellar astrophysical parameters like their luminosity, temperature or chemical composition. Each step in the processing using earlier data has it own logic regarding the products it will deliver, but must also be seen as part of the validation of the whole process.

The data flow between the different DPCs in the DPAC takes place via a Main Database (MDB) hosted by the SOC. Intermediate processed data from the DPCs flows into the MDB; the DPCs can then extract the data from this for further processing. Thus the Main Database acts as the hub of DPAC's data processing system and must have an efficient and reliable connection with all the DPCs. During operations, the Consortium plan is to version this database at regular intervals, typically every six months, corresponding roughly to the period for the satellite to scan the whole celestial sphere. Each version of the MDB is derived from the data in the previous version, supplemented with the processing of the newest observations.

This document is internally organised to address the key elements of the data processing and to explain how the DPAC will face this challenge. Its goal is threefold:

- to respond to the AO requirements, describing the principles and methods to be employed in the processing and explain how we will develop, test and implement the processing;
- to describe the Consortium organizational and management structure, demonstrating that the means to carry out the task have been correctly estimated and how they will be secured over the Consortium lifetime;
- to serve as a reference document for the DPAC itself, especially to new members who will join the DPAC in the future.

Part I discusses the context and boundary conditions of the Gaia data processing, recalling the main objectives of the mission, the observing strategy of the satellite, and the kind of raw data that will be produced by the on-board instruments and sent to ground. The second part describes the Gaia data processing itself, from the initial data treatment to the final determination of the astrophysical parameters for each source, including the end-to-end simulations needed for development, testing and verification planned by the Consortium. The last chapter of this part outlines the overall system architecture for the implementation of the data processing and the organisation of the data exchange. The third part describes the DPAC itself, presenting each of the CUs, DPCs and the management structure, as well as summarizing the resources to be employed. The top level work packages are detailed in a series of appendices. Other appendices provide the compliance matrix of this document with the AO requirements, the information (deliverables) that the DPAC requires from ESA, and the qualifications of the members of the DPAC management team.

This proposal is driven by the ambition to provide the best one can draw from the Gaia's raw measurements, the ambition to deliver before 2020 more than an observation catalogue but an organised data base containing all the Gaia products and finally the ambition to renew and amplify the remarkable achievement of Hipparcos still unchallenged ten years after its completion. We are confident that the DPAC has the means to achieve these goals and trust that the elements elaborated in this document will convince ESA that our proposal is realistic, efficient and robust.

Part I

Presentation and objectives

1 Overview of the Gaia Science Case

1.1 Space Astrometry and Hipparcos legacy

A clear night sky, free from city smog and light pollution, provides an awesome spectacle which has inspired the human race through the ages. The way it changes regularly over the year was used for tracking the seasons in early agriculture; the fixed relative positions of the stars were used for navigating across the oceans and for orientation on land. Those apparently fixed positions also provided the reference for observing the wanderers of the sky, the planets, Moon and Sun. The extensive study of these objects led to the formulation of basic laws describing gravitationally bound orbits by Kepler, which in turn led to the laws of classical mechanics by Newton. Improving the description of the planetary orbits, so that their positions could be predicted more accurately, required ever increasing measurement accuracies. These requirements stimulated the development of high performance specialized instrumentation which required precision engineering, from the initial large quadrants to the recent Hipparcos and Gaia satellites. Requirements on lenses and mirrors for these instruments were, as demonstrated by the work of Fraunhofer, a major stimulant in studies of optics.

By the second half of the 18th century those in the forefront of celestial research were considering methods for measuring the distances to the stars. The most direct way of measuring the distance of an object that can't be reached is through triangulation: using a known-length baseline and measuring the difference in direction when the object is observed at either end of this baseline. This difference is called the parallax of the object; the smaller the parallax, the greater the distance. The largest baseline astronomers have readily available is the diameter of the orbit of the Earth around the Sun.

Measurement accuracies at the end of the 18th century were around several arcseconds with which it was not possible to detect stellar parallaxes. However, this did allow the discovery detection of aberration and nutation, and the confirmation of the finite speed of light. What had also been detected were the so-called proper motions of stars: although the positions of the stars on the sky appear fixed over an individual lifetime, they were actually observed to be moving, though very slowly.

It was one of the stars which showed a relatively large proper motion, 61 Cyg, which led to the detection of the first parallax in the middle of the 19th century by Bessel. Bessel had correctly argued that stars showing large proper motions were more likely to be among the closest to the Sun. Bessel used a special telescope: a Heliometer built by Fraunhofer, an instrument designed for very accurate position measurements. In some ways this instrument prefigures the technique used 140 years later on Hipparcos: cut the objective or main mirror in two, and allow each half to project a different part of the sky on the same detector. By measuring differential rather than absolute positions, very precise relative measurements could be obtained.

In the Heliometer used by Bessel the relative positions of two stars were measured through shifting one half of the objective with respect to the other half such that images of the two stars would coincide. The shift applied to the lens was a measure of the distance on the sky between the two stars. Assuming one star to be relatively nearby, and the other relatively far away, this difference will primarily show the displacements of the nearby star. Those displacements are the proper motion – linear as a function of time – and a motion induced by the Earth's motion around the Sun, which describes a small ellipse over a year. Bessel concluded that the parallax of 61 Cyg A was 314 ±20 mas (compared to the current best estimate of 287.18 ±1.51 mas).

By the end of the 19th century it had become clear how distant stars are, with the largest parallax measured at around 700 mas. A parallax of 1000 mas is caused by a star at a distance from the Sun 206 264 times further than the average distance of the Earth from the Sun. This distance is referred to as one parsec. It takes light more than three years to travel this distance.

An important development that led to significant improvements of measurement precision was the introduction of photographic plates and subsequently of measuring machines to extract the positional information from those plates. Following an initiative by the Observatoire de Paris, the first all-sky photographic catalogue was created, followed by an all-sky reference map, the Carte du Ciel. The photographic-plate field nicely covered the Pleiades cluster, which was therefore often used as an object to test these Astrograph telescopes. Plates of the Pleiades cluster taken towards the end of the 19th century, combined with measurements by Bessel obtained around 1840, showed for the first time the existence of a group of stars sharing the same proper motion. Around 1900 it became in addition possible to take photographic plates at different wavelengths. This allowed for the measurements of stellar colours. The relation between brightness and colour for those stars that appeared to share the same proper motion in the Pleiades field led Hertzsprung, and independently Russell, to the discovery of the relation between temperature and brightness for stars, a relation now known as the HR-Diagram.

Photographic techniques, used on specially built long-focus telescopes, dominated the parallax determinations for the first 70 years of the 20th century. The main problem with those measurements was their differential nature: on the small field of a photographic plate, usually less than one degree diameter, parallaxes can only be determined relative to selected reference stars. However, these reference stars themselves also have parallax displacements, which had to be estimated (in a statistical sense) from the stellar colours and magnitudes. Hence the *precision* of these measurements was always much higher than their *accuracy*, and data obtained by different observatories more often than not disagreed well beyond the claimed accuracies. By the 1970s this method of ground-based parallax determinations appeared to have reached its limit.

Two main driving forces led to the subsequent development of space astrometry: the need to be able to observe the entire sky with one instrument, and the need to measure absolute rather than differential parallaxes. Both could only be achieved by means of a dedicated satellite, and the first ideas of such a mission were developed by Lacroute in the 1960s. The original design was for a relatively small mission, covering a few thousand stars only, and was put forward to CNES in August 1967. Lacroute introduced the idea of observing through two sufficiently well-separated apertures onto the same focal plane. As parallax displacement directions at a particular epoch are a function of position on the sky, well-separated fields of view will be subject to different displacement directions, which ultimately, by combining data collected over the entire sky and over a period of at least 18 months, allows for the full determination of the parallaxes of the observed stars.

CNES did not take on the challenge of building the first astrometric satellite, but around 1975 ESA became interested. With the introduction of novel observing techniques, which had been pioneered on transit instruments by Erik Høg in the 1960s, a mission emerged that would be capable of measuring around 100 000 stars over a 30 months period: Hipparcos was born. Hipparcos was designed as a survey satellite, scanning each part of the sky in at least two directions every six months. The scan was made in two fields of view, separated by 58 degrees as projected on the sky. The two fields of view were combined into one optical system through the beam combiner, a mirror first polished, then cut in two, and glued together at an angle of 151 degrees, thus providing a basic angle of 58 degrees between the fields of view. In this way both fields of view were projected on the same focal plane. (More on the early design of Hipparcos can be found in [vL97, Kov98]).

By describing a great circle on the sky, the differential data in the focal plane could be related to differential data on a great circle. The accuracy of this depended on two assumptions: the stability of the basic angle (both random and systematic errors had to be well below 1 mas), and the ability to reconstruct the rotation phase of the satellite with sufficient accuracy. The main problem was that the astrometric data which the mission was trying to retrieve was also the data used to reconstruct the along-scan rotation phase of the satellite. This mutual dependence between catalogue construction and attitude determination has led to various iterative reduction schemes for the Hipparcos data. This includes the 3-step great-circle based reduction as used in the published data [vL97] as well as the recent global iterative solution as applied in the re-reduction by van Leeuwen [vLF05, vL05].

The attitude reconstruction plays a very important and critical role in the reduction of the Hipparcos as well as the Gaia data. It is through the attitude reconstruction that the observations from the two fields of view become properly linked. It is therefore not helpful if one field of view contains many more stars than the other. In order to become properly linked, both fields of view will need to contribute more or less equally to the attitude modelling. To accommodate this for the Hipparcos mission, an Input Catalogue was created such that the stars were more or less evenly distributed over the sky. The preparation of the Hipparcos Input Catalogue, in which the errors on the predicted star positions needed to be less than 300 mas, was led by Catherine Turon at Meudon Observatory. Starting from more than 200 proposed observing programmes, the Input Catalogue contained 118 218 stars brighter than magnitude 12.5, and was nearly complete for stars brighter than magnitude 8.

Hipparcos was launched on 8 August 1989 from Kourou. Considering the limitations of the computer hardware in the 1980s, the processing of the Hipparcos data was not going to be simple. A 3-step reduction scheme designed by Lindegren enabled the reduction of the data with sufficient accuracy to reach the original goal of the mission, namely to measure parallaxes with accuracies of up to 2 mas. This accuracy was set by the expected noise level on the attitude reconstruction and instrument parameter calibration, and was at the level of the photon noise for stars of 11 magnitude only. Despite an orbit anomaly and the resulting problems for mission operations and data processing, the mission results showed accuracies as good as 0.5 mas (photon-noise limited down to magnitude 9 to 9.5). These results were obtained by two parallel data processing consortia, FAST (with institutes from France, Germany, Italy, The Netherlands and USA, led by Jean Kovalevsky) and NDAC (Sweden, Denmark and the United Kingdom, led by Lennart Lindegren), and the final published result is a weighted mean of the two consortia products. The publication of this first ever space astrometry mission aroused international interest and was widely acclaimed beyond the inner circle of specialists as a landmark in astronomical science. The published catalogue shown in Fig. 1 and its associated database has fostered nearly three thousands refereed and proceeding papers since 1997 in many fields, including stellar physics, galactic kinematics, reference frames and double stars.

Since the publication, a critical examination [vL05] of the data processing has shown that the results could have been even better and this brought to light some imperfections in the data analysis model. A new reduction – using a global iterative procedure not dependent on great circles – has been completed which will lead to a new catalogue in which the accuracies are photon-noise limited down to magnitude 3 to 4, with accuracies as high as 0.15 to 0.2 mas. These accuracies reflect the much improved understanding and modelling of the satellites attitude, in which the noise was reduced from 2 mas in the published data to around 0.5 mas in the new reduction. The noise in the published data was dominated by modelling inaccuracies, causing significant error correlations in the astrometric data. In the new reduction these correlations have been reduced by a factor 40 to a completely insignificant level.

Apart from astrometric data, Hipparcos also provided a photometric catalogue, with epoch photometry in a broad passband plus B and V photometry for objects brighter than magnitude 10.5. The latter was obtained from the star mapper slits in the focal plane. The astrometry from transits through those slits was used for the on-board attitude control and on-ground attitude reconstruction (in particular for the cross-scan

Figure 1: The 17 volumes of the printed version of the Hipparcos Catalogue.

positions and velocities). It was in addition used to construct an all-sky astrometric catalogue down to magnitude 11.0 to 11.5, the so-called Tycho catalogue, prepared by the Tycho consortium led by Erik Høg.

1.2 Objectives of Gaia

1.2.1 Beyond Hipparcos

The Hipparcos Catalogue was not yet published when the scientists involved in finishing the data processing were already thinking of a new, much more ambitious space mission combining a high-accuracy astrometric and photometric survey of faint stars. Without invoking too much change from the Hipparcos concept it was clear that CCD detectors had reached a level of maturity to allow a change from the old technology of photomultipliers used in Hipparcos into the more efficient and space qualified electronic detectors available in 1995. This led to ROEMER as the first proposed successor of Hipparcos submitted in 1993 for the Third Medium Size ESA mission [BGH^{+}93], [Høg94]. ROEMER was essentially a Hipparcos replica operated with CCDs and with the capability of observing thousands stars simultaneously in the sub-milliarcsec accuracy. This translated into a system with errors 20 times smaller than Hipparcos and a five magnitude fainter limit. Already in the scientific proposal were all the ingredients that make the Gaia mission so valuable for stellar physics, galactic kinematics, the distance scale and the reference frame. The proposal was rated very highly by the ESA advisory committees and scientifically was ranked the highest of all medium-sized missions by the Astronomy Working Group. However, other considerations led it to being considered as only a possible post Horizon 2000 Cornerstone Mission.

At the same time a group led by L. Lindegren and M. Perryman were considering a more appealing mission concept for the Horizon 2000 Follow Up, proposed to reach the 10–20 microarcsec accuracy for the 50 million stars brighter than $V = 15$ [LPB$^+$93], [LP94]. The concept was submitted to the ESA Horizon 2000+ Survey Committee in 1994 and consisted of a scanning Fizeau interferometer (in fact three identical interferometers stacked on top of each other) able to do global astrometry from space in a similar way as Hipparcos, but two orders of magnitude better in accuracy. While the project evolved significantly during the study phase (to the point that the pairs of entrance pupils of the interferometers merged) the overall scientific objectives remained unchanged and have been the key elements in the success of the Gaia proposal. Gaia was approved by ESA's Science Programme Committee in 2000, and redesigned in 2002 as a consequence of a cost-reduction exercise.

The Gaia payload consists of three distinct instruments mounted on a single optical bench, the design of which is detailed elsewhere in this volume. Unlike HST and SIM, which are pointing instruments observing a preselected list of objects, Gaia is a scanning satellite that will repeatedly survey in a systematic way the whole sky, tying together without regional errors widely separated sources. The main performances of Gaia expressed with just a few numbers are staggering and account for the vast scientific harvest awaited from the mission: a survey to $V = 20$ of all point sources (more than one billion objects), with an astrometric accuracy of 25 μas at $V = 15$ and 7 μas for the few million stars brighter than $V = 13$; radial velocities down to $V = 17$, with an accuracy ranging from 2 to 10 $\mathrm{km\,s^{-1}}$; multi-epoch spectrophotometry sampling from the visible to the near IR. The expected performances are listed in Tab. 1 and are more notable when compared to Hipparcos, given the lasting impact of this latter mission has had on astrophysics.

Beyond sheer measurement accuracy, a major strength of Gaia follows from (i) its capability to perform an all-sky and sensitivity limited absolute astrometric survey, (ii) the unique combination into a single spacecraft of the three major electronic detectors carrying out nearly contemporaneous observations, (iii) the huge number of objects and observations which allow the accuracy on single objects to be achieved on very large samples, thus yield statistical significance. This last feature is immensely valuable for astrophysics and not shared by a mission like SIM, a pointed mission able to concentrate only on relatively few pre-selected sources with high astrophysical potential.

Table 1: Astrometric performances of Gaia compared to Hipparcos

	HIPPARCOS	Gaia	
Magnitude limit	12	20 - 21	
Completeness limit	7.3 - 9	20	
Number of objects	120 000	35×10^6	$V < 15$
		350×10^6	$V < 18$
		1.3×10^9	$V < 20$
Astrometric accuracy	1 mas ($V < 9$)	7 μas	$V < 12$
	1-3 mas($V > 9$)	25 μas	$V = 15$
		300 μas	$V = 20$
$\sigma_\pi/\pi < 1\%$	150 stars	11×10^6 stars	
$\sigma_\pi/\pi < 5\%$	6,200 stars	77×10^6 stars	
$\sigma_\pi/\pi < 10\%$	21,000 stars	150×10^6 stars	
Radial velocity	–	$2 - 10$ km s^{-1}	$V < 17$
Spectro-photometry	–	$\simeq$ 25-colour	$V < 20$
Low resolution spectroscopy	–	R = 11,500	$V < 16 - 17$

1.2.2 Gaia: Science

The immediate objectives of space astrometry are, in principle, the same as those of ground based astrometry: to determine the apparent positions of celestial bodies over time and derive from them astrophysically important parameters such as distances, proper motions and motions within double and multiple star systems. However, the current generation of CCD-based astrometric instruments can also be used as stable photometers to acquire photometric measurements. Such an approach has appeared in every proposal of a scanning mission having astrometry as a first goal, such as the German DIVA mission or the different versions of the US led project FAME . In addition, Gaia carries a dedicated spectrometre for determining radial velocities for at least 100 million stars.

Whereas the direct product of the Gaia mission will be a highly accurate astrometric and photometric survey to V = 20 mag, the science goals are much broader and account for the support of a large scientific community. A top-level summary is as follows:

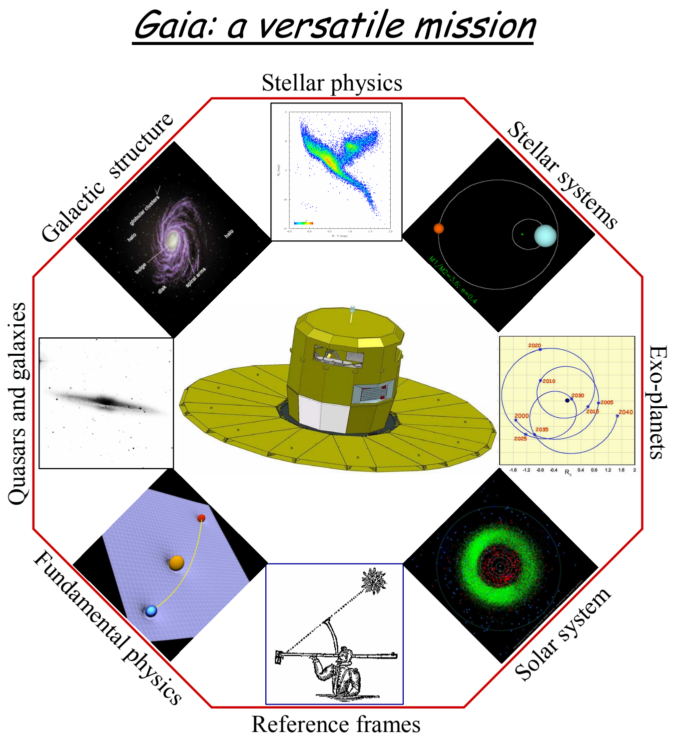


Figure 2: Sketch of the Gaia science capabilities showing the various areas where Gaia will make significant contributions.

- Mapping the Milky Way in three dimensions (parallaxes, positions, extinction)
- Galactic kinematics (proper motions and radial velocities)
- Formation and evolution of the Milky Way
- Stellar physics (classification, M, L, $\log g$, $T_{\rm eff}$, [Fe/H])
- Distance scale (geometric distances to Cepheids and RR Lyrae stars, HR diagram)
- Age of the Universe (cluster diagrams, distances, luminosities)
- Dark matter (stars as tracers of gravitational potential)
- Reference frame (quasars, absolute astrometry)
- Extrasolar planet detection ($\sim M_J$, astrometry and photometric transits)
- Fundamental physics (relativistic parameters $\gamma \sim\ 5\times10^{-7}$, $\beta\sim5\times10^{-4}$)
- Solar system science (taxonomy, masses, orbits for 5×10^5 bodies)

Gaia will measure the positions, distances, space motions, and many physical characteristics of about one billion stars in our Galaxy and beyond. Gaia will provide the detailed 3-D distributions and space motions of all these stars, complete to 20th magnitude. The measurement precision, reaching a few millionths of a second of arc, will be unprecedented. This will allow our Galaxy to be mapped, for the first time, in three dimensions. Ten million stars will be measured with a distance accuracy of better than 1 percent and a 150 million to better than 10 percent. Compared to Hipparcos, Gaia will improve parallax and proper-motion accuracy by almost a 100 times and the number of stars observed 10 000 times. In addition it measures radial velocities and spectrophotometry for all the sources. Gaia will survey a vast population of solar system bodies (major planets, natural satellites, comets, and asteroids, including several thousand near-Earth objects) and extragalactic objects (half a million quasars and thousands of supernovas). In addressing all these fields, Gaia covers a significant part of modern astrophysics [PdBG^{+}01],[Mig05b].

1.2.2.1 Stellar physics

The study of stellar structure and evolution provides fundamental information on the properties of matter under extreme physical conditions as well as on the evolution of galaxies and on cosmology. Gaia will provide accurate distances to massive numbers of stars, which combined with photometric and other data give luminosities, surface temperatures, abundances, masses and ages for very many stars. Effects that will be probed include the size and properties of convective stellar interiors (in combination with astroseismic data) and the diffusion of chemical elements in radiative zones. The availability of high-precision fundamental data for large stellar samples, including rare objects (such as extremely metal-poor stars and rapid evolutionary phases), will greatly advance the theoretical modelling of stellar interiors. Multi-epoch, multi-colour photometry of all stars brighter than 20 mag plus multi-epoch spectral information for the 20 million brightest stars, will provide a description of stellar stability and variability across the HR diagram. Multi-epoch radial velocities will also be very useful for monitoring the radial pulsations of variable stars, including the classical Cepheids, which will have precisions per epoch of ~ 5 km s^{-1} out to several tens of kiloparsecs. Gaia will thus have a major impact on our knowledge of the distance scale in the Universe by providing accurate distances and physical parameters for primary distance indicators in the Milky Way and nearby Local Group galaxies.

The Hipparcos mission has shown the capabilities of space astrometry to detect and resolve binary stars with a separation smaller than the size of the diffraction pattern. Gaia will have similar detection capabilities and will be able to resolve binaries with an apparent separation of less than 20 mas. The Radial Velocity Spectrometre (RVS) will observe about 10^6 spectroscopic binaries and about 10^5 eclipsing binaries. Of these 10^5 eclipsing binaries, 25% will be double-lined spectroscopic binaries (SB2) for which it will be possible to derive the masses of the two components [SÖ5]. Combined with the distance and luminosity calibration this will put the investigation of the M-L relation on much safer grounds than is the case with today's mass estimates. Gaia will detect a majority of the $\sim$ 10 million binaries closer to 250 pc from the Sun, and further away an estimated 60 million will be identified.

1.2.2.2 Galactic structure and evolution

While data from Gaia will produce an unprecedented astrometric catalogue of 10^9 stars, in many respects the Galaxy could be considered as *the* primary science target of the Gaia mission. Stellar distributions in position and velocity are constrained by the Galactic gravitational potential, with the distribution of young stars also depending on the time- and position-dependent star formation rate. Measures of interstellar extinction will trace the ISM. For the brightest stars our kinematic knowledge will be complete, and these include important tracers: the radial velocities of K giants as far as $\sim$ 20 kpc will be measured with a precision better than $\sim$10 km s^{-1}; the kinematics of stars at the tip of the red giant branch – together with asymptotic giant branch stars (AGB) and CH-type carbon stars – will be observable to $\sim$50–75 kpc in the outer halo; the brightest and youngest stars in the spiral arms will have radial velocity precisions better than $\sim$5 km s^{-1} up to $\sim$2.5 kpc (for a B5V star). The kinematic mapping of our Galaxy that will be accomplished with Gaia data will allow us to reach a thorough understanding of the large-scale dynamical processes shaping our Galaxy, including instabilities that transport angular momentum (e.g., bars, warps and spiral arms), as can be done with no other galaxy. The large, complete and unbiased sample of remote tracers in the halo permit a more accurate estimate of the mass of the Milky Way halo [WE99, WVT$^+$05]. Together with kinematics, the measurement of stellar metallicities and ages will allow an identification of the stellar populations that are the building blocks of the Galaxy. Furthermore this will allow us to trace past mergers with other galaxies and thus allow a historical reconstruction of the formation and evolution of the Milky Way. At present, such detailed data are only available for the immediate vicinity of the Sun. Understanding the Galaxy as a whole requires the observation of large, unbiased stellar samples over a substantial fraction of its volume. Gaia's comprehensive astrometric, photometric and spectroscopic survey will achieve just this, and thereby revolutionize Galactic studies.

1.2.2.3 Extrasolar planets

Currently more than 200 planets have been detected around nearby stars, almost all of them from the radial-velocity variations induced in their parent stars. The spectroscopic method is, however, limited by the number and types of stars that can be surveyed and its bias to short orbital period companions. In contrast, Gaia will survey hundreds of thousands of nearby stars for planetary companions using the astrometric variation of stellar position ([LSSC00]). The astrometric method is applicable to all spectral types, allows unambiguous determination of orbital inclinations and planetary masses, and is more sensitive to longer orbital periods (up to 5–10 years).

Gaia's singular contribution to extrasolar planet science is its potential to find thousands of new planets, increasing our statistical knowledge in a way that no other space mission or ground based programme can achieve, and thereby allowing proposed models of planet formation and evolution to be tested ([Soz05]). It will also provide invaluable inputs for future missions, such as Darwin/TPF, that will search for terrestrial planets in the habitable zones of their host stars.

1.2.2.4 General Relativity

Gaia will follow the bending of star light by the Sun and major planets over the entire celestial sphere, and therefore directly observe the structure of space-time. At an accuracy of a few microarcsec, relativistic effects can no longer be considered as small corrections to a Newtonian model and General Relativity becomes a fundamental part of the mathematical framework for the data processing ([Kli03b]). It is one of the exciting goals of the mission to be able to make the first global test of General Relativity. Gaia appears capable of measuring departures from The General Relativity value ($\gamma = 1$) of the PPN parameter γ of order of a few parts in 10^{-7}([VLB+03]). These accuracies compare quite favorably with recent findings of scalar-tensor cosmological models, which predict for γ a present-time deviation from the General Relativity value between 10^{-5} and 10^{-7}. Also, the magnitude of the deflection when observing at right angles to the Sun is about 4000 microarcsec, while for a ray grazing Jupiter it is 16000 microarcsec. Combining millions of measurements will map these effects to unprecedented accuracy. For the first time it will be possible to measure more subtle effects such as the quadrupole moments of Jupiter, which according to General Relativity give an amplitude of 240 microarcsec ([CM06]).

1.2.2.5 Solar System

Solar System objects appear in the Gaia proposal at a high level in the science program achievable within the current design, with a relevance to virtually all categories of minor planets. Solar System studies will receive a massive impetus through the repeated observations of several hundred thousands of minor planets of every type (main belt, Trojans, near-Earth and trans-Neptunian objects). However, the astrometric observation of rapidly moving objects is very challenging for the data processing so dedicated recognition algorithms will be implemented. This systematic, sensitivity-limited survey covering the whole celestial sphere will be the most direct impact of Gaia on Solar System science. But more importantly, Gaia will observe a given asteroid about 15 times per year with an accuracy better than 1 mas (per observation), 500 times better than that of most present large surveys. In addition, Gaia will simultaneously obtain spectrophotometric data at every epoch. The following and impressive list gives the main areas in which the combination of the Gaia measurements will most benefit to Solar System science:

- Orbits: virtually all objects observed $\times 30$ better than now $\Rightarrow$ proper elements, dynamical families

- Masses from close encounters (some ∼ 100 masses expected)
- Diameters for over 1000 asteroids ⇒ shape, density
- Photometric data in several bands ⇒ albedo, taxonomic classification
- Light curves over five years ⇒ rotation, pole, shape
- Space distribution vs. physical properties
- Perihelion precession for 300 planets ⇒ GR testing, solar J2

2 Overview of the Gaia Data Processing

This section describes the global nature of the data processing, its complexity for Gaia and its necessary stepwise structure.

2.1 Objectives of the data processing

2.1.1 Overview

Gaia will produce an impressive (by current standards) volume of raw data with about 50 GB of uncompressed science data per day yielding at mission completion a telemetry volume of roughly 100 TB. The greater part of this data volume represents photon counts registered by the video chains of the CCDs in the astrometric, photometric, and RVS fields of Gaia (together some 10^9 pixels). Although all the science content is there, it is deeply hidden in the data coding and measurement process. Transforming the raw data into scientifically meaningful quantities such as parallaxes, magnitudes and radial velocities is what the activities of DPAC are all about. It is a huge and challenging task in many respects: mathematics, physical modelling, computer programming, data management, storage and access, international funding and a bit of sociology.

On a global scale the Gaia data processing problem can be viewed as an iterative adjustment of a very large number of instrumental, astrometric and photometric parameters to the entire mission data set. A subset of the adjusted parameters will form the final science data from the mission.

2.1.2 Gaia raw data

Gaia's two telescopes scan the sky at a constant rate of 60 arcsec/s. As the satellite rotates the image of an object will move across the Focal Plane Assembly (FPA), crossing first one of the Sky Mappers (SM), then the Astro Field (AF), thereafter the Blue and Red Photometers (BP and RP), and finally the Radial-Velocity Spectrometre (RVS) field (see Fig. 6). The science data consists of FPA-generated video samples combined in data packets for each object before being compressed within the Video Processing Units (VPU). These star packets contain, in a predefined arrangement, the counts recorded in one of the two SMs around every detected source plus the content of the windows for the subsequent 9 Astro CCDs and the 2 photometers. A typical star packet will comprise a header followed by the observation values. The header describes the object and some general data associated with the observation, such as the detection time, the across-scan (AC) position, the estimated magnitude and the scan rate. The content of the *value component* depends on the CCD and usually also on the object magnitude, as specified by the sampling strategy. For the AF CCDs this would be counts binned in the across-scan direction to reduce both the noise and the telemetry volume. For the photometers the read windows are transmitted with numerical binning in the AC direction for faint stars. The RVS spectra are transmitted in dedicated star packets as they are read out, without binning.

The star packets are recorded in the Payload Data Handling Unit (PDHU) before being downlinked to the ground when ground contact is established. The telemetry flow itself is divided into packet streams that are transmitted to the ground antenna, nominally Cebreros in Spain.

In addition to the science data, a much smaller volume of ancillary data is also transmitted to the ground. This include satellite housekeeping data, the results of on-board processing and the measurements made by the Basic Angle Monitoring (BAM) and Wavefront Sensor (WFS) devices.

The telemetry data will be initially processed by ESOC before being sent on to ESAC which is hosting the Mission Database. This initial ground processing provides diagnostic information on the state of the satellite and its instruments, but does not change the science content of the telemetry described above, which represents the raw data from the Gaia instrument.

2.1.3 Final Science Data

The goal of the data processing is to transform the raw data into the Final Science Data. In simplistic terms, the Final Science Data will consist of an astrometric-spectrophotometric catalogue of approximately 10^9 sources based on the ~ 70 observations made of each source. (The term ‘source’ designates any object detected and subsequently observed by Gaia, whether it is a single star, a binary or multiple star, an extragalactic object, or a solar-system body.) The following describes the minimum targeted content of the Final Science Data, which we can call the **Primary Science Data**, and is defined in broad terms in the SMP and AO:

Identifier: Each source will have a unique alphanumeric identifier.

Astrometry: For each stellar and extragalactic object the astrometric data will nominally consist of five astrometric parameters (position at epoch, proper motion components and parallax). Additional parameters will be needed for the (numerous) cases where the source has a more complex nature than that of a single isolated point source. For example, in the case of a close binary system, the orbital elements of the system will be derived. These astrometric parameters will be absolute, in the sense that they will be defined with respect to a unique inertial reference frame. For solar-system objects the catalogue will include the observed positions together with their orbital elements. The latter are in particular needed to thread into a single identifier the $\sim$60 observations of a particular source and also to confirm the detection of new sources. Uncertainties (and correlation coefficients when relevant) will be provided for all the astrometric quantities.

Spectrophotometry: The most precise photometry for each source will be based on the Astro field observations, made from unfiltered light passing through the optical system of Gaia. This very broad passband, covering the spectral range from 350 nm to 1000 nm, is called the G band. For each source a mission-averaged G magnitude will be provided plus a measure of its variability. For sources with significant variability the epoch photometry will also be provided. Colour information will be available from the red (RP) and blue (BP) spectrophotometric fields for all sources – these data will be in the form of mission-averaged low-resolution spectra. Again, for variable sources the epoch spectrophotometric data will be provided. Uncertainties will be provided for all estimated quantities.

Spectroscopy: For sources brighter than about 17th magnitude in G the RVS field will provide spectroscopic observations, the principal aim being the determination of radial velocities. Final mission-averaged radial velocities and spectra will be provided for the sufficiently bright sources, together with the uncertainties and (if applicable) the temporal variability of the radial velocities.

Classification: Proper treatment of the data demands that some minimal astrophysical characterization of each source be carried out. Thus an indication of the nature of the source will be given, based on the combined astrometric, photometric and spectroscopic data. This will include (but not be limited to) whether the source is stellar or extragalactic. Possible duplicity will be indicated. For the purposes of defining the astrometric reference frame, QSOs will be identified. Solar-system objects will mainly be recognized by their large apparent motion detectable during a single FPA crossing. Due to their high science priority, stars with extrasolar planetary candidates will also be identified.

Astrophysical Parameters: The scientific motivation of Gaia also demands that some further characterization of the stellar sources be made. Based on the mission-averaged spectrophotometric data and astrometric data, the effective temperature, metallicity ([Fe/H]), surface gravity ($\log g$), and interstellar extinction will be estimated. It may not be possible to reliably estimate all of these quantities for all sources, depending on their magnitude and location in the HR diagram, but best estimates and uncertainties will be provided.

The above describes the minimum set of science quantities making up the Primary Science Data **as mandated by the SMP**; these have the highest priority in terms of data processing goals and are addressed by the core Consortium activities. However, this is only a subset of the Final Science Products. Other products are necessarily produced for the purpose of validating the Primary Science Products, are inevitable or necessary by-products of the data processing, or are provided as a service to the wider scientific community to render the complete set of Final Science Products accessible and useful. Many if not most of these products are indeed 'science products' in the sense that they are scientifically interesting by themselves. An excellent example is the PPN parameter γ, which appears as a global parameter in the astrometric reduction and must be solved simultaneously with the parallaxes. To determine it later would require the entire global astrometric solution to be redone, as the parallaxes are highly correlated with the value of γ.

As a rule, the Final Science Products derived as a service to the community include scientific quantities that demand processing of the raw pixel-level data and/or a 'global' processing approach. Other examples include: Mass determinations of a small subset of solar-system objects using the Gaia astrometry and a global reduction method; spin rates, axis orientations and size determinations for a large number of asteroids from the raw pixel data; a complete variability analysis of variable sources to determine periods and classify the variability type. The combination of astrometric, photometric and spectroscopic data will allow a thorough analysis of binary systems, including eclipsing binaries, that will prove useful for validating the quality of the photometry, spectroscopy and even the astrometry. The spectroscopic data for brighter sources will allow abundance analysis of some specific elements, as well as an estimate of stellar rotation.

Among the Final Data Products will also be a large set of data generated during the processing and then later used in further processing, the so-called intermediate data products. Some will be of no immediate astrophysical interest, but nonetheless essential for a complete characterization of the instrument and observation process. These include the attitude data and all geometric, photometric and spectroscopic instrument calibrations. Others, however, may have scientific content. All of these quantities will be made available as well, in order to allow sufficient transparency of how our final results were arrived at, and in order to allow re-processing of any part of the data reduction that members of the broader scientific community might deem necessary. For "scientific processing" this is fundamental: the science products of the DPAC processing must be amenable to independent verification. Also, we do not expect the DPAC to produce the final word for some of the higher-level data products, such as classifications. Any intermediate products used to derive these will also be made available as part of the Final Science Data to allow future refinements.

Scientists who now comprise the DPAC have worked for several years on the division between what should or should not be in the Gaia catalogue. These views have been discussed with ESA through the Gaia project scientists and presented to ESA committees. During these discussions our thinking about the Gaia catalogue was not only based on internal and ESA thinking about Gaia, but also on the data processing experience of Hipparcos and its catalogue production and on the working knowledge of DPAC members from other large-scale surveys. The motivations for having the structure we propose is based on the following facts:

- the importance of determining calibrators for astrometry, photometry and RVS;
- the importance of having stringent procedures for validation of the calibrations;

- the heaviness of the computational effort;
- the presence in the time frame 2015–2020 of other large-scale surveys;
- the lack of reliable observations matching Gaia accuracy to provide an external validation of the results.

While we have made some effort in deciding what scientific processing the DPAC should and should not do, based on the criteria mentioned above, it is clear that prioritisation is necessary. Should it be difficult to carry out additional tasks within the strict schedule, a decision will be made by the DPAC Executive, at the appropriate time and with the appropriate consultations, to adapt the objectives to the resources. However at the present time the final data products presented in this document are included in the DPAC baseline proposal and should be part of the Final Catalogue content. The Consortium has gathered into a single organisation the means (human, financial and above all organisational) to carry out the full data processing in a very consistent way and within the shortest schedule possible. Although one can argue that what the DPAC can do could obviously be done by other groups, there is no guarantee that this will be done on a comparable timescale, with the same overall consistency and with every step carefully documented. The AO for CU9 to come later will give the opportunity to better define the actual limit of the processing and how some of the tasks we describe (very few indeed) may overlap with the final data access.

2.1.4 Intermediate data releases

The final goal of the data processing is to produce the astrometric-spectrophotometric catalogue described above with final accuracies meeting the mission goals. Given the timeline specified by ESA in the Science Management Plan, this catalogue will be available around 2020. However, it is fully understood by the DPAC that one will produce one or more intermediate catalogues in the course of operations. In addition to its scientific value, such an effort is well motivated both as a means to spur the development of advanced catalogue access tools and to prepare the larger scientific community for the final Gaia catalogue. The content, level of validation and documentation of intermediate data releases will be specified together with the GST, as indicated in the SMP. Whether other kinds of intermediate results could be made available on a more continuous basis (e.g. epoch photometry, variability indices, clean spectra, minor planet positions) will be considered in due course together with the GST. Such a decision will have to be weighed carefully to avoid disseminating insufficiently validated data, a situation that will have a bad impact on the image of the mission, and also to avoid diverting the DPAC resources into the preparations of these releases.

2.2 Structure of the data processing

2.2.1 Outline of the Data Processing task

The Gaia data processing consists of several iterative processes dealing with astrometry, photometry and spectroscopy. These iterations reflect the fact that, as a survey satellite reaching accuracy and completeness levels never obtained before, it has to be self-calibrating in many aspects of its observations (see Sect. 5.6 for a more precise statement). The most obvious in this respect is the astrometry: were the attitude perfectly known independently of the observations (say from a set of ultra accurate gyros mounted on the spacecraft or any other attitude sensors based on a subset of well positioned stars or quasars), the positions of the stars would be rather easily obtained. Conversely a good initial catalogue at the few μas level would tell us quickly how the satellite rotates and make the data processing much simpler, but this is not what global space astrometry is all about. Gaia has neither of these and must be fully self-calibrating, solving the attitude, instrument parameters and star positions (and much more) from the transit data.

Therefore the same data that will ultimately form the astrometric catalogue are also used to reconstruct the scan-phase of the satellite (along-scan attitude), which is the reference frame for the astrometric measurements, and to determine the instrument calibration parameters. Similar considerations, but with less entanglement, apply to the spectroscopic and photometric data, where reference stars will be used to make the initial calibrations and determine the photometric system. Additional complications arise from the fact that these three main iterative solutions are interlinked and rely on common calibration data, making the Astrometric Global Iterative Solution (AGIS) the core of the data processing.

Iterations affect all three main steps in the data analysis: the pre-processing and calibration of the raw observations (CCD images), the iterative solutions for the astrometry, spectroscopy and photometry, and the various shell processes such as variability, automatic classification and double stars. Information on duplicity as accumulated over the mission, for example, will affect the pre-processing retrospectively, and will thus affect the astrometric and photometric solutions. Similar considerations apply to the detection of variability.

Data exchange plays an important role in ensuring compatibility amongst all groups. It is envisaged that the data communicated between groups and databases will follow rigid procedures based on Interface Control Documents (ICD), defining formats and changes to formats that will need to be established on an individual basis. This will also apply to the way in which software will be developed and maintained. Clear procedures following a software configuration and versioning control standard will be devised and implemented, and applied at all processing levels, central and distributed. These procedures will allow algorithm providers to implement their code according to the standard, and will also provide them with procedures to follow pertaining to the source of data, the toolboxes and parameter database to be used.

2.2.2 The major processing steps

The processing starts with the arrival of a data set sent to the ground and develops through successive steps. These steps are described in more detail in this document. The following just aims to summarize the main steps and to highlight their impact on the overall DPAC organisation.

2.2.2.1 Initial Data Treatment

The Initial Data Treatment (IDT) transforms the most recently arrived telemetry flow into a more convenient form, determines basic image parameters, links the new observation with previous observations (sources) in the main data base and derives various auxiliary data. The first part consists of uncompressing, rearranging and reformatting the data to create raw information ready for storage in the main database. This part is not strictly data processing since there is no change or addition to the information content.

To insert the observations into the Mission Data Base (MDB) they must be properly identified with sources already observed and entered in the MDB at an earlier time. (This process is referred to as *cross matching*). Observations that have no corresponding MDB source are entered as new sources. This phase needs to have access to (i) a low-precision attitude, which will be based on a smoothed version of the onboard attitude, and (ii) the 1D or 2D centroiding for each source, which is equivalent to the transit time of the center of the image for every source on every CCD. This is the fundamental astrometric measurement. At the same time an estimate of the source intensity is made for the photometry. The transit times (centroids) are transformed into local geometric coordinates, and preliminary (rough) sky coordinates, using the available calibrations and attitude. Moving sources (primarily solar system objects) will be recognised by combining the attitude information, the geometric calibration and the transit times on the different astrometric CCDs during a single focal plane crossing. Transit times and fluxes will be updated every six months as better calibration files become available.

The result of the IDT is the insertion of the observations into the main data base, properly linked to already identified sources in the sky.

2.2.2.2 Astrometric Core Solution

The astrometric core solution is the cornerstone of the data processing since it provides the calibration and the attitude solution needed for most of the other treatments, in addition to the astrometric solution of about 100 million *primary sources*. The main equations to be solved relate in a most general way the observed position on a detector to a comprehensive astrometric and instrument model as,

$$O = S + A + G + C + \varepsilon \tag{1}$$

where

- O is the observed one-dimensional location of the source at the instant determined by the centroiding algorithm applied to the observed counts.
- S is the astrometric model. For the primary stars this comprises only the five astrometric parameters $(\alpha_0, \delta_0, \pi, \mu_\alpha, \mu_\delta)$ respectively denoting the position, parallax and proper motion components.
- A represents the parameters used to model the instrument attitude over a given interval of time. They are, for example, the cubic spline coefficients of the quaternion describing the orientation of the instrument in the celestial reference frame.
- G represents the global parameters such as the PPN parameters or other relevant parameters needed to fix the reference frame of the observations.
- C comprises all the parameters needed for the instrument modelling: geometric calibration parameters (both intra- and inter-CCD), chromaticity effects, basic angle, and CTI offsets. In practice, from the processing viewpoint, there is no real difference between the C and G parameters.
- ε is the Gaussian white noise which can be estimated from the photon counts and centroiding for every observation. It is used to weigh the equations. A test is performed at the end to validate the assumption on the noise.

The way the system is solved by successive iterations and the convergence ensured is described, along with results of extensive testing, in Sect. 5.1.

2.2.2.3 Photometry

Broad-band photometry will result from the fluxes measured in the Astro CCDs during the star transits, while the dedicated low-dispersion prisms will provide dispersed images in blue and red bands. The data processing will comprise two main phases: (i) extensive calibration of the photometric responses and wavelength calibrations of the photometers; (ii) computation of calibrated photometry for every sources. The first phase is a global treatment of reference sources, which are supplied initially by ground-based observations. (Later they will be redefined internally so that the whole system is self-calibrating.) The second phase of the data processing is treatment on a star-by-star basis. Photometric processing will provide colour information which is used for colour dependent calibration and for the chromaticity correction of the astrometry. The photometric results must, therefore, be available for the processing of the next astrometric data set. Tests for variability are performed on the epoch photometry and will become more and more significant with the accumulation of data.

The variability analysis will start from fully calibrated fluxes, with the G-band being the most reliable indicator of variability. This data will be extracted from the main data-base with a delay of one data reduction cycle (Sect. 7.3) compared to the photometric treatment. The complete analysis of the variability (type of variability, periodic or double-mode oscillation searches) will combine the photometric and spectroscopic data (and use the astrometric position as well) to achieve a global analysis of the variability phenomena in stellar evolution (see Sect. 5.2.5).

2.2.2.4 Spectroscopy

The spectroscopic data are acquired by the spectroscopic instrument (RVS) and associated detectors. The main input for the processing consists of the raw spectra in the form of photon counts recorded over $\sim$1100 pixels. The main output of the processing will be the (time-dependent) calibration parameters of the RVS giving its response and wavelength mapping onto the pixels, the calibrated and cleaned spectra at each observation epoch for the 100 million sources observable, and the radial (line-of-sight) and rotational velocities. The system has no internal calibration. Calibration must be carried out with the data, specifically using (i) some standard reference sources and (ii) an iterative procedure based on bright and stable sources. Using the initial instrument parameters available for this cycle, the radial velocity of the bright sources is determined, then the selection of the reference source is refined. This last subset is then used to solve the calibration model (primarily the wavelength mapping). The system is iterated until convergence. Then the calibration model is applied to all other sources to derive the radial velocity and generate the calibrated spectra which are further analysed for astrophysical content. As the mission progresses all sources will have several transits available allowing a summation of the individual scans (testing beforehand for variability, e.g. spectroscopic binaries). The radial velocity will be determined from the combined spectra, again by cross-correlation with an appropriate template. The template will be selected using the auxiliary information coming from the photometric processing.

2.2.2.5 Object Processing

Once the data have gone through the set of global processing it remains much to do. First the astrometry of the non primary stars, several hundred millions, that must be solved at every cycle as more observations are available and better calibration can be used. This is also from the residuals that many non single stars (and planets) will be first spotted and sent to a dedicated treatment. In this case several models will be tried to determine either an accelerated motion or some or all the orbital parameters. This is typical of the object processing where many individually small processings will have to be run billions times in total. The information gathered will be sent back to the start of the processing to update the initial data and this non single stars will be removed from the calibration data. The same will happen with solar system objects which are detected outside the main processing chains and solved for astrometry and photometry before the information is returned to the data base. The determination of the astrophysical parameters requires very specialised algorithms (automated classification based on many parameters, temperature and chemical composition determination) and it uses all kind of data resulting from the earlier global processing. The results is primary classification, but also flagging of sources for the iterations(e.g QSO are recognised during the classification and this is crucial to define the clean sample to build the inertial frame, temperatures are needed to select the proper synthetic spectra for radial velocity determination).

2.3 The challenges

Here we describe the expected end-of-mission precision of the main science products, and the principle challenges of the data processing that must be overcome to reach these performances. We also discuss the technical challenges associated with the data processing itself, leaving aside the significant technical challenges that the realization of the Gaia instrument itself must overcome.

2.3.1 Target performance

Performance requirements on the accuracy of the final data products are specified in the Mission Requirements Document (MRD). The target performance for the data processing, as described below, is to go beyond these requirements.

2.3.1.1 Astrometric performance: First and foremost, Gaia is an astrometric mission. Consequently, the astrometric accuracy goals for Gaia are also the most challenging. As already mentioned in section 1.2, the accuracy is almost two orders of magnitude higher than that achieved by the Hipparcos mission.

The end-of-mission precision of the astrometric parameters of individual stars depends strongly on their magnitude and colour, and to a lesser extent on their location in the sky. Representative sky-averaged values for the expected parallax precision are displayed in Tab. 2. The corresponding figures for the coordinates (right ascension and declination) at a mid-mission epoch and for the two components of the annual proper motions are similar but slightly smaller (i.e. better, by about 15 and 25 percent, respectively, for a 5-year mission).

The astrometric data reduction challenges are manifold. They start with the treat-

Table 2: End-of-mission parallax precision in microarcseconds (μas). Representative values are shown for unreddened stars of the indicated spectral types and V magnitudes, as given in ESA's Mission Requirements Document (third column).

Star type	V magnitude	2005 formal requirement
B1V	< 10	7
	15	25
	20	300
G2V	< 10	7
	15	24
	20	300
M6V	< 10	7
	15	12
	20	100

ment of the raw image data, where the astrometric centroid of CCD images has to be determined to the order of one percent of the pixel size or better – in the presence of photon noise and CCD radiation damage effects. This is a delicate signal processing task which has to be performed for 10^{12} images. If it took 1 millisecond to process one image, the processing time for just one pass through the data (on a single processor) would take 30 years. Needless to say, the adopted approach is much faster.

Two more pieces of information are needed to give the (pixel) centroids some astronomical meaning: Firstly, for each and every moment during the five years of scientific measurements the instantaneous orientation (attitude) of the rotating Gaia instrument in space must be determined to an accuracy of about 100 microarcseconds. Since the spacecraft and optics have a diameter of 3 meters, this implies determining the relative location of any two parts on its outer periphery to about 1.5 nanometers with respect to inertial outer space. Secondly, the path of the light through the optics and onto the CCD detectors must — in some sense — be known to similar precision. For the spacecraft designers this means to guarantee that the geometric configuration of the whole meter-sized instrument be stable at the nanometre level over a period of several hours (the rotation period). For the data reduction this means to implicitly establish the geometry to the nanometre level. This calibration task, as well as the attitude determination, must be done with nothing but Gaia's own CCD images of stars as input.

So far we have discussed only the precision of individual measurements and the astrometric parameters for individual stars. Some scientifically important quantities can in principle be derived to even higher precision from the averaging over many stars observed by Gaia. Examples are the space motion of the Magellanic clouds and of star clusters, or the average parallax of some particular group of stars. This can only be done if the accidental errors of the Gaia data indeed cancel out in the formation of such averages, i.e. if there are no systematic errors in the data processing results. Control of systematics introduced by the instrument and/or the data processing methods is, therefore, a critical issue for the data processing. It means that for some aspects of the attitude and geometric calibration the ultimate goals go beyond the already stupendous numbers given above. One particular aspect of high relevance for the question of systematic errors lies outside Gaia: In order to correctly interpret Gaia's raw measurements in terms of astronomical positions and motions it is necessary to model the path of the light rays through the bent space-time of the solar system to the order of 1 microarcsecond. Likewise, we must know the three-dimensional velocity of the Gaia spacecraft with respect to the solar-system's centre of mass to about 1 millimeter per second. The former challenge creates the task of producing an unprecedented high-precision general-relativistic theory of light propagation; the latter challenge will probably necessitate astrometric observations of the Gaia spacecraft from ground-based optical observatories during the mission. Ironically, the reduction of those auxiliary ground-based observations can reach the required absolute precision only after a preliminary Gaia star astrometric catalogue has been produced.

2.3.1.2 Photometric performance: The main merit of the Gaia photometry is that it will cover the entire sky on the same photometric system, yielding a photometric catalogue of unprecedented homogeneity and depth. From the Astro measurements of unfiltered (white) light, Gaia will produce G-magnitudes, while the spectral energy distribution of each source will be sampled by a dedicated spectrophotometric instrument providing low resolution spectra in the blue (BP) and the red (RP), see Sect. 3.2.

The estimated precisions for the G-magnitudes per focal plane transit and at the end of the mission are shown in Fig. 3. Taking into account the photon noise from the source and the sky-background as well as the read-out noise, precisions of ~ 10 and ~ 1 millimag are achievable at $V \sim 19$ per transit and at the end of the mission, respectively. This implies that the precision of the G measurements is ultimately limited by the calibration errors. In addition, the G passband yields the best signal-to-noise ratio for variability detection among all of the Gaia measurements.

The signal-to-noise ratio for the BP and RP spectra at the end of the mission for several main sequence stars has been estimated from the photon noise of the source, the background and the read-out noise. These precisions are shown in Fig. 4 for several apparent magnitudes and absorptions. The main limitation in case of bright objects is due to the calibration errors, which are not included in the figure.

The end-of-mission source spectra will be used to derive final astrophysical parameters (APs) for each source. The astrophysical parametrisation performance based on the expected final BP and RP spectra was evaluated using the Figure of Merit

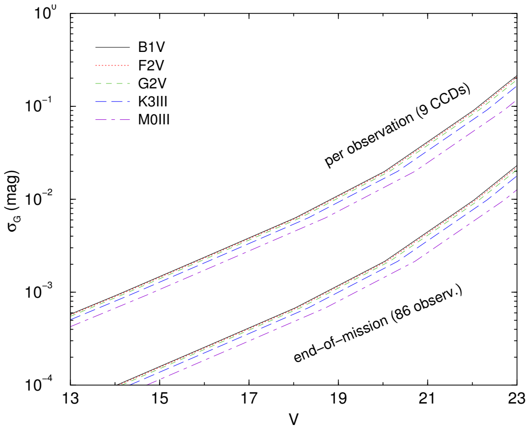


Figure 3: Estimated precision of the G magnitude, as a function of V, per focal plane transit and at the end of the mission without the calibration errors.

method described in [JHBea06]. The method takes into account the precision of the measurements and the sensitivity of each spectral element to the changes of the source astrophysical parameters. The information from the derived parallax has been included as 'a priori' information. Fig. 5 shows the predicted errors for an F-type subgiant star, a key target for the analysis of the halo, thick and thin disks. At $G = 17$, the estimated precisions are about 0.05 mag for absorption, 0.25 dex for $\log g$, 1.5% for the temperature and 0.1 dex for [M/H] for solar chemical composition.

The precision estimates from the Figure of Merit method represent the achievable performance on the assumption that the instrument noise model is correct and that the synthetic spectra which are used to calculate the sensitivity of the measurements to the astrophysical parameters represent true stars. In addition this method only takes into account local degeneracies between astrophysical parameters such as the one between effective temperature and absorption. Hence the method assumes that sources can be correctly classified according to spectral type based on the photometry. In practice we will have to deal with global degeneracies where, for example, a reddened O-star may be confused with a K-dwarf (at low spectral resolution). These global degeneracies require the development of methods that are capable of both identifying the spectral type of stars and providing refined estimates of their astrophysical parameters. How this can be achieved in practice is described in section 5.5.

The principal challenges of the photometric data processing are:

- Developing an automated and robust calibration procedure for the G-band fluxes measured from the AF images. The calibration has to correct for large scale (CCD to CCD) as well as for small scale (from one column to the next in a CCD) response variations. The evolution in time of the instrument response curves (mirrors, CCD QE, etc.) and the effects of charge transfer inefficiency – leading to increasing signal loss during the mission – have to be accounted for

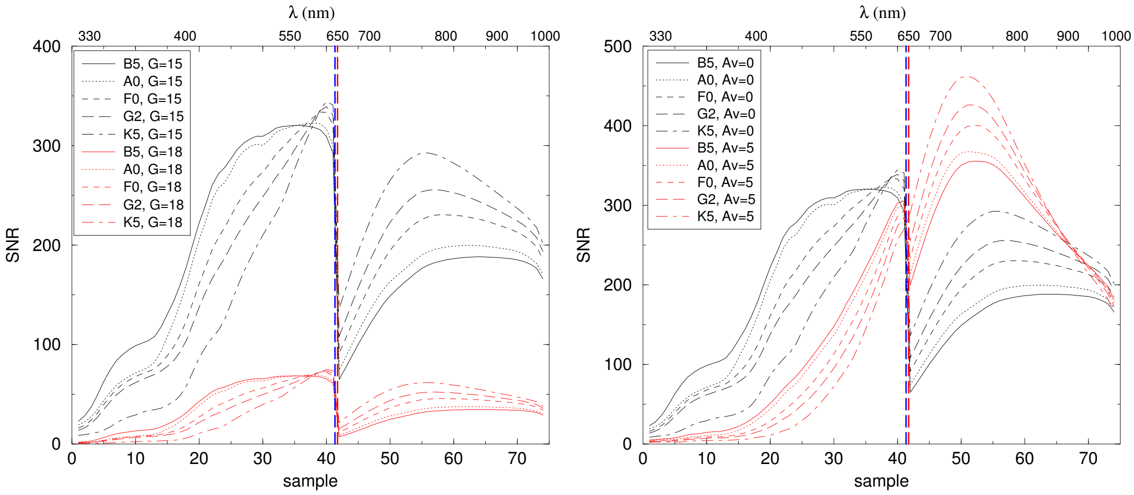


Figure 4: Estimated end-of-mission signal-to-noise ratio for the BP and RP spectra for several main sequence stars without including the calibration errors. Left panel: for unreddened stars with $G = 15{,}18$; right panel: for two different values of A_V and $G = 15$. In each panel the BP spectra occupy the left half and the RP spectra the right half.

in the internal flux calibration. The calibration procedure will have to be set up for each of the CCD gates that will be used to restrict the integration time for bright stars. Sections 5.2 and 5.2.2 provide more details on the photometric calibration issues.

- Implementing methods for the extraction of BP/RP spectra and for making sky background measurements from the raw image. These must be robust in crowded regions where the dispersed images and measurement windows from different sources overlap.

- As described in section 5.2.1 the measured signal in BP and RP will depend on many parameters (dispersion curve, PSF/LSF, response curves, CTI, etc.) which will evolve during the mission and which differ from one position to the next in the AC direction. The latter is especially important for the prism dispersion curves which vary by around ±5% in the AC direction. This means that the combination of all the measured BP/RP spectra for a particular source into a single average end-of-mission spectrum is a difficult data processing challenge requiring a detailed understanding of the signal content. The variation of the signal parameters in time and across the focal plane will also complicate any variability analysis of the spectra.

- All photometric measurements will have to be tied to an absolute flux scale and this requires the development of external flux calibration methods that are capable of dealing with the vast array of different sources (stars across the

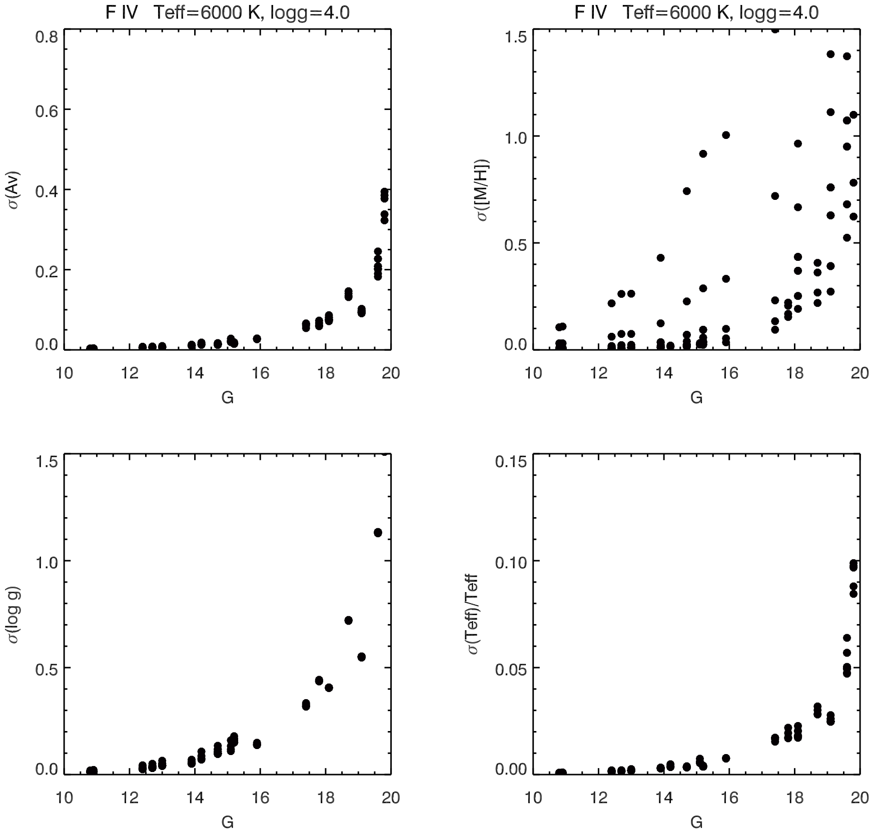


Figure 5: Precision of the main stellar astrophysical parameters for an F IV star with several [M/H] values and at different locations in the Galaxy (different distances and absorptions), estimated from simulated BP/RP spectra using the Figure of Merit method described in [JHBea06], which includes the parallax information. The calibration errors are not included.

whole HR-diagram, quasars, galaxies, solar system objects, etc.). As part of this work extensive ground based preparations, consisting of spectrophotometric observations of standard sources over the wavelength range 300–1100 nm, are required (see section 5.6).

- In addition to the fluxes the wavelength scale for the BP/RP spectra must be calibrated. This will involve a wavelength and focal plane position dependent geometric calibration. Identifying the most suitable sources and/or features in the spectra to determine the zero-point of the wavelength scale is a major challenge (see section 5.6).

More details on the technical demands of the photometric data processing can be found in [Bro06a] and [Bro05].

2.3.1.3 Radial velocity performances: The Radial Velocity Spectrometre (RVS) will collect the spectra of about 250 million stars down to magnitude $G_{RVS} \simeq 17$ (for the redder stars). Each source will be repeatedly observed (on average 40 times over the 5 years of the mission). With about 10 billion spectra, the "Gaia-RVS spectroscopic catalogue" will be the largest of its time. Table 3 presents the Gaia specifications for the end-of-mission radial velocity performance.

Table 3: Gaia specifications for the end-of-mission radial velocity performance (in km s^{-1}) as function of V magnitude.

	V	RV specifications km s^{-1}
B1V	12.0	15
G2V	16.5	15
K1IIIMP	17.0	15
B1V	7.0	1
G2V	13.0	1
K1IIIMP	13.5	1

The main challenges of the spectroscopic processing are:

- **Disentangling the spectroscopic observations.** The RVS is an integral field spectrograph, without slit or fibers, dispersing all the light entering its field of view. As a consequence, the spectra of neighbouring sources will overlap. This will happen frequently in the dense parts of the Milky-Way such as the Disk and the Bulge (in particular in the low extinction windows). In case of overlap of spectra, the contributions of the different sources will have to be separated. The method foreseen is to iteratively model and subtract all the sources contaminating a given object, considering alternatively each source as the primary source of interest and then as a contaminating object (see 5.3.2).

- **Calibration of the RVS.** Because of the special operational conditions of the RVS, it was not possible to equip it with calibration devices (such as calibration lamps or absorption cells). Therefore, instead of using dedicated calibration exposures to calibrate the characteristics of the spectrograph, the RVS will be self-calibrating, i.e. it will be calibrated using its own observations of reference sources. Some of these reference sources will be ground-based standards, but the larger fraction will be characterized and qualified as reference stars using the RVS observations themselves. The process of characterising the sources, identifying those suited as calibrators and using the appropriate ones to calibrate the spectrograph will be iterative. This process is called the Spectroscopic Global Iterative Solution (SGIS) and is described in section 5.3.1.

- **Analysis of 10 billion spectra.** The "Gaia-RVS spectroscopic catalogue" will be larger by several orders of magnitudes than the largest existing spectroscopic catalogue. One single second of computing time to analyse one spectrum translates into 300 years for the analysis of the 10 billion spectra (fortunately spread over many computers). The size of the RVS sample already represents a challenge in itself, so a dedicated and optimized algorithm will be developed to analyse the spectra in a reasonable amount of time. In addition to the simple number of spectra, the complexity will come from the heterogeneous nature of the data to analyse: a very broad range of magnitudes and therefore of signal-to-noise ratios, different sampling strategies (see 5.3.2), many epochs of observations, all possible types of stars at any possible stage of their evolutions. This diversity of observations will require us to implement flexibility and robustness in the spectroscopic analysis algorithm.

2.3.2 Technical challenges

For a space astronomy mission Gaia produces an unprecedented volume of data, 40 to 50 GB every day for 5 years leading to almost 100 TB of compressed data. The data volume itself is fairly daunting but certainly manageable with the technology which should be available in the 2015–2020 time frame.

The real challenge for Gaia lies in the processing of this data. Each piece of data will have to be accessed repeatedly and frequently as the catalogue is iterated toward its final accuracy. Access will need to be spatially optimized for some algorithms while others require temporal access. Data access and data management are a major challenge for Gaia. If not handled correctly, access to the data and writing of results will cause an insurmountable bottleneck for the processing tasks.

The shear complexity of the processing (as discussed in Sect. II) has led us to break the system down into manageable chunks which may be constructed and operated autonomously in distinct institutions. The actual products are not, however, so neatly partitioned, resulting in a complex processing system with many dependencies. The construction of the software system for any one of the subsystems is itself a challenge: few IT projects with a five year horizon proceed or end as expected. When we take the coupling of the subsystems into account the problem is increased. While we believe we have a management approach to support the construction of this system but we are well aware of the technical and sociological difficulties which lie in our path [O'M05].

The distribution of the processing is a boon because it alleviates any one institution from having to construct the world's largest supercomputer (which would be needed to do all processing in one place). The bane of the situation is that data need to be shipped continuously to and from the various data processing centres. The data volume to ship increases with the mission lifetime. We estimate[1] a possible 200–300TB of data to be moved in 2017. We are optimistic about the availability of high speed networks when we need them, but it is not clear they will be affordable to our project (the physical transport of storage devices remains a viable option).

On the processing front, let us recall that Gaia sees 10^9 objects about 80 times. Each viewing yields 10 Astro samples plus spectra and photometric data so we may assume a total of roughly 10^{11} data samples (equivalent to images). If we assume all processing (including database search and I/O) requires one second per image, the data processing would take 10^{11} seconds or over 3000 years on a single computer. Yet the Gaia processing is significantly more complex, requiring iterative reprocessing of the data. Doing this in a reasonable amount of time (i.e. months) is our greatest technical challenge. To reprocess the data in one month we would need to process 400 objects, or about 40000 samples, per second. The current rough FLOP (FLoating-point OPeration) estimate for all Gaia processing is 10^{21} FLOPs with much of that coming at the end of the mission. Even distributed over five years this would require a $6x10^{12}$ FLOP/s machine. That is equivalent to approximately 3000 high end desktops. Hence whether the computing power is central or scattered across Europe we have no choice on the processing model: it must be massively distributed.

---

[1] The Main Database is estimated to be about this size toward the mission end but could be larger

3 Gaia Instruments, measurements and raw data

This section describes the functionality of the satellite instrumentation: what is detected and the data sent to ground.

3.1 Astrometry

Gaia will derive astrometric measurements from the data obtained in the astrometric part of the focal plane instrument. The raw data will be collected while the satellite slowly rotates about an axis perpendicular to the two viewing directions of the telescope. The two Gaia fields of view (separated by 106.5 degrees) continuously move ('scan') over the sky in a pre-defined pattern called the nominal scanning law, , which is described in Sect. 3.5 and illustrated by Fig. 14 in that section. The satellite rotation causes the star images to move across the CCD detectors on the focal plane. The accumulating photo-electric charges are actively transported across the CCDs in synchrony with the images, the so-called TDI (time delay integration) mode of CCD operation. The charges reach the read-out register of a CCD when the corresponding image leaves that CCD at its trailing edge.

Fig. 6 shows the present (and probably final) layout of the joint focal plane of all Gaia instruments. The main astrometric observations are obtained with the 62 CCDs labelled 'AF' (Astro Field). For data rate and read-out noise reasons, the AF CCDs are not read out entirely. Instead, only small 'windows' around the target star images are processed. The vast majority of the CCD pixel data (those containing dark sky) are electronically dumped. Due to the rectangular apertures of Gaia's telescopes the bright cores of the diffraction images are box-shaped, and the far, faint diffraction wings are cross-shaped. The image structure is shown in the left-hand panel of Fig. 46.

In both dimensions (called 'across scan' and 'along scan') the image cores are well resolved at the CCD pixel level. However, again for the sake of reducing the data rate and read-out noise, up to 12 pixels in the across-scan direction are combined at CCD readout (on-chip binning) resulting in one-dimensional 'images'. Figure 7 shows representative examples of the 'raw' AF data, for sources at two different magnitudes.

Perhaps surprisingly, the across-scan binning improves the overall astrometric performance of Gaia, although it causes a complete loss of all across-scan information in the images. The resulting astrometric measurements from the AF CCDs are thus one-dimensional: two-dimensional astrometric information is gained by scanning every source in many grossly differing scan directions. The observability of time-dependent astrometric effects like parallax, proper motion and binary-star revolution is guaranteed by the chosen nominal scanning law, which provides a globally optimized temporal and directional distribution of the individual visits to each sky field. Over the five years of Gaia's science mission every star in the sky on average will be vis-

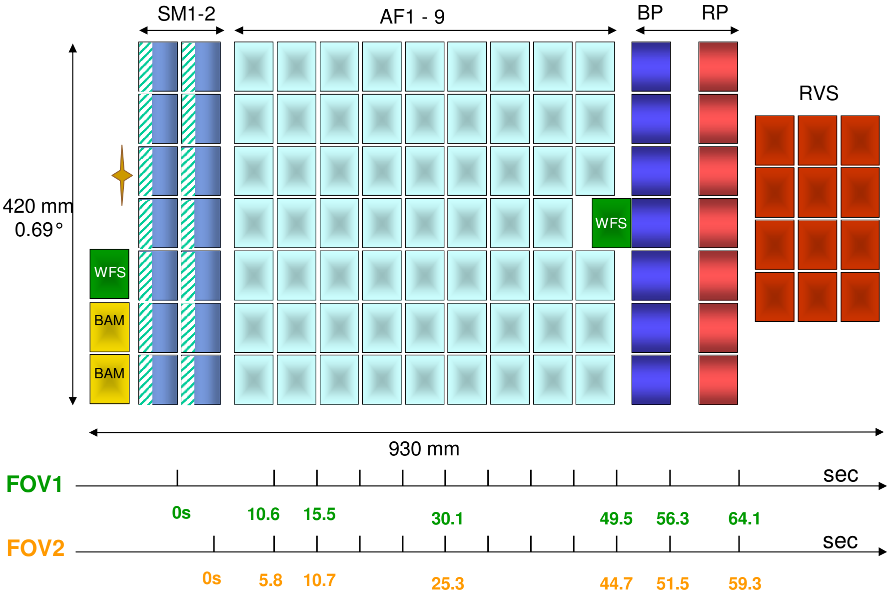


Figure 6: The Gaia Focal Plane Array (FPA). Each coloured rectangle indicates one CCD of approximately 4 cm by 6 cm size. The direction of star image motion is indicated at the bottom. It takes a star image about 4.4 sec to cross one of the CCDs and this defines the effective exposure time for the individual scientific observations. The various acronyms denote the scientific instruments of Gaia which, from left to right, are: the sky mappers (SM), the main astrometric field (AF), the blue photometre (BP), the red photometre (RP) and the radial-velocity spectrograph (RVS). In addition, there are detectors for auxiliary instruments: the basic angle monitoring (BAM) system and the wavefront sensors (WFS). Note the enormous size of the overall assembly. In this figure the 'across-scan' (AC) and 'along-scan' (AL) directions defined in the text are vertical and horizontal, respectively. The two bottom lines give the time to reach the different parts of the field from SM1 or SM2.

ited about 80 times by one of the two fields of view. This corresponds to about 700 individual astrometric observations (nine per field-of-view transit).

In order to separate the windows containing actual star images from the majority of dark-sky CCD data, the on-board system in charge of CCD controlling and science data handling must know in advance where and when a star image will leave an AF CCD. The 14 CCDs labelled 'SM' (star mapper) in Fig. 6 are therefore read out completely and the resulting data stream is subjected to an autonomous image detection process. Next, a numerical centroiding process precisely determines the centres of all detected SM images which are sufficiently bright to warrant an AF observation. Using real-time knowledge of Gaia's rotational motion (from the on-board attitude control system), the times and locations of the forthcoming exits of the same image from the suite of AF CCDs are predicted. The first window produced from these predictions, collected in one the AF1 CCDs, is used to check ('confirm') the reality of the detected SM image (i.e. to discriminate against cosmic-ray events, noise spikes and other detection process artifacts). If this check is passed, the other eight windows are read out, and the whole suite from SM to AF9 is stored for transmission to the ground. Each of the two strips of SM CCDs (SM1 and SM2 in Fig. 6) sees only one of Gaia's fields of view (necessary to decide to which field a detected image belongs) while all other CCDs receive superposed light from the two fields.

The SM and AF1 image centroids are also fed into the attitude control system for the precise determination and control of the rotation rate of Gaia around all three axes, which in turn is used in the window prediction process described before.[2] The SM images have the same structure as the AF images, but are sampled differently. In particular, in order to determine the rotation rate and window locations across scan, the read-out scheme must conserve across-scan resolution. Only a mild binning is therefore applied to SM (2×2 pixels) and AF1 (1×2 pixels).

The ASM and AF data are collected by CCDs without filtering the light. In addition to their primary astrometric use, they are also used for broad-band photometry. For the spectral response see the discussion of the unfiltered photometric band, called 'G', in Sect. 3.2 and Fig. 8.

A certain inhomogeneity of the sky coverage is unavoidable, even with the optimized nominal scanning law of Gaia. The major location-induced effect is a dependence on ecliptic latitude for the number of observations, displayed in Fig. 15 in Sect. 3.5. Stars at about ±45 degrees ecliptic latitude receive more visits by Gaia, stars at low ecliptic latitudes relatively few. The precision of the astrometric parameters varies accordingly.[3] More details on the location dependence of the precision can be found in [gai00, Section 8.1].

Another unavoidable consequence of the nominal scanning law is an uneven distribution of the observations over time. The field-of-view transits of a given star occur in tight groups separated by weeks or even months. Within each group, the time separations of the transits alternate between 1.8 hours and 4.2 hours. The size of a group may range from only one transit to slightly more than a dozen (maximum total duration of the group about 40 hours). Small groups are more frequent, although most groups contain at least two transits. Note that each field-of-view transit produces one SM and nine AF observations.

It is beyond the scope of the present document to describe the mathematical foundation of the nominal scanning law, of the necessity of the two superposed fields-of-view, and of the transformation from the raw image centroids to the final astrometric

---

[2] Both the TDI mode and the window prediction from the SM images are also used for the photometric and spectroscopic instruments of Gaia.

[3] It is a slightly unfortunate coincidence that the galactic centre region lies very close to the ecliptic.

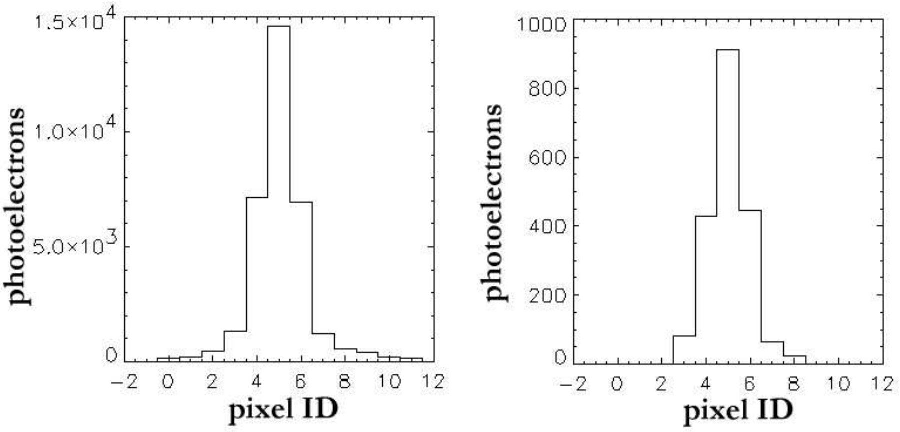


Figure 7: Examples of raw, one-dimensional AF images of two G0V stellar point sources at 15th and 18th magnitude, assuming that the source position is centered on pixel 5, near the center of the AF field. The window for the brighter source has 12 pixels (left plot), and 6 pixels for the fainter (right plot). The total assumed noise is 5.9 photoelectrons, including a sky background at 22.5 mag/square arcsecond (contributing 2 photoelectrons).

parameters (i.e. positions, proper motions, parallaxes, binary orbits). Information can be found in [EFN$^+$97, Vol. 3], in [gai00] and references therein, and in [Lin98]. The chosen spacecraft design and mission operation scheme guarantee that Gaia's raw measurements can be transformed into the aimed-at, completely independent, extremely-high-precision astrometric end results. These results will be 'absolute' in the sense that neither their internal consistency nor their precision will depend on prior knowledge of the astrometric sky.

The basic astrometric measurements of Gaia are the along-scan centroids of the AF images, which are determined on-ground from the (one-dimensional) set of CCD samples in one AF window.[4] Those centroids are in units of CCD samples, counted from the first sample in the window under consideration. The basic task of the astrometric data reduction is to define a translation of such 'sample coordinates' into (one-dimensional) position measurements on the sky, with the goal of deriving the astrometric parameters of the observed stars. This translation required derivation of two very big sets of intermediate unknown quantities:

- A model for the effective optical projection of an image centroid (in 'sample coordinates') through the optical system of the Gaia telescope onto the sky – the so-called geometric calibration
- A model for the rotational motion of that system – the satellite attitude – as function of time.

---

[4] The across-scan information from the SM images is used, too, but it plays a subsidiary role only.

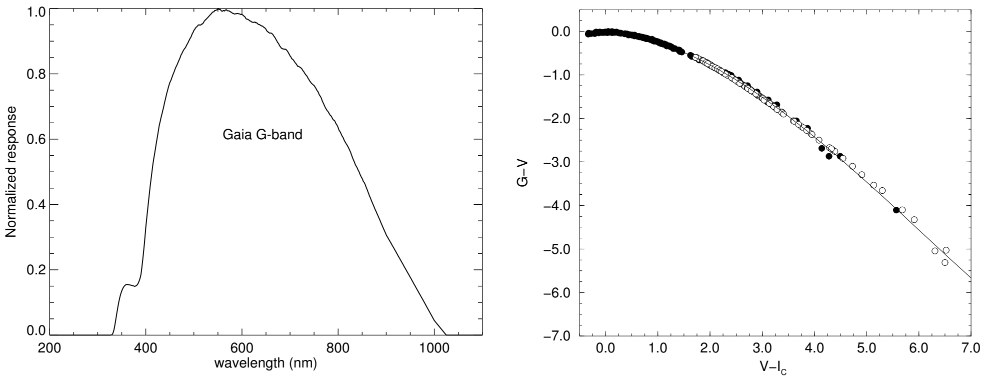


Figure 8: The left panel shows the G passband as defined by the telescope mirror coating and AF CCD QE. The right panel shows the relation between $G-V$ and $V-I_C$. Filled symbols show unreddened stars of all spectral types. Open symbols correspond to the same stars reddened by $A_V = 5$ mag. Reddening vectors run parallel to the colour-colour relationship.

Both models must have at least the same precision as the best individual centroids, i.e. a few dozen microarcseconds. In other words, the Gaia data reduction must (in some sense) determine the geometry of the telescope and focal plane to the order of a nanometre, and the rotational motion of the telescope in space to the same order. Since the only conceivable handle to attack this task are the stars in the sky and Gaia's own measurements, the astrometric data reduction must determine the attitude, geometric calibration and astrometric star parameters at the same time in one giant, self-consistent adjustment process.

3.2 Photometry

Gaia will collect broad-band photometric measurements from the data obtained in the astrometric field of the focal plane while multi-colour photometry is obtained from the dedicated photometric instrument.

The white-light photometric band, called 'G', is defined by the telescope mirror coatings (6 mirrors with silver coating) and the quantum efficiency of the AF CCDs. The G-band magnitude for a particular source is defined according to:

$$G[\mathrm{mag}] = -2.5\log_{10}\left(\frac{s}{s_{\mathrm{A0V}}}\right), \tag{2}$$

where

$$s[e^-] = (D\times H)\,\tau\int N(\lambda)T^6_{\mathrm{Ag}}(\lambda)Q_{\mathrm{AF}}(\lambda)d\lambda \tag{3}$$

is the number of photons received in the G-band from a source with spectral energy distribution $N(\lambda)$ (in units of photons s^{-1} m^{-2} nm^{-1}) and s_{A0V} is the value of s for an unreddened A0V star with $V = 0$. The Kurucz A0V star spectrum from the Gaia parameter database is used here. The other terms in Eq. 3 are the AL and AC sizes D and H of the entrance pupil, the integration time τ, the mirror reflectivity $T_{\rm Ag}$ for a silver coating, and the quantum efficiency $Q_{\rm AF}$ for an AF CCD. Fig. 8 shows the G passband and the relation between G and the Johnson V-band as a function of $V - I_C$ colour, covering all spectral types for unreddened stars and stars with A_V =5 mag.

Gaia will provide distances at the 10% accuracy level for some 100–200 million stars, which, when combined with estimates of the G magnitude and the interstellar extinction, will yield unprecedented absolute magnitudes in both accuracy and number.

Multi-colour photometry is provided by the dedicated photometric instrument of Gaia. It consists of two low-resolution fused-silica prisms dispersing all the light entering the field-of-view in the along-scan direction. One disperser — called BP for Blue Photometre — operates in the wavelength range 330–680 nm; the other — called RP for Red Photometre — covers the wavelength range 640–1050 nm. Both prisms have broad-band filter coatings for blocking light outside the bands defined above. The dispersion of the prisms ranges from 3 to 29 nm/pixel for BP and from 7 to 15 nm/pixel for RP. The prisms are located between the last telescope mirror and the focal plane. These simultaneous semi-photometric measurements of the spectral energy distribution yield key astrophysical information, such as the type of object (star, QSO, solar system object, ...) and its parameterisation (temperature, gravity, chemical composition, and absorption in case of stars; redshift in case of QSOs and so on).

An example of the expected signal that will reach the detector is shown in Fig. 9. The simulated spectra shown in this figure include realistic dispersion curves, response curves, and PSFs as a function of wavelength (for details see [Bro06b] and section 5.2.1), but the effects of measurement noise are not included.

An example of the data that can be expected from the BP and RP photometers at a single observation is shown in Fig. 10. Here realistic noise has been included, though the background is removed. In general only one-dimensional spectra are measured due to the binning in the across-scan direction. The windows around the dispersed image that are binned and read out are shown in Fig. 9. They consist of 60 samples along the spectra, each sample consisting of 12 binned pixels, which includes sufficient samples for measuring the sky background. The BP and RP spectra will be used to estimate the colours of the sources and to derive their astrophysical parameters. In interpreting this data it is important to note that the spatial extent of the PSF combined with the dispersion curve causes the light from each wavelength to be smeared out over several pixels in the spectrum. This causes a decrease of the 'spectral purity' of the photometric data in the sense that particular photometric bands which are useful for the derivation of specific astrophysical parameters will be contaminated by light from nearby wavelengths outside the band. The number of independent 'photometric bands' is approximately 18 [Bro06b].

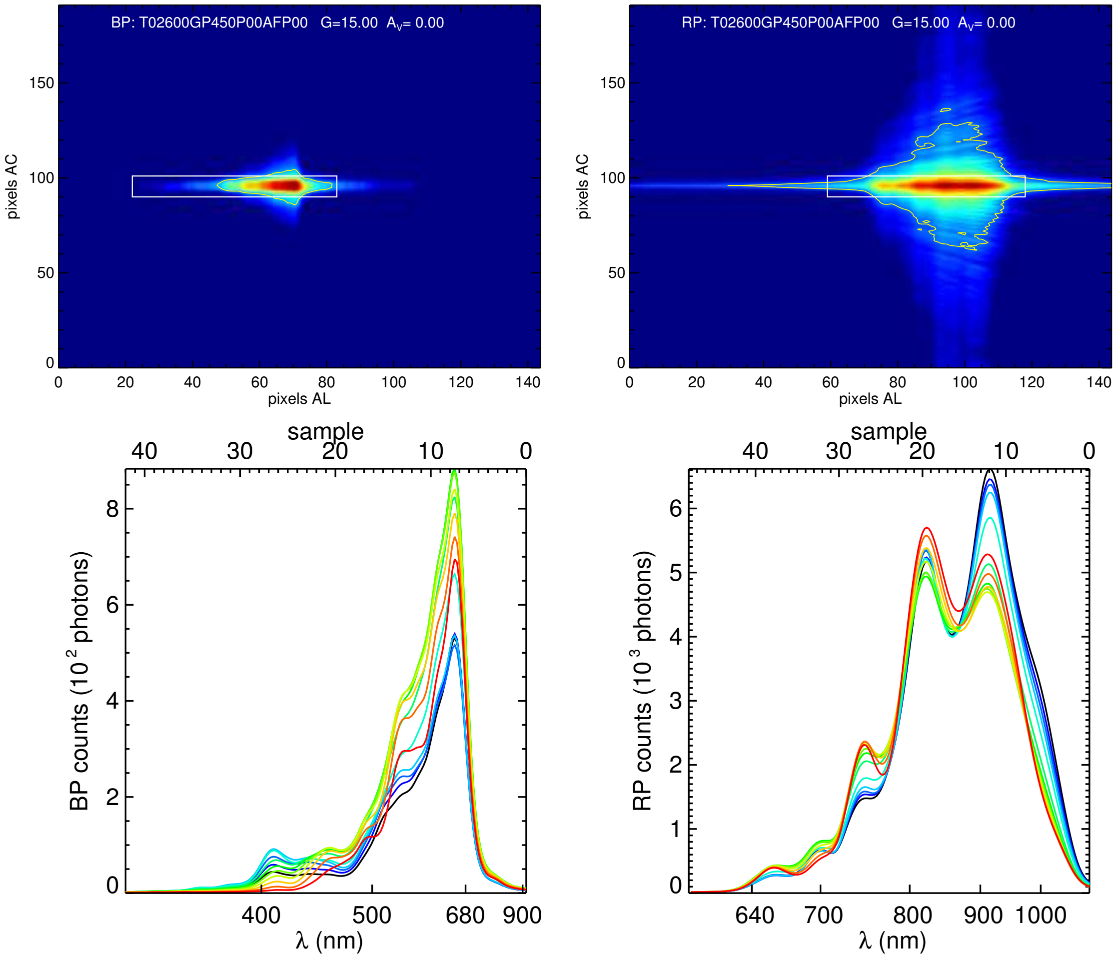


Figure 9: Top panels: simulated dispersed BP (left) and RP (right) images for a $G = 15$ M6V star. The colour scale shows the photon counts in each pixel. The yellow contours show where the sky background level starts and the white rectangles are the windows around the dispersed images that will be read out and transmitted to ground. The flux inside the rectangles is summed in the vertical direction to obtain a 1D spectrum. Bottom panels: the 1D BP (left) and RP (right) spectra of the same star that will be available on the ground for data processing. The different coloured lines show the spectra for different values of $\log g$ increasing from -1.0 to $+5.0$ (-1.0 is black and then $\log g$ increases from blue through cyan, green, and yellow, to red).

At the end of mission the ~ 70 observations of each source will be added, permitting a factor of ~ 4 in oversampling in wavelength and a reduction in the noise, approaching the ideal signal shown in Fig. 10. This procedure of course required that an accurate wavelength and flux calibration are achieved. The expected end-of-mission performances on the photometry are summarized in section 2.3.

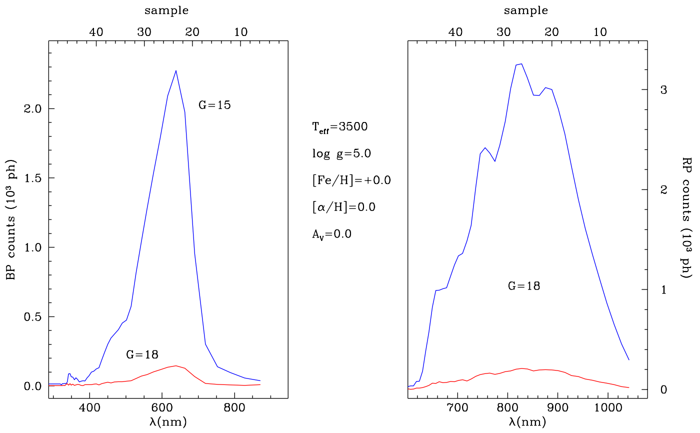


Figure 10: Simulated single-epoch BP and RP spectra for a M2V star for magnitudes $G = 15$ and 18. The vertical scale is in photon counts per pixel. Noise contributions include Poisson noise on the source and sky counts, CCD read-out noise, dark current, plus the overall video chain noise. The total detection noise is assumed to be 5.7 e^-, and calibration errors are included with an a priori tentative value of 0.030 mag per transit. An additional error margin of a factor of 1.2 is included. The noise model is taken from [JHBdB06].

3.3 Spectroscopy

3.3.1 Introduction

The acquisition of radial velocities for the largest possible number of Gaia targets constitutes the first goal of the RVS. As the third component of the velocity vector, radial velocities are crucial for studying, without *a priori* assumptions, the kinematical properties of the Galaxy. The radial velocities are also necessary to correct the astrometric measurements of some 10^5 “nearby” stars [AH98] for the effect of perspective acceleration: an apparent astrometric displacement with a quadratic time dependence induced by the motion of the source along the line of sight. Finally, multi-epoch radial velocities will be very useful for spotting transient phenomena and for detecting and characterizing multiple systems [SÖ5] and variable stars.

The RVS wavelength range contains a mix of strong and weak lines from different chemical species. This variety of spectral transitions will provide several stellar and interstellar parameters in addition to the radial velocities: rotational velocities, atmospheric parameters[5] (effective temperature, surface gravity, metallicity), individ-

---

[5] The atmospheric parameters will be derived using jointly the astrometric, photometric and spectroscopic information.

ual abundances of several key tracers of the chemical history of the Galaxy (e.g. Si, Mg) and diagnostics of 'peculiar' stellar behaviour such as mass loss or stellar activity. The RVS domain also contains a Diffuse Interstellar Band (DIB) located at 862 nm, which, unlike other known DIBs, appears to be a reliable tracer of interstellar reddening [Mun99, Mun00]. The DIB will complement the photometric observations for the derivation of the map of the Galactic interstellar reddening.

3.3.2 RVS design

The RVS is a near infrared ([847, 874] nm), medium resolution spectrograph: $10\,500 < R = \lambda/\Delta\lambda < 12\,500$. It is illuminated by the same two telescopes as the astrometric and photometric instruments. It is a slitless and fiberless instrument, dispersing all the light entering its 0.22×0.39 deg^2 field of view. The RVS focal plane is located in the same plane (at the edge) as the astro-photometric focal plane and therefore also repeatedly scans the celestial sphere. During the 5 years of the mission the RVS will observe a source about 40 times on average[6]. The RVS optics are dioptric and the dispersive element is a grating. It is made of 3 (AL) × 4 (AC) CCDs operated in Time Delay Integration (TDI) mode.

3.3.3 Spectra

In late type stars, the ionized Calcium triplet (849.80, 854.21, 866.21 nm) is the dominant feature in the RVS wavelength range ([847, 874] nm). The intensity of the Calcium triplet decreases with surface gravity, but remains very strong in dwarf stars. RVS spectra also contain many weak, unblended (or moderately blended) lines of different chemical species, in particular of some alpha elements (e.g. Si I, Mg I). The spectra of the coolest stars exhibit no strong molecular bandheads, but many molecular transitions of CN and TiO are visible. The spectra of early type stars are dominated by lines of the end of the Hydrogen Paschen series. The intensities of these lines decrease with increasing gravity. Other lines are also present in the spectra of hot stars, including Ca II, N I, He I and He II. Figure 11 shows an example of a synthetic RVS-like spectrum of a solar metallicity G0V star ($T_{\rm eff} = 5950$K, $\log g = 4.5$, $[{\rm Fe/H}] = 0.0$) without noise. Figure 12 shows a Gibis simulation of a single CCD observation for the same spectral type and an apparent magnitude $V = 10$.

3.3.4 Sampling strategy

The RVS will be operated in windowed mode, as are the other Gaia instruments. The windows are 1104 pixels long (AL) by 10 pixels wide (AC). The length of the

---

[6] assuming 20% dead-time during the mission.

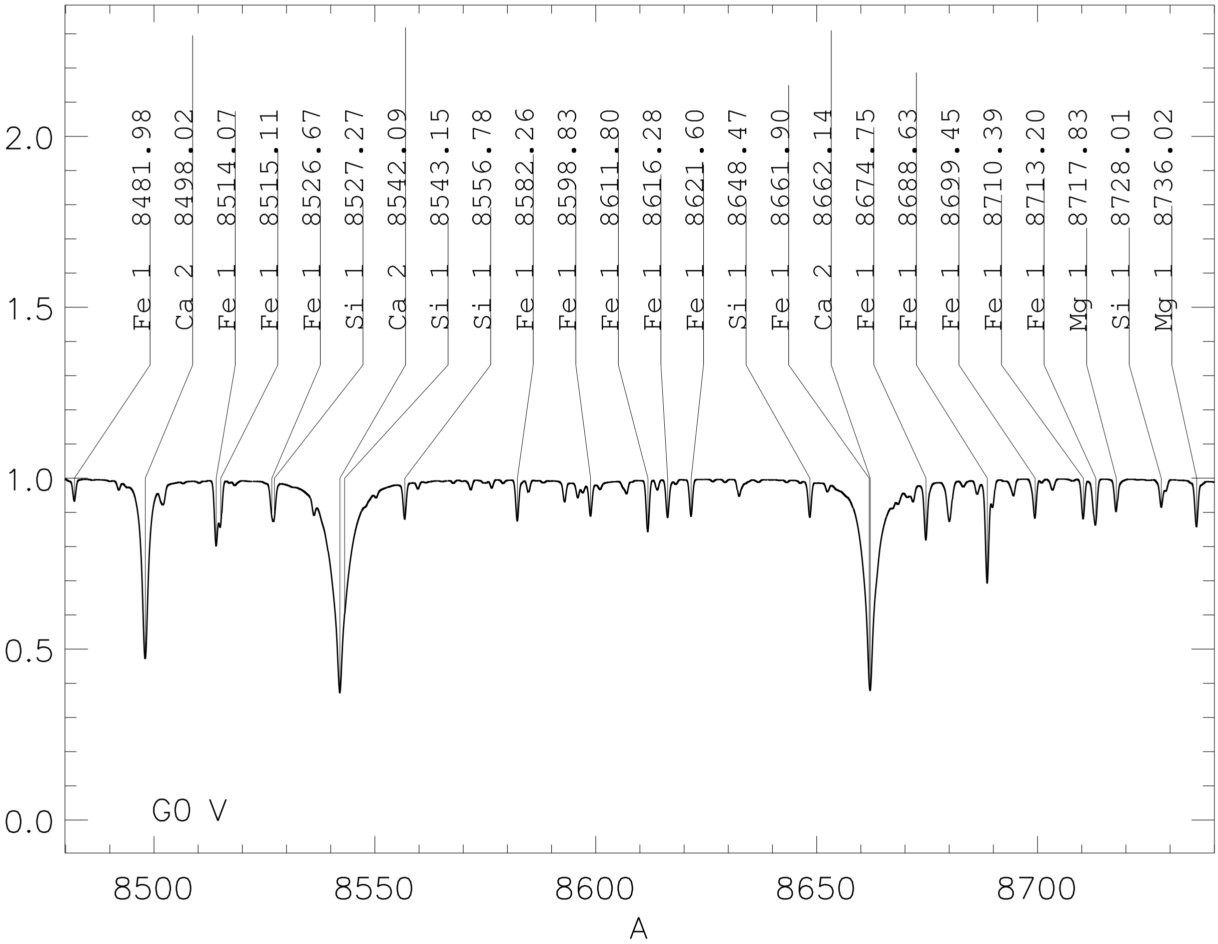


Figure 11: Synthetic RVS-like spectrum of a solar metallicity G0V star

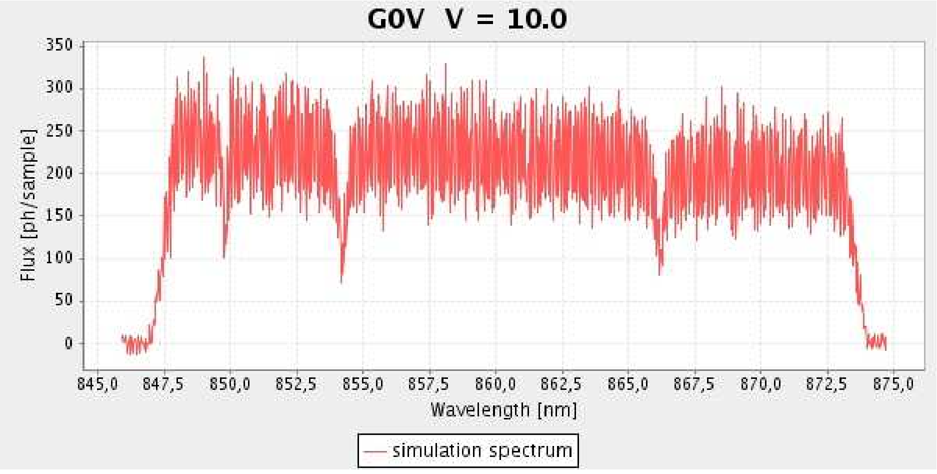


Figure 12: Gibis simulation of a single CCD observation of a V=10 G0V star

windows includes pixels out of the "filter bandwidth" in order to measure the background. The spectra will be sampled with three different modes:

1. The brightest stars, $4.75 \leq G_{RVS} \leq 7$, will be recorded with samples of 1 × 1 pixel. The corresponding spectra will therefore be made of 1104 × 10 samples (AL×AC). In addition, for calibration purposes a small fraction of the fainter stars (over the whole RVS magnitude range) will be observed with this same sampling. These stars are referred to as Calibration Faint Stars (CFS). The CFS will be randomly selected by the on-board algorithms. The CFS status will not be given to a star for the whole mission, but only on a transit-by-transit basis. A star selected as CFS at a given transit will have the same probability (but not more) as its neighbours to be selected as CFS at a subsequent transit.
2. The spectra of the stars in the magnitude range $7 \leq G_{RVS} \leq 10$ (except CFS) will be recorded with samples of 1 (AL) × 10 (AC) pixels. The corresponding spectra will be made of 1104 × 1 samples.
3. The spectra of the stars fainter than $G_{RVS} = 10$ (except CFS) will be recorded with samples of 3 (AL) × 10 (AC) pixels. Larger samples (more binning) are used for the "faint" stars in order to reduce the total read-out noise as well as the telemetry flux. The corresponding spectra will be made of 368 × 1 samples.

The average AL spectral width of a pixel is 0.26 Å. Therefore, stars brighter than $G_{RVS} \simeq 10$ will be slightly oversampled, while the stars fainter than this limit will be under-sampled by a factor $\sim$2.

The RVS is an integral field spectrograph. As a consequence, the spectra of neighbouring stars (with similar across scan positions) will overlap, generating a conflict between overlapping windows. In the case of conflict between two windows, the window of the star with the highest priority (usually the brightest one) will be recorded as if there were no conflict (keeping its rectangular shape), while the star of lower priority will be truncated according to the degree of overlap between the two spectra, resulting, most of the time, in an L-shape window. In case of conflict between more than two stars, the shape of the windows of the stars of lower priorities can be more complex.

3.4 Orbit, tracking performances, ephemeris

The processing of the Gaia observations requires a prior knowledge of the position and velocity vectors of the spacecraft with respect to the barycentre of the solar system. The velocity is required in order to correct the sources' (stars or planets) apparent positions for the aberration. Distances to the solar system barycentre are also needed to determine the stellar parallax with the proper scale, since the parallactic effect increases with the distance to the reference point. Similarly, the computation of the deflection of light by the Sun and the major planets cannot be done without a rather accurate knowledge of the distance and angular position of the planets. Finally, the observations of solar system objects place severe constraints on the knowledge of the positions of Gaia from which we determine its orbital elements. The data processing baseline assumes that these elements will be available from an external source prior to the mission (solar system ephemeris) or during the mission (Gaia final orbit). Therefore they won't be determined as additional calibration parameters fitted to the data, although this option is retained for the spacecraft velocity in the case that the orbit tracking accuracy proves insufficient. In the following it is convenient to consider separately the ephemeris of solar system bodies (major and minor planets and satellites) and that of the Gaia orbit, which is deduced primarily from the active orbit tracking at ESOC.

The accuracy requirements have been discussed in [Kli03a] and set in [Mig05a] taking into account the target astrometric accuracy. This is done such that the remaining uncertainty coming from the computation of the position and velocity of the spacecraft or of the planets has a negligible impact on the accuracy budget. A safety margin of one order of magnitude has been applied in the requirement assessment.

3.4.1 Solar system ephemeris

The astrometric data processing will be carried out in the BCRS (Barycentric Celestial Reference System, Sect. 5.1.1) with the origin at the barycentre of the solar system and reference directions provided by the ICRS (International Celestial Reference System). Regarding the solar system ephemeris the requirements are :

1. Velocity of the Earth in the BCRS to 2.5 $\mathrm{mm\,s^{-1}}$ for each component (1-σ error) and no systematic over the mission length larger than 1 $\mathrm{mm\,s^{-1}}$,
2. Position of the Earth in the BCRS to 0.15 km over each component(1-σ error).

and it is highly desirable to have the position of the other planets (except Mercury) in the $\sim$ 2–5 km range. One must bear in mind that the requirements express what is needed to carry out the data processing without accuracy loss, but they tell us nothing about the extent of the degradation if they are not met. The damage is not the same if one cannot compute an accurate aberration of every star or if one has to drop a handful of observations in the vicinity of Jupiter. The evaluation of the light deflection by the giant planets applies only to stars observed in the immediate vicinity of the planets (Jupiter and Saturn) and the requirements have been assessed for the most demanding cases of the bright stars. At worst, these observations will be rejected if the requirements are not met, and only a small number of observations (although very valuable for specific applications) would be lost. On the other hand, the light deflection by the Sun is computed for every source at any observable angle from the Sun (that is to say always larger than 45 degrees). Fortunately the relevant requirement in position ($\sim$ 3000 km) is easily met.

3.4.1.1 Source of Ephemeris

There are at the moment two main world centres in a position to provide high quality ephemerides for the solar system: the NASA/JPL in the United States and the IMCCE (Institut de Mécanique Céleste et de Calcul des Ephémérides) in Europe. The planetary solutions belong to two groups: (i) numerical solutions, based on direct numerical integration of equations of motion; (ii) analytical solutions constructed from expansions of the solutions into Fourier and Poisson series. In the first class we find the reference solutions (DExxx, LExxx) provided by the JPL for the planets and the Moon, while the Paris group has a long tradition of analytical solution, like VSOP (Variations Séculaires des Orbites Planétaires)[Bre82, PB88] or TOP (Théorie du mouvement des quatre grosses planètes)[Sim83]. The most recent versions of these ephemeris are referred to the BCRS and the next version of VSOP will use TCB as the time independent parameter. Without entering into the details, the two sources have comparable performances and represent the state of the art in this field. Either could be customized to meet the Gaia needs. While the JPL versions are automatically distributed in the form of Chebyshev polynomials, the analytical theories of the IMCCE appear as long tables of trigonometric terms with amplitudes, phases and arguments. A computer efficient access interface must be constructed for an intensive use.

Given this near equivalence, it was decided by the Relativity and Reference Frame Working Group during the study phase to favour a European source and to ask the IMCCE to be the provider of the primary source of solar system ephemeris for Gaia. This decision has been subsequently agreed by the Institute and this task integrated in the DPAC activities. It happens that in parallel the same group had started the development of a fully new generation of ephemeris, named INPOP an acronym for *Intégrateur Numérique Planétaire de l'Observatoire de Paris*, with higher accuracy, fitted to a new set of observations and with TCB as an independent time argument [FLS+05]. As a result of this timely development, the DPAC baseline is now to implement this new solution as the main source of the solar system ephemeris for Gaia. In case some unexpected difficulties could delay the finalisation of INPOP, the last issue of VSOP will be readily available as a backup.

3.4.1.2 Implementation

The calls to the ephemeris are extremely frequent in the data processing and could be very demanding in terms of computing resources. Thus some form of pre-computed ephemeris is mandatory, reducing the final computation to a mere interpolation. For this purpose the numerical representation and the access has been optimised and

customised for the data processing needs. This has been prepared with a mock-up version used in the simulation and in the GDAAS prototype [Mig03a], [Mig03b]. Starting from the analytical development of the theories it was possible to generate Chebyshev representations by selecting the degree and the period of validity of each block so that the numerical truncation was compatible with the accuracy requirements. The data files for the major planets (Venus to Neptune) and Gaia have all the same setting and a call generates for any time the position and velocity vector of the bodies.
For consistency reasons it is important that the same ephemeris is used everywhere in the data processing. To this end the ephemeris is part of the common data accessible in the Parameter Database and dedicated Java classes have been written in the Gaia Toolbox which provide easy access to everybody. The version used during the preparation and for the simulations will be replaced by new files based on the official ephemeris delivery of the IMCCE. This delivery will take place in two steps: early 2007 for a test version and 2009 for the final version.

3.4.2 Ephemeris of minor planets

The ephemeris for minor planets has a very different nature: There are many more objects – several hundred thousands – but the accuracy needs are low (around one arcsecond). Accesses will be extremely frequent during the Initial Data Treatment (IDT) (Sect. 4.1) in order to decide whether a source recently observed is a known solar system object rather than a stellar source. Although the number of sources is large, it remains much smaller than the number of Gaia targets. Thus searching for a positive identification for each observations will result in a positive match in only one in every 3000 accesses ($3000 = 10^9/3 \times 10^5$). A search algorithm is being developed to retrieve efficiently all the known planets located in the areas of the sky scanned by Gaia's FOVs at any time. The main input will come from a list of orbital elements of the minor planets, updated during the mission by the Gaia discovery of new planets. The computation will be organised to minimize the retrieval time during the IDT.

3.4.3 Orbit tracking

The very accurate position and velocity vectors of Gaia in the BCRS are mandatory for the Gaia data processing. Unlike the solar system ephemeris this must be obtained from real time observations of the spacecraft. The three main requirements are given in [Mig05a] as follows:

1. Random error (1-σ for each Cartesian component to compute the along-track aberration) on the velocity of Gaia from the tracking to within $2.5\,\mathrm{mm\,s^{-1}}$,

2. No systematic error over the mission length on the velocity of Gaia from the tracking larger than $1\,\mathrm{mm\,s^{-1}}$,

3. Position of Gaia from the tracking to within 0.15 km (1-σ error for each Cartesian component)

The tracking capabilities have been analysed at ESOC and are reported in [Hec06]. The baseline tracking solution (one station, one range point per pass) as it is available at the time of writing does not meet the requirements, marginally for the velocity and significantly for the position. This assessment is based with the following assumptions:

- Once the operational orbit is reached, tracking exclusively from one ground station (New Norcia assumed).
- The duration of the tracking pass is restricted to 8 hours maximum.
- Doppler measurement frequency: one every 10 minutes.
- Two Range measurements per tracking pass at the beginning and at the end, cases with one point at the beginning and with Doppler only are included.
- Doppler noise 0.03 mm s^{-1} (1-σ) above 15 deg elevation.
- Range Bias 20 m (1-σ).
- Range noise 20 m (1-σ) above 15 deg elevation.

The reference case yields $\sigma_v = 8$ mm s^{-1} and $\sigma_p = 8$ km. For both values there are specific periods during which the errors are particularly large. The average 1-σ error in velocity is about 4 mm s^{-1}, not dramatically far from the requirements. For the position this is still 4 km, significantly above the requirements. There are still several options not included in the reference solution, like having two stations with Doppler and above all the use of Delta Differential One-way Range (ΔDOR) measurements, which would fill the requirements at any time. Finally, and specifically for the position, regular angular astrometric observations of the spacecraft would improve considerably the positional and velocity accuracy. With significant additional complexity, the data processing could also add some general parameters to produce a smooth representation of the velocity to improve on the tracking data. Although the situation is not yet fully clarified, the preparation of the data processing assumes that the Gaia position and velocity vectors will be known in the BCRS/ICRF with an accuracy very close to the requirements.

3.4.4 Lissajous orbit

Following the technical and scientific requirements of the Gaia mission the orbit of the satellite has been chosen to be a Lissajous orbit around the Lagrange point L_2 of

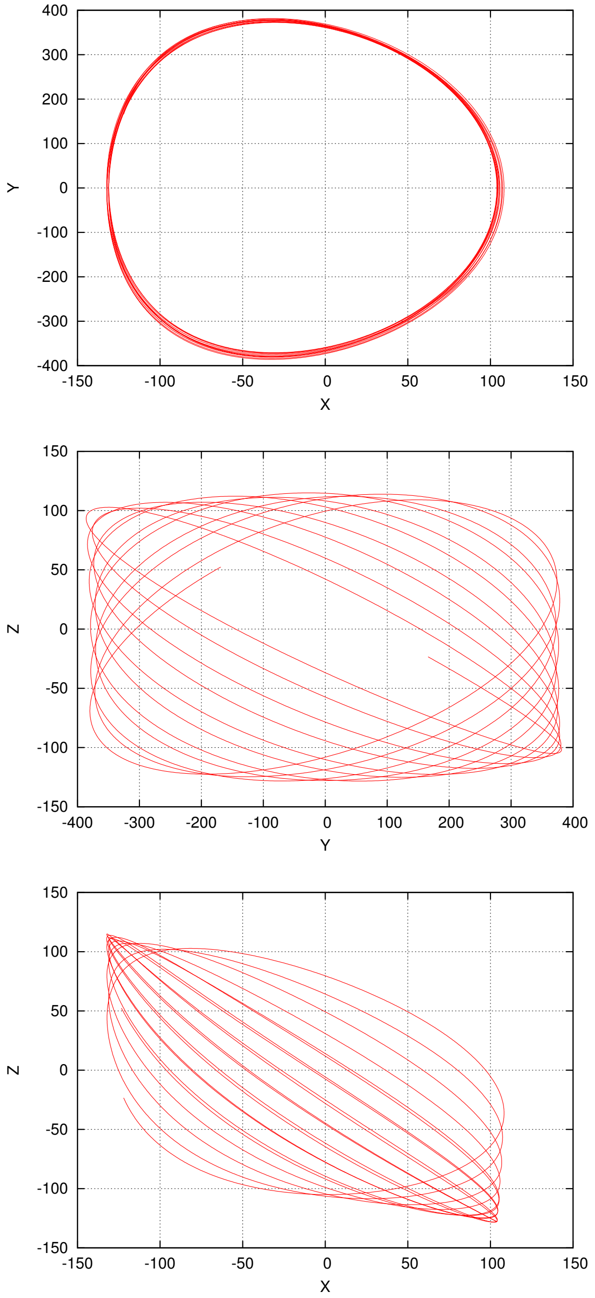


Figure 13: Example of numerically constructed realistic Lissajous orbit with the size close to that recommended for Gaia (see Section 3.4.4 for details). The coordinates, in units of thousand of kilometers from L_2, are shown in the co-rotating coordinates: the X-axis is directed from the Earth-Moon barycentre to L_2, the Y-axis lies within the plane spanned by the X-axis and the BCRS velocity of L_2, and the Z-axis complements the triad to the right-hand one.

the Sun–Earth system [Hec06]. Although the high-accuracy calculations related to the orbital dynamics (launch and transfer orbits, orbit and attitude control etc.) will be done by ESOC, it seemed to be advantageous (for the purposes of simulations, for example) that the Gaia scientific community have some capability in computing realistic Lissajous orbits around L_2. To this end two studies were carried out. First, the first-order (linearised) theory of Lissajous orbits were summarized in [Mig03b] and a version of Gaia ephemeris (i.eȧ set of Chebyshev polynomials representing the position and velocity of the Gaia satellite with respect to the BCRS) based on this first-order framework has been prepared. This first-order orbit shows many of the quantitative characteristics of the real Lissajous orbit, but is dynamically incompatible with the ephemeris used for the rest of the Solar system. In a second study, a fully realistic Lissajous orbit has been calculated based on the numerical integration of relevant equations of motion with the positions and velocities of the Sun and planets taken from a given ephemeris [Kli05]. The procedure is a simpler version of that adopted in ESOC [Hec06] and consists of a numerical bisection of the velocity in the so-called escape direction from the initial guess given by the linearised theory.

Figure 13 gives an example of a realistic numerically constructed Lissajous orbit around L_2. The size of the orbit is close to that adopted for Gaia. The orbit is integrated for ten subsequent segments of 200 days (a total of 2000 days). Between the segments the velocity was adjusted so that the Lissajous orbit remains quasi-stable. The relatively large segment length of 200 days is possible here because of the simple dynamical model: the post-Newtonian equations of motion without solar pressure and without noise. In the upper plot showing the dependence of Y on X the non-linear character of the orbit can be seen easily: a linearised orbit would be a pure ellipse in this projection. This new orbit will be combined with the ephemeris of the Earth and implemented at the same time as the first delivery of the solar system ephemeris by IMCCE. This orbit will then be the baseline for the simulations and any other work using the Gaia orbit in DPAC. The Chebyshev version will be accessible through the GaiaTools.

3.5 Scanning

Gaia will perform its observations from a controlled, Lissajous-type orbit around the L2 Lagrange point of the Sun and Earth-Moon system. During its 5-year operational lifetime, the satellite will continuously spin around its axis, with a constant speed of 60 arcsec s^{-1} (Fig. 14). As a result, over a period of 6 hours (the spin period), the two fields of view will scan across all objects located along the great circle ‘perpendicular to’ the spin axis. As a result of the basic angle of 106.5° separating the fields of view on the sky, objects transit the second field of view with a time delay of 106.5 minutes compared to the first field of view.

Gaia’s spin axis does not point to a fixed direction in space but is carefully controlled

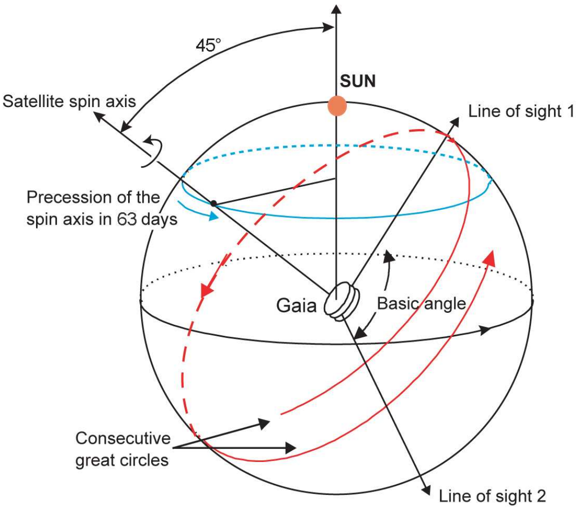


Figure 14: Gaia's two fields-of-view scan the sky according to a carefully prescribed, 'revolving scanning law'. The constant spin rate of 60 arcsec s^{-1} corresponds to 6-hour great-circle scans. The angle between the slowly precessing spin axis and the Sun is maintained at 45°. The basic angle is 106.5°.

so as to precess slowly on the sky. As a result, the great circle that is mapped out by the two fields of view every 6 hours changes slowly with time, allowing repeated full sky coverage over the mission lifetime.

The 'scanning law' prescribes how the satellite's spin axis evolves with time during the mission. The optimum scanning law (i) maximizes the angle ξ between the Sun and the spin axis (the Solar-aspect angle) at all times and (ii) maximizes the uniformity of the sky coverage after 5 years of operation. The first requirement results from the fact that the parallactic displacement of transiting stars is proportional to $\sin \xi$; a higher value of ξ thus leads to larger measurable parallaxes and higher end-of-mission astrometric accuracies. Thermal stability and power requirements, however, limit ξ to about 45°. The best strategy is thus to let the spin axis precess around the solar direction with a fixed angle of 45°. This combination of a spinning satellite, scanning the sky along great circles, and a precession of the spin axis is referred to as 'revolving scanning', and was used for the Hipparcos mission. The actual speed of precession of the spin axis on the sky should be small enough that consecutive great-circle scans overlap sufficiently, and large enough that all stars on the sky transit the fields sufficiently often.

The above requirements have been worked out in detail for Gaia, leading to an optimum nominal scanning law. For a spin rate of 60 arcsec s^{-1} and a solar aspect angle of 45°, the precession speed is such that 5 years of operation corresponds to 29 revolutions of the spin axis around the solar direction; the precessional period thus equals 63 days. On average, each object on the sky is observed about 70 times in the astrometric and photometric fields and 40 times in the spectroscopic field (two

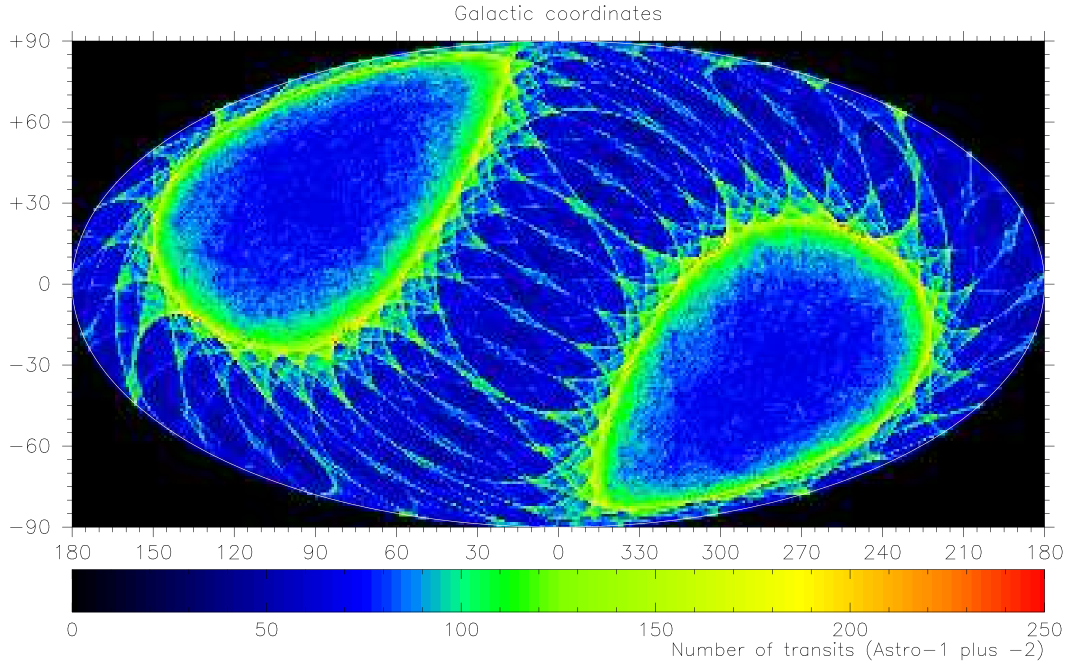


Figure 15: During its operational lifetime, Gaia will continuously scan the sky, roughly along great circles, according to a carefully selected, pre-defined scanning law. The characteristics of this law, combined with the across-scan dimension of the fields of view, result in the above pattern for the distribution of the predicted number of astrometric and photometric transits on the sky in galactic coordinates. The distribution of spectroscopic transits is identical, but with absolute numbers smaller by a factor 4/7, following the 4 CCD rows used for the RVS instrument compared to 7 CCD rows used for astrometry and photometry.

viewing directions combined and 20% total dead time assumed in both cases). Direction dependencies in galactic coordinates are illustrated in Fig. 15.

The on-board object detection and selection logic cannot cope with arbitrarily dense fields. In areas of high star densities (exceeding about a million objects per square degree), a dedicated scanning law is therefore optionally available. This mode, referred to as the modified scanning law, consists of applying a reduced precession speed in combination with a random selection of faint stars to be observed. Temporary activation of this mode, when encountering for example Baade's Window, will result in an increased number of passes over the same region of sky and will thus alleviate the payload-induced density restrictions. The precise operational strategy of this mode is currently under definition.

3.6 Telemetry flow

3.6.1 Telemetry organisation and flow

A brief description of telemetry structure and contents, organisation, and the envisaged onboard-to-ground transmission scheme is given in the following. It must be noted that the detailed telemetry system design is an ongoing industrial activity in the current phase B2 and, consequently, some of the presented details may still change by the time the implementation phase C commences in April 2007.

Gaia's telemetry falls into three basic categories

- Periodic and non-periodic Service Module (SVM) and Payload Module (PLM) housekeeping data
- Attitude data from the Attitude and Orbit Control Subsystem (AOCS)
- Observation data

The housekeeping data comprises engineering quantities that define the status of all essential satellite and payload systems. Examples are voltages of electrical components, temperature sensor readings, counters and diagnostic information from the CDMU/PDHU, and many more. Typical sampling rates for most the Housekeeping (HK) parameters should be around 1 Hz which will result in a low-rate data stream not exceeding a few kbit/s. It is expected that the large majority of HK parameters will be irrelevant for the data processing. Nonetheless, a few selected ones (e.g. the parameter describing the variations of the basic angle between the two telescope viewing directions) will clearly be of relevance for the data reduction. The detailed content and structure of the HK parameters have not been established. By its nature and similarity to other missions this is believed to be a standard, uncritical engineering task.

Attitude data forms a second stream of information of constant low rate of a few kbit/s. Unlike HK parameters, however, attitude data is of paramount importance for the astrometric core processing (Sect. 5.1). As the system design of the AOCS is still ongoing the detailed structure and content of the onboard attitude data are unknown at present. In this situation the data simulation efforts (Sect. 6) assume the availability of a quaternion representation at a constant rate (e.g. 1 Hz). This assumption appears valid: In case Gaia's AOCS system is not capable of generating quaternions directly, a corresponding on-ground process could do this instead.

The bulk of Gaia's telemetry (see Sect. 3.6.2) will consist of raw CCD samples representing the elementary observation data acquired by the astrometric (Sect. 3.1), photometric (Sect. 3.2), and spectroscopic (Sect. 3.3) payload modules. Each stellar object that transits the focal plane will undergo a sequence of detection in SM1/SM2, confirmation in AF1, and astrometric, photometric, and spectroscopic measurements

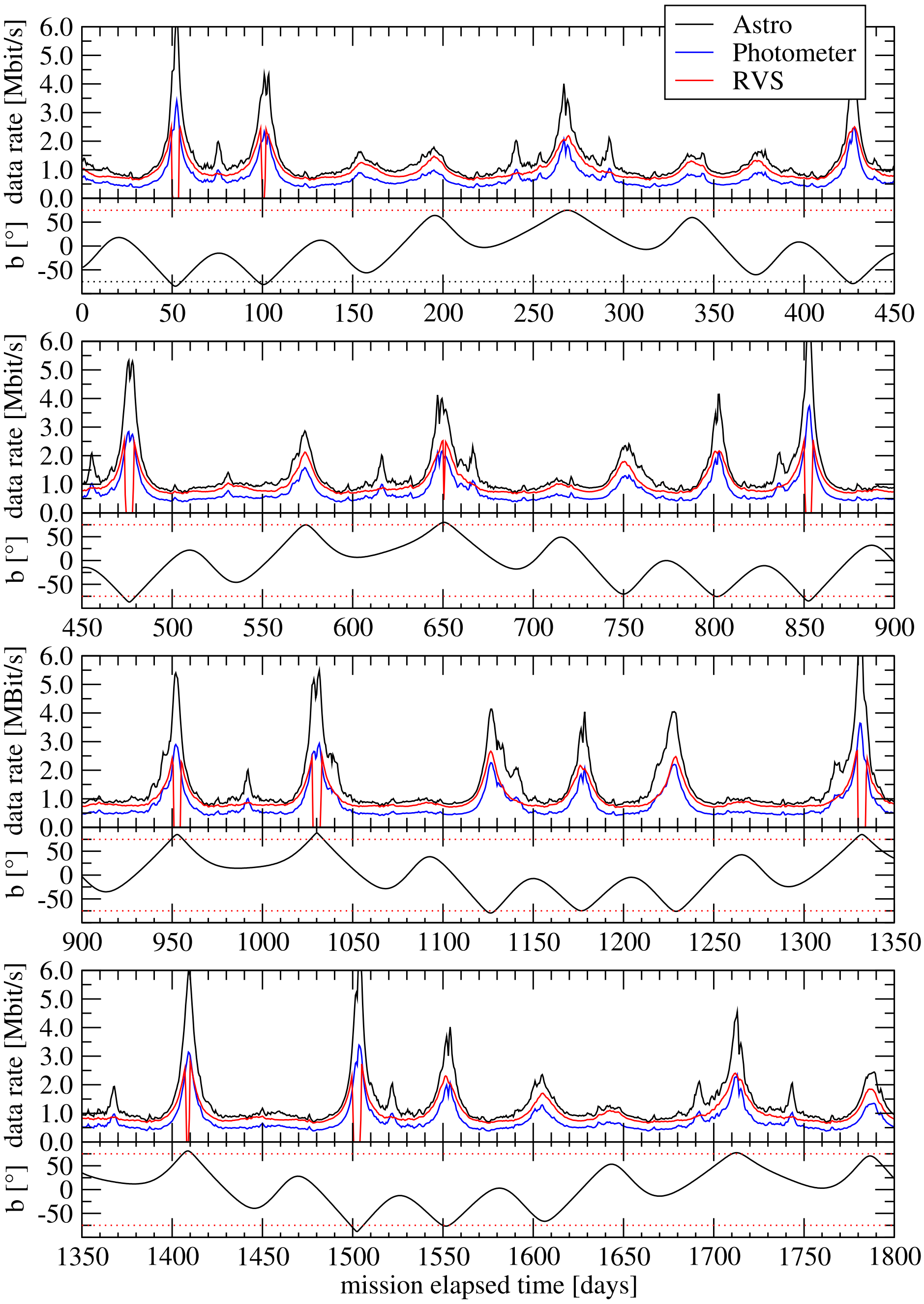


Figure 16: Mean data rates from Astro, Photometers, and RVS (for one particular RVS data model and operating scenario - see [Lam05] for details) over the 5 year nominal mission duration. The plots beneath each of the four partial rate curve graphs show the galactic latitude of the spin axis (+Z of the SRS frame) on the same time axis. The dotted horizontal red lines mark the ±75° limits.

in the subsequent main fields AF1–9, BP+RP and RVS respectively. Transits over the

RVS CCD strips will only occur for a subset of all objects due to the narrower AC height of that instrument.

Each object that was detected and confirmed will be assigned tracking windows by the onboard software. The windows define rectangular areas in CCD pixels around the object's PSF for readout by the CCDs in the remaining transits over strips AF2–9, BP/RP, and RVS. Windows vary in size and binning mode (summing of pixels in AL/AC direction during the readout) depending on the object's apparent magnitude according to a defined sampling scheme [HdB06]. At the end of a transit the read out CCD samples are grouped, compressed, augmented by attributes from the detection process, and finally organised into so-called star packets. Any RVS data from a transit will be separated from the corresponding SM/AF/RP/BP data and put in a separate RVS star packet for that object. Completed star packets will be stored in the Solid State Mass Memory (SSMM) before transmission to the ground. The SSMM is organised and operated as a file system in which the individual star packets fill clusters, pre-allocated data structures of a fixed size. The clusters in turn form linked LIFO queues with assigned downlink priorities. Which queue a star packet will be assigned to depends on its magnitude and possibly a small number of other attributes. Roughly speaking, brighter stars get higher priorities than fainter ones.

During routine observation phases the SSMM is continuously filled with incoming star packets. When contact with the ground station is established the SSMM is gradually emptied by downlinking data in order of decreasing priority. This multi-stage process is roughly composed of the following steps:

- Grouping of clusters with star packets into Telemetry (TM) Source Packets (typical source packet sizes are a few 10 kbit)
- Augmentation of Source Packets with Reed-Solomon check symbols
- Splitting of Source Packets into Transfer Frames (the length of a transfer frame is of the order of 10 kbit)
- Transmission of transfer frames to the ground through the medium-gain antenna at a rate of 5–8 Mbit/s

Received transfer frames on the ground are reorganised into Source Packets for onward ground transmission to MOC/SOC (Sect. 3.6.3).

3.6.2 Telemetry volumes and rates

The total volume of telemetry data that Gaia will generate during its operational lifetime is determined by the following four main factors:

- Object numbers: The maximum sensitivity of the respective payload modules in combination with capacity limitations of different onboard systems limit the maximum number of celestial objects that Gaia will be able to observe. This figure is around 1200 million for the astrometric instrument and the photometers and around 450 million for RVS [Lam05, Fig. 1].

- Scanning: The NSL (and MSL) determine the frequencies and counts with which the 1200 million objects will be observed during the operational phase. The non-uniformity in coverage (Fig. 15) will have a small effect on the total data volume due to the very large stellar density variations across the sky. The initial conditions of the NSL also have an influence.
- Operations: Operational scenarios may include the disabling of payload modules or activations of other telemetry-reducing measures under specific conditions. An example could be the disregarding of faint-star observations in cases of space-ground communication problems, etc. Also, satellite maintenance outages and all other non-nominal periods in which no observations can be carried out will influence the total data volume.
- Telemetry data structures: It is clear that the sampling scheme ([HdB06]) has the largest impact on the total data volume. The amount of data generated per object and transmitted to the ground in a star packet (Sect. 3.6.1) depends on the object's brightness. Approximate star packets sizes [EA06] will be:

Magnitude range	Packet size (uncompressed)	Packet size (compressed)
Astro + Photometers		
G=6–13	51 kbit	25 kbit
G=13–16	5.6 kbit	2.8 kbit
G=16–18	4.4 kbit	2.2 kbit
G=18–20	4.3 kbit	2.0 kbit
RVS		
G=6–7	530 kbit	212 kbit
G=7–10	53 kbit	21 kbit
G=10–17.5	18 kbit	7 kbit

For accurately predicting total volumes, time-resolved data rates and average data rates, the above four factors have to be incorporated into a single, homogeneous modelling framework [Lam04, Lam05, EA06]. This yields the following key results: Table 4 shows per instrument the mean uncompressed data rates, total accumulated end-of-mission data volumes as well as these numbers referred to SM+AF+BP+RP. RVS is quite rate-intensive: The rates and volumes amount to about 60% of what the star mappers, the astrometric field, and the photometers combined generate. The total uncompressed accumulated data volume at the end of the mission is about

Table 4: Mean uncompressed data rates and total data volumes for a 5 year nominal mission

Instrument	mean rate [Mbit/s]	total volume [TB]	rel. to SM+AF+BP+RP
SM+AF+BP+RP	3.1	62	1.0
RVS	1.8	35	0.6
Total	4.9	97	1.6

100 TB. This figure takes into account data loss that has occurred due to SSMM saturation.

The telemetry rates of Astro (Astro-1 and Astro-2 combined) and the Photometers (RP and BP combined) over the nominal mission duration are depicted in Fig. 16. The single curve below each of the four rate graphs shows the galactic latitude of the spin axis ($+Z$ of the SRS frame) in units of degrees. The curve reflects the very large stellar density variations across the sky: Maxima in the data rates are correlated with galactic pole pointing of the spin axis, hence, scans along (or at shallow angles to) the galactic plane where the density of stars is very high.

About 25 galactic plane scans will occur over the five years. In these periods the on-board telemetry volume exceeds the capacity of the daily downlink, thus, the SSMM cannot be fully emptied during a single ground station pass. As a result the SSMM fills gradually up until it gets full and data loss becomes inevitable. In this situation newly acquired data is not simply lost but a priority-driven deletion scheme discards marked data clusters to make room for star packets of higher priority. After memory saturation has occurred it can take up to several tens of days to fully empty the SSMM again. This time depends strongly on the chosen SSMM size, the available downlink rate, and the length of the daily communication period. The latter is given by the duration of visibility of the satellite from the ground station. Fig. 17 shows this for Gaia's prime ground station Cebreros, Spain. It can be seen that the visibility varies between 7.5 and 14 h over the course of a year with minima in the summer when L2 is situated low over horizon. It is hoped that optimum use of the available visibility can be made. However, cost considerations may dictate a maximum constant use of merely 8 h per day. Alternatively, New Norcia, Australia is considered as a backup station to supplement the core coverage of Cebreros in times of galactic plane scans when maximum daily downlink capacity is needed most.

Apart from the ground station visibility period, the raw downlink rate through the medium gain antenna is the most influential parameter that determines how much data can be downlinked and, ultimately, the total end-of-mission data loss. The minimisation of this quantity is clearly desirable in the interest of maximising the science return from Gaia (astrometric accuracies quoted in Sect. 1.2.2 depend on dead times). The task is a complex trade-off between numerous parameters and factors some of which have been mentioned above. The most important ones are:

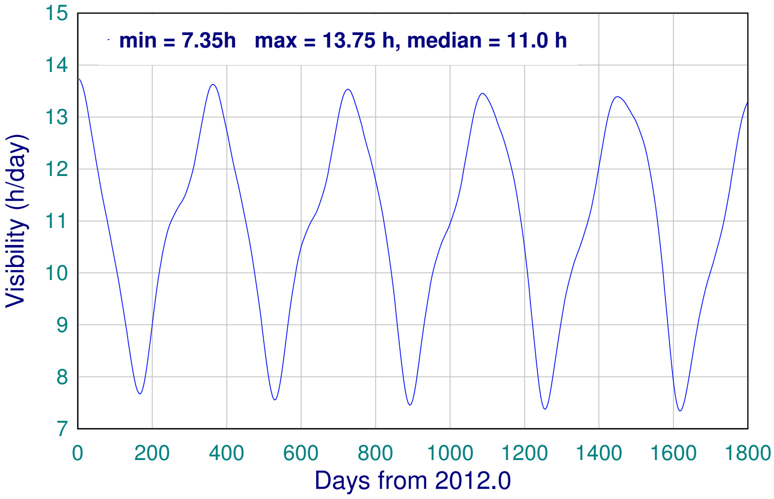


Figure 17: Daily visibility of Gaia from ESA's Cebreros ground station over the duration of the mission

- Effective downlink rate
- Duration of daily space-ground communication period
- SSMM size
- Per-object size as per sampling scheme

This analysis has been performed in previous project phases [Lam04] already but is still ongoing in the light of continuing industrial design optimisations and the start of more detailed ground segment planning and development activities.

3.6.3 Telemetry transmission to SOC

The DPAC interface for scientific data is the Science Operations Centre (SOC) located at the European Space Astronomy Centre (ESAC) near Madrid in Spain. The DPAC acknowledges that the transmission of telemetry from MOC to SOC or ground station to SOC is an internal ESA issue.

The MOC–SOC Interface Requirements Document covers all aspects of the transmission of data from MOC to SOC [Hoa07]. Timely delivery of the data needed for FL, essential for instrument health monitoring, should be guaranteed by the priority scheme [B*et al*06] .

Part II

Gaia Data processing

4 Preliminary processing

This section presents the organisation of the data reception, initial processing, ingestion and initial quality control.

4.1 Initial Data Treatment

As new data arrives from the spacecraft, it must be unpacked, organised, and processed to a sufficient level that it can enter the astrometric, the photometric, and the spectroscopic processing cycles. In particular image parameters must be extracted from the astrometric observations, the satellite attitude reconstructed, and all transits matched against the catalogue of stellar sources and the solar system ephemerides. This is the task of the Initial Data Treatment (IDT).

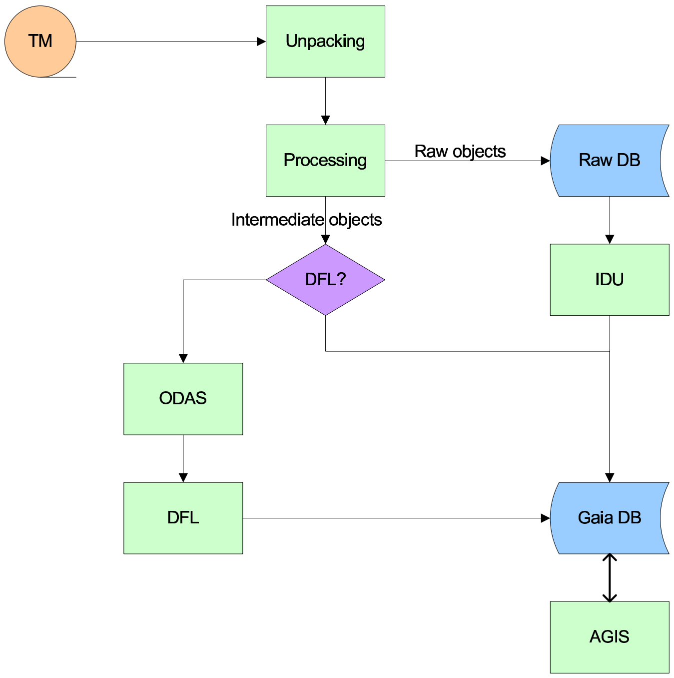


Figure 18: Overview of the IDT and related processes. Unpacked telemetry is processed, the observations stored in the Raw DB, while the image parameters go to the Gaia DB after possibly having entered the First Look processes. From time to time the observations are reprocessed in IDU

The IDT must run daily, partly to keep up with the data volume, and partly to allow the earliest possible health checks on the spacecraft operations. These checks are carried out by the First Look as described in Sect. 4.2.

In Fig. 18 we show the major components of the IDT, and the processes closely related. Unpacked telemetry is processed, and the observations stored in a data base of *raw objects*, while the extracted parameters (the *intermediate objects*) go to the Gaia DB, from where the astrometric data will enter AGIS. The first time we process data from a particular day, the intermediate data will also pass through the First Look processes ODAS and DFL. At intervals of six months, the processing of the observations is repeated in the Intermediate Data Updating (IDU) as better calibration parameters become available. The details of the processing is different for the astrometric, the photometric, and the spectroscopic observations, as described below.

4.1.1 Data reception and unpacking

The telemetry arrives daily at SOC (cf. Sect. 3.6.3), where the IDT, FL, and also AGIS will run, and from there data will be exchanged with the other DPCs. Due to variations in the star density on the sky, it will not always be possible to empty the on board storage during a visibility period, and data may arrive with a delay of several days. There is no way we will know how much data may arrive later, so we will not wait for it to arrive, but simply process all available data on a daily basis.

What is important for the IDT is that housekeeping data has a very high priority, and that a priority scheme for the stellar data gives higher priority to the brighter sources. We will therefore always have sufficient data for a smooth IDT run, almost up to the time of the last telemetry package. This run must include the on board attitude, and enough observations to establish a first, and accurate, on ground attitude.

Apart from the scientific telemetry, also housekeeping data, control commands, time correlation data, and orbit data will be received at SOC, and will be needed in the IDT.

An observed transit is split on board in two different telemetry files. The first part contains the transit across the SM, AF, BP, and RP CCDs, while the other part, only observed for a small fraction of the sources, consists of the transit across the RVS CCDs. Both parts are labelled with the same *Transit Identifier*, but will arrive independent of one another on ground.

An unpacking job makes the formal checks on the telemetry packages, and runs the decompression. In order to facilitate the distribution of the processing later on, the decompression process will store the data in different files according to data type, priority, and observation time. The volume of uncompressed data is on average 65 GB daily, of which 42 GB is from RVS, cf. Tab. 4.

4.1.2 Housekeeping data and auxiliary data

The housekeeping data (Sect. 3.6.1) will, in general, not be processed in the IDT, but simply stored in the data base for future reference. Other kinds of auxiliary data, like the control commands issued from ground, the reconstructed orbit of Gaia, and the time correlation data must also be stored. Daily processes are, however, needed for the timing data and for the attitude. These processes must run before the treatment of the scientific data.

The time correlation data are processed in order to monitor the on board clock, and in order to establish the transformation to TCB, to which all observations must be referred. The on board attitude determinations form our starting point for the attitude reconstruction. The individual determinations are of modest accuracy (several arcseconds), but with a simple filtering we can at least reach a level of one arcsecond. The filtered attitude, including the error estimate for each value, is added to the IDT attitude file.

4.1.3 A first on ground attitude reconstruction

An accurate knowledge of the attitude is needed in IDT itself for a reliable cross matching and for assigning a spatial index[7] to each transit, and it is needed as a starting point in AGIS. We assume that the filtered on board attitude is not sufficient, and will therefore set up a simple on ground attitude reconstruction, which must run once for any time interval.

We use the astrometric observations of brighter stars (G 8–16, say), which are relatively few and fast to process. In dense areas, we may limit the number of stars per time interval for each field of view, to obtain a balance in the weights for both directions. After the normal image parameter extraction, they are cross matched with the source catalogue (or an Attitude Star Catalogue subset of the full catalogue) using the initial attitude. Finally we run an attitude determination on this set of observations, and replace the initial values in the IDT attitude file.

After these preparatory steps have been completed, the processing of the large volumes of observations can begin.

4.1.4 Processing of spectrometric data

The RVS processing in IDT deals with a large data rate, some 42 GB/day of telemetry, but will do little actual data analysis. A subset, of both bright and faint sources, will be selected for a simple treatment, to allow First Look to monitor the conditions and operations of the RVS instrument. For further spectrometric processing, IDT must provide a cross matching, spatial indices for the two fields of view, and a flux

---

[7] The spatial index is a number identifying a small area on the sky. Considering yearly proper motions, window sizes, and the signal from the other field of view, we will need areas measuring some 15 arcseconds. A possible choice is the hierarchical triangular mesh (HTM). In order to take disturbing signals from the other field of view into account in the later processing steps, we will need spatial indices for both viewing directions

estimate for the RVS band. These quantities are derived exclusively from the transits of the astrometric and photometric field, and not from the RVS spectra themselves.

The IDT output for RVS will be a set of files of raw data, where only the spatial indices are added, while RVS flux and cross matching will be written to separate files as they may later need updating. The total amount of output will be similar to the volume of telemetry, i.e. 42 GB/day.

4.1.5 Processing of photometric data

IDT will process the SM and AF data to derive G band fluxes, which will later be used in the photometric processing for calibration and analysis. This forms part of the image parameter extraction, and is discussed in Sect. 4.1.6.

For the BP and RP spectra, IDT will run a colour extraction module, which will estimate the RVS flux, estimate colour information for defining the relevant PSF/LSF for SM and AF, and estimate the colours needed for the chromaticity as specified by the astrometric instrument model (cf. Sect. 5.1.2).

In principle, the photometric analysis requires an accurate transit prediction in order to work well. The along scan prediction is needed for the wavelength, and the across scan prediction for taking flux loss into account. A good transit prediction requires astrometric data which is not available during IDT, and we will therefore only obtain fairly crude colours at this stage. All these colours will later be updated by later photometric processing.

The IDT output for BP and RP will be a set of files of raw data, where we only add the Transit Identifiers and the spatial indices, whereas the cross matching information comes in a separate file. The data volume will be of the order 8 GB/day, as for the telemetry.

4.1.6 Processing of astrometric data

The AF and SM windows are analysed by fitting a PSF (for 2D windows) or an LSF (1D windows) to the samples. This gives us the accumulated flux, the transit time, and for two dimensional windows also the pixel coordinate in the across scan direction.

As explained in detail in the astrometric instrument model (Sect. 5.1.2), we model the signal (in electrons) in the normal 1D case in terms of an LSF, L, an amplitude (in electrons), α, a background (in electrons per sample), β, an image centroid (in TDI periods), κ, and a noise contribution: $N_k = \alpha L(k-\kappa) + \beta + \text{noise}$, where N_k is the signal (in electrons) derived from sample k after correcting for bias and gain. (See Eq. 8, and for the 2D case Eq. 7).

The centroid is derived by a maximum likelihood method ([Lin00]), and in our case the likelihood function becomes:

$$l(\kappa,\alpha) = \prod_{k=1}^{n} \frac{[\alpha L(k-\kappa)+\beta+r^2]^{N_k}}{N_k!} e^{-[\alpha L(k-\kappa)+\beta+r^2]}. \tag{4}$$

The centroid, κ, gives directly the transit time for the *accumulated charge* across the trailing edge of the CCD, in units of TDI periods, and for the 2D case the AC centroid, μ, gives the pixel coordinate again for the charge at the edge. Rather than the charge, the parameter we need is the mean position of the image during the integration. The charge is accumulated over a number of TDI periods, typically 4500 but fewer in case of gate activation, and results from the motion of the image within the sliding window during that time. If n_l CCD lines are active, the transit time , t, and flux, f, calculated by IDT is:

$$t = (\kappa - n_l/2)\tau, \tag{5}$$

and

$$f = \alpha/(n_l\tau), \tag{6}$$

where τ is the duration of one TDI period. As a first approximation, this transit time corresponds to the transit of the image across the central pixel line in the active CCD area. This approximation fails if the image motion within the window is irregular when the window enters and leaves the CCD, but corrections for effects like irregularities in the attitude or optical distortions do not belong in IDT.

Apart from the centroiding, the image analysis will also include a check for image structure, which may indicate duplicity or an extended or disturbed image.

We notice that for the commonly used one dimensional windows, the fraction of the flux lost in the across scan direction, due to AC motion and lack of centring, can only be determined from the photometric calibration, and only once the AC pixel coordinate can be predicted, and no correction for this loss will therefore be applied during IDT. Similarly, no corrections will be applied here for variations in sensitivity among the CCDs or from one area of a CCD to another.

For an accurate centroiding, and especially for an accurate flux determination, we must know the local sky background. This is in general not possible for the individual transits. Some of the sky mapper samples will mostly contain background signal, but the sky mappers only cover one field of view each. We will therefore use the sky background measured in both sky mappers for transits near in time, and when possible also measure it in the larger AF windows.

When the image parameters have been determined for each of the elementary CCD transits, a check will be made to see if the source is a *moving object*, i.e. a solar system member, or if the transit suffers from some disturbance so all or some part of it should be flagged as defective.

To be able to carry out the cross matching and to assign a permanent spatial index to each observation, we need a celestial position for the observed source at the subarcsecond accuracy level. The SM observations are always 2D, so the centroiding gives both a transit time and an AC pixel coordinate. Applying the current geometrical calibration, the position on the CCD is transformed to field angles. From the on ground attitude, just determined, and the relativistic model (aberration and light deflection), we derive the celestial position.

The transformation to field angles from pixel coordinates consists of a reference transformation and the application of a set of calibration coefficients. The reference transformation need not be very realistic, and simply calculates the focal plane coordinates from the nominal coordinates for each CCD and the pixel coordinates of the observation, followed by a gnomic projection. The focal plane coordinates derived here as well as the gnomic projection, should only be understood as convenient approximations. There is no way we will ever know the exact focal plane coordinates, or the exact optical projection. Finally, the most recent geometrical and chromatic calibrations will bridge the gap between this approximation and the real angles. The remaining steps to reach celestial coordinates are fairly trivial as we do not need the highest accuracy during IDT.

4.1.7 Outline of the processing

In Fig. 19 we show the data flow for the main IDT pipeline taking care of the photometric and astrometric observations. We begin with the photometric process because the colours are needed for the PSF model for the SM and AF centroiding. For the field angles and the moving objects we need geometrical calibration and the chromaticity correction. The attitude also enters the moving objects detection, and it enters when we calculate the celestial position. With the position we can assign a spatial index, e.g. HTM, and we can then store the raw objects (for astrometry and photometry), and the intermediate objects (astrometry).

The RVS processing (not shown) will need the flux estimate from the photometric process, and the spatial indices from the astrometric process, and will therefore not start until these processes have completed. It may, however, run during cross matching.

On average some 50 million transits must be processed on a daily basis, and sometimes many more. The transits are processed independently, and it is therefore obvious to choose a distributed processing for the treatment of the observations.

The IDT will produce several data sets, but the more massive are the sets of *raw data* and of *intermediate data*. The raw data contains the essential part of the telemetry with only minor additions like a spatial index. Once generated, the raw data is never updated. It contains essentially the same information as the telemetry, and will therefore also be of around 65 GB/day.

The intermediate data contains the reduced astrometric observations (transit times and fluxes) in a form suited for the further data processing in AGIS, and will give an additional 18 GB/day.

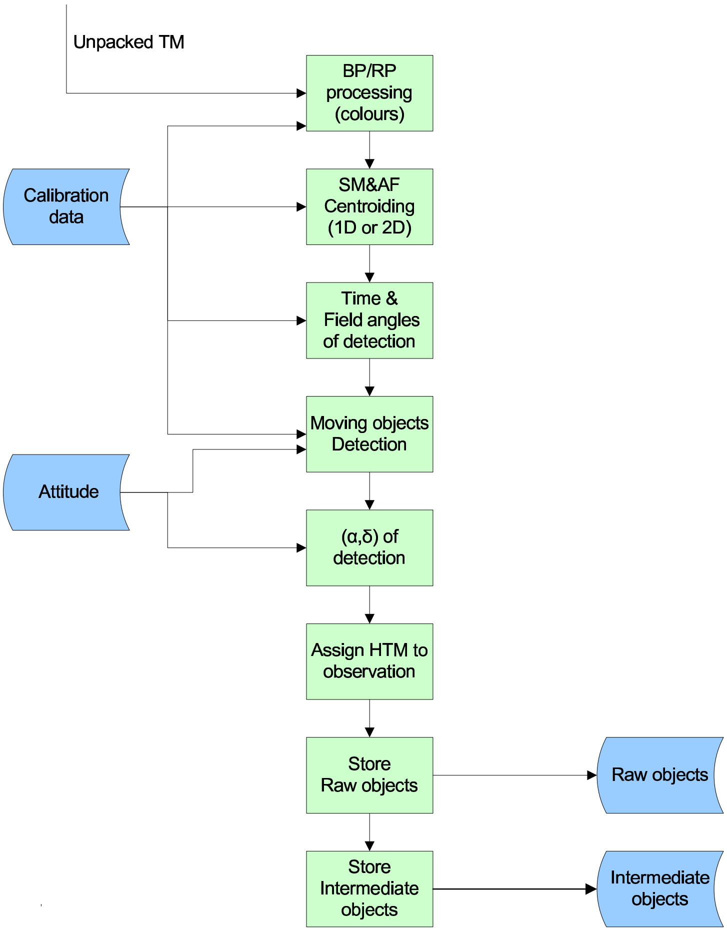


Figure 19: Simplified overview of the main IDT pipeline, showing the steps of the photometric and astrometric processing. Two important processes are not shown here: The initial attitude updating is indicated only by its result (attitude input box at left). It forms a side loop to the IDT process, running after the transit time and field angle calculation. It uses the small subset of very bright transits for a cross-match with a specific, small attitude star catalogue and subsequently performs an attitude solution from them. The second omitted process is the main cross-matching, running before the "store intermediate objects" step. It produces both the main cross-match table (not shown in the figure), and also creates additional entries for the main catalogue source list from the remaining unmatched transits. In this way it ultimately defines the final list of Gaia sources (after iterations through the IDU cross-matching, see Section Sect. 5.1.6).

4.1.8 Cross matching

The cross matching establishes the link between the intermediate data objects and the catalogue of sources. The link will take the form of a *Match Table*, giving the transit identifier, the spatial index, and the source identifier. The cross matching also needs the solar system ephemerides, and ephemerides of orbiting stellar systems. New sources must be added to the main source catalogue.

Especially during the first years of the mission, the source catalogue and the ephemerides will be incomplete, and we cannot avoid some small fraction of incorrect identifications. Solar system objects may occasionally be mistaken for stars; stars with higher proper motion than assumed may generate additional entries; double stars may sometimes be resolved and sometimes not; etc. These errors will gradually be weeded out as the source data improves, and as the cross matching is repeated.

Some attention must be paid to the list of sources. It will not be time efficient to update the source catalogue in the main database on a daily basis, so instead IDT will work with the same main catalogue for some period of time. As new sources are encountered, they are added to a catalogue of new sources rather than the main catalogue. From time to time the new sources are merged with the source catalogue of the main database, and an updated version of the main catalogue will be extracted for the coming IDT runs.

4.1.9 Updating intermediate data

In IDT we are forced to make approximations and to use preliminary values for calibrations and PSFs, and disturbances from sources in the immediate vicinity will not yet be known. The intermediate data will therefore need updating from time to time. This updating is discussed in Sect. 5.1.6.

4.2 First Look

Astrometric space missions like Gaia and Hipparcos have to simultaneously determine a tremendous number of parameters concerning astrometry and other stellar properties, the satellites attitude as well as the geometric and photometric calibration of the instrument. Additionally, in the case of Gaia, general-relativistic effects have to be incorporated in the data reduction.

To reach the targeted level of precision for these missions, many months of observational data have to be incorporated in a global, coherent and interleaved data reduction. By no means neither the instrument nor the data health can be checked at the desired level of precision by standard procedures applied to typical space missions. Obviously it is undesirable not to know the measurement precision and instrument stability until more than half a year of the mission has elapsed. If any unperceived, subtle effect would arise during that time this would affect all data and result in a loss of many months of data.

For this reason a 'Detailed First Look' (DFL) has to be installed to judge the level of precision of the stellar, attitude and instrument parameters and achieve its targeted level by means of a sophisticated monitoring and evaluation of the observational data. This in-depth scientific assessment of the quality of the Gaia observations shall be performed within about 24 hours after its reception at the Science Operation Center (SOC). The diagnostic output will be accessible in a database. If modifications to the satellite operations appear necessary, this will be communicated to the SOC.

The DFL includes diagnostics for the astrometric, photometric, and spectroscopic instruments of Gaia. Since Gaia's main goal is high-precision astrometry, the astrometric tasks will be described in more detail here, but many of the photometric tasks or RVS tasks can be treated analogously.

DFL is divided into two major groups of diagnostics:

- the examination of one-day calibration (ODC) processes and
- the statistical investigation of the quality of the data processed by the Initial Data Treatment (IDT).

The ODC processes comprise one-day calibrations for astrometry, photometry, radial-velocity measurements, the PSF/LSF, and the CCDs.

Like AGIS the astrometric ODC, which is called One-Day Astrometric Solution (ODAS), aims at an astrometric self-calibration of Gaia, but takes into account the special problems of the restriction in the time basis; the input data consist of a contiguous set of observations down to a magnitude of about G=16 from a period of 14 to 24 hours each day. During one day the scanning plane of Gaia will vary only slightly ($\approx 4^\circ$), so that scan directions change only little. Therefore, only the along-scan coordinates along a Reference Great Circle (RGC) can be measured with high precision, across-scan information is only provided by the sky mappers, unbinned windows on the AF CCDs, and the finite change of scanning directions near the nodes of the scanning law. No proper motions, parallaxes, or global parameters will be determined by the ODAS. Besides the determination of source positions, very precise values for the geometric calibration of the CCDs, and an attitude (OGA2) being orders of magnitude more accurate than the On-ground Attitude (OGA1) provided by IDT will be available.

The so-called Ring Solution will be used as the basic module of the ODAS. It is a direct (non-iterative) solution of the astrometric problem restricted to about 24 hours. Therefore, the correlations between source positions, attitude, and geometric calibration can be fully taken into account. The main diagnostic ODAS output for the DFL consists of residues and normalised residues for each elementary observation.

DFL will investigate the normalised residues in detail and for many different subsamples (magnitude intervals, colour intervals, position on the focal plane/CCD, field of view, time) in order to detect deviations from the assumed error model with μas accuracy, which means that we can precisely test whether the centroiding works fine and the attitude is sufficiently stable.

The second class of DFL diagnostics utilizes observational data directly after processing by the IDT. This includes the investigation of image centroids and widths for different magnitudes, including those below the limit used for ODC calibrations. This would e.g. allow to find out whether low-level signal disturbances spoil the faint images by checking whether the on-board centroiding also works sufficiently for the faint stars. In addition DFL will investigate the detection statistics (does the on-board detection algorithm work under all circumstances) and important properties of the electronic detectors (noise, electronic ADC noise, low-energy cosmic rays, straylight or background). The statistics of the data gathered during one day will be compared to a-priori expectations and to the data from previous days.

The goal of the FL tasks is to detect both major flaws and small deviations from the assumed performance. Major flaws could e.g. require the switching to backup instruments, or the restart or rewriting of on-board software components — events that hopefully occur never or very rarely. Smaller deviations (down to the μas level in astrometry) from the expected data quality will almost inevitably happen during the whole mission. Many of the resulting measures fall rather into the category optimisation and the slight change of some on-board parameters. It could e.g. also mean the optimisation of the attitude control strategy, if the attitude is not smooth enough on the μas level. Other examples could be that the PSF/LSF provides hints that the focusing is not optimal (so that small corrections should be applied), or that the PSF overly varies with the position in the focal plane, so that the optical alignment needs slight adjustments.

The reaction to such events needs several days (at least if they occur for the first time or rarely). In order not to lose a significant amount of additional time for measurements, it is necessary to perform these diagnostics on a daily basis with all stars down to a certain magnitude limit ($G \approx 16$ for astrometry and photometry, $G \approx 10$ for RVS) plus significant samples of fainter stars distributed as uniformly over the sky as possible (in order to allow meaningful time analyses). After correction measures or optimisation actions have been performed, the daily diagnostics will allow immediate assessment whether the expected improvements have been achieved.

One of the major predictable problems of Gaia's calibration is the treatment of radiation-induced CCD damages. The centroid bias and the total flux of any Gaia image on the CCDs will depend on a huge number of parameters, among them the total number of traps, the position in the focal plane, the position within a CCD, the magnitude of the source, the time since charge injection, and the time since previous sources moved through the same CCD column. The radiation-induced centroid shifts have to be taken into account by a large number of additional geometric calibration parameters (including those mentioned above).

Other processes where the astrometric DFL can help to optimise the strategy are the efficient use of the CCD gates in order to avoid saturation but simultaneously use the full-well capacity of the CCDs as completely as possible. This strategy may change over the mission due to the degradation of the CCDs.

It is one of the most important tasks of ODAS to determine these parameters with an already relatively high precision, while the final treatment is in the responsibility of the AGIS. The flux deficiency due to radiation damage is part of the photometric (One-day) calibration. And finally, the PSF/LSF calibration will determine the changes in image shape after radiation damage.

The photometry ODC uses bright standard stars, well distributed in position on the sky, effective temperature and surface gravity. Each day the calibration parameters are compared with data measured during the previous days in order to check their stability.

For the RVS the ODC consists of the extraction of the spectra from the RVS IDT, a single-transit analysis, and a calibration pipeline for all bright stars, which uses reference stars consisting of ground-based standards or suitable/stable stars identified by the calibration pipeline itself. Similar to the astrometric case, a faint-star analysis is performed. For those stars with sufficient SNR an additional single-star analysis is conducted. All sorts of calibration parameters and diagnostics will be compared with the values from previous days in order to assess the stability and quality of the spectra and the stability of the instrument.

For all scientific instruments, checks of the functioning of the on-board software, CCD diagnostics, etc. are performed. Of particular importance is whether on-board source detection, image centroiding and window propagation work properly. For this purpose it must be checked whether under all circumstances the right number of sources is found for stars of all relevant magnitudes, and whether all images and spectra are well centered in the windows downloaded to the ground.

All diagnostics performed by the FL software will be judged by the First Look Scientists, experts in the scientific aspects of the Gaia mission and members of the DPAC. Their goal is the monitoring, optimisation and calibration of the Gaia payload, as well as the identification of problems and required actions. They are located at their home institutes (except during the commissioning phase, see below) and will directly communicate with the SOC, providing input into payload operation activities.

The decisional process towards such activities will be initiated by the First Look Scientist, possibly after consultation with relevant experts within the DPAC. Proposed actions will be forwarded to the SOC , assessed by the Science Operations Manager of the SOC and fed into the Science Mission Planning System (SMPS) at the SOC to produce Payload Operations Requests . These requests will be sent to the MOC and, on approval by the Spacecraft Operations Manager and Mission Manager, should be uploaded to the spacecraft. In critical cases an ad-hoc assessment board will be formed by ESA to advise the Mission Manager. In these cases the process may take longer than the few days mentioned earlier.

The management details of these procedures will be specified in the Gaia Science Operations Scenario Document and in the Gaia MOC-SOC Interface Document, both to be produced by ESA/SOC.

During the commissioning phase, the First Look Scientist team will be located at the SOC in order to perform a first analysis of the Gaia data and to provide a first high-precision in-flight calibration of the scientific instruments.

Besides optimisations of Gaia's scientific instruments, output of FL will also provide immediate check of all IDT tasks, so that the experience can directly be used to perform corrections and optimisations. Moreover, the calibration output of the ODC tasks is much more precise than what has been previously measured, at least until the first global solution has been performed. Therefore, these data can be used as high-precision starting for the global solutions. In the case of astrometry, these are the geometric calibration of the CCDs, an improved on-ground attitude (OGA2), and very good source positions.

Although not directly connected to the goals of the First Look, for purely practical reasons the Science Alerts software will run within the First Look processing framework. The flux-based Science Alerts aim at quickly detecting supernovas etc. in the Gaia observations, with the purpose of allowing follow-up observations with other instruments. To be useful they need to be produced on the same time scale as the FL diagnostics, viz. within about a day. Running the Science Alerts software within the IDT/FL framework in particular avoids the necessity of establishing a guaranteed daily data transfer of IDT results from the SOC to one of the other DPCs. However, the scientific responsibility for the development of Science Alerts software and for the assessment and usage of respective outputs rests with the photometric and spectroscopic scientific teams (see Sect. 8.6 and work package descriptions for GWP-T-517-00000 and GWP-S-650-0000 in Sect. B).

5 General data processing : purpose and methods

This section describes the theoretical approach of the on ground data processing necessary to transform the satellite data into the final data products. Where appropriate, mathematical models of the measurements and adopted methods are presented together with a functional analysis identifying the important elements or steps in the processing. Whereas the section provides the necessary information needed to understand the content of the top-level SW modules it does not cover the practical implementation and data-handling aspects.

5.1 Astrometric core solution

A main product of the Gaia mission will be astrometric information about all the objects observed throughout the mission. In some cases (such as a faint variable star or extra-galactic supernova observed only in a few scans near its maximum) these data may not be much more than a single-epoch position (t, α, δ) and a corresponding error ellipse; but for the majority of objects the astrometric data will include many more parameters needed to characterize its position as a function of time. For a typical stellar objects, at least the standard five *astrometric parameters*—position (α, δ), parallax (π) and proper motion $(\mu_{\alpha*}, \mu_\delta)$—will be provided. In any case, it is of paramount importance that these data are given in a single, internally highly consistent reference frame for the positions and proper motions, and on an absolute scale for the trigonometric parallaxes. The uniqueness of the reference frame, and the absoluteness of the parallaxes, are prerequisites for any further scientific analysis of the data in terms of distances and motions. Similarly for the solar-system objects: their observations are eventually reduced into a set of orbital elements which must be given in the same reference frame. Moreover, it is a requirement that this reference frame should coincide, as accurately as possible, with the International Celestial Reference System (ICRS) defined in terms of extra-galactic radio sources.

It is the purpose of the astrometric core solution to provide the basic data needed to connect any Gaia observation directly to this reference frame and parallax scale. To get a feeling for what is involved, consider a single passage of some unknown point-like object across one of the CCDs in the astrometric field. In the CCD output signal, the object will be seen as a temporary increase of the detected flux in a few consecutive sample readouts. Using the on-board timing of the samples, and its subsequent correlation with a ground-based clock, it will be possible to derive the accurate time at which the centre of the object moved across a fiducial reference line on the CCD, expressed on some convenient time scale. The resulting quantity, referred to as the *transit time* of the object on the CCD, constitutes the basic observational datum for all the astrometry. Which factors, apart from unavoidable measurement noise, influence this observed transit time?

The instantaneous celestial direction of the incoming light rays is obviously a primary determinant of the observed transit time, but equally important is the instantaneous celestial orientation of the instrument itself, i.e., the *attitude*. However, the attitude only gives the pointing of the instrument axes, whereas the precise location and orientation of the actual CCDs relative to these axes require further specification. This is provided by the *geometric calibration* of the instrument. This takes into account not only the physical geometry of the CCDs, but also optical distortion including the differential mapping of the two viewing directions—and thus the basic angle. The geometric calibration will actually include many more subtle effects operating on the build-up, transfer, readout and electronic processing of the charge images on the CCDs, as in practice they cannot be completely separated from a physical or optical displacement. Given a knowledge of the attitude and geometric calibration it is thus possible to reconstruct the accurate celestial direction of the incident light rays, at least in the along-scan direction. Interpreting this direction in terms of the astrometric parameters of the object requires additional transformations, taking into account the barycentric motion of the satellite (stellar aberration), gravitational light deflection, and possibly other effects. Although these are considered to be essentially known even at the accuracy levels required for Gaia, the possibility of a further *global* modelling of the observations in terms of such effects is foreseen.

The determination of the astrometric parameters of the objects is thus very intimately linked with the determination of numerous additional parameters representing the attitude of the instrument as function of time and its geometric calibration (possibly also some global parameters, such as PPN parameters describing the metric). None of these can be determined to sufficient accuracy by independent measurements. For example, the geometric calibration is needed over the whole astrometric focal plane to an accuracy of order $30\,\mu$as per CCD transit, or 5 nm in linear measure; this can in reality only be achieved by using the vast number of high-precision observations collected in the course of the normal observations. This leads to the concept of Gaia as a *self-calibrating* instrument, in the sense that the normal science observations provide the main input not only for calculating the astrometric parameters, but also for all the other—attitude, geometric calibration and global—parameters. This does not mean that ground-based calibrations are not needed: indeed they are necessary as a first approximation for the in-flight calibration, and to allow a correct modelling of the errors in this process.

Having established a sufficiently accurate mathematical and statistical model representing the observed transit times in terms of the astrometric, attitude, calibration, and global parameters, it is *in principle* straightforward to derive the most likely parameter values for the given observations. *In practice* the problem is a daunting one because of its size (in terms of the number of observational data points and number of parameters involved), complexity (in terms of any given parameter affecting a very large number of data points), initial uncertainties concerning many details of the instrument modelling, and the need to make the solution very robust against modelling errors.

Robustness is particularly important with respect to the modelling of the stellar parameters. The five-parameter astrometric model used as a basis for the astrometric core solution is strictly valid only for stars in uniform space motion relative to the solar system barycentre. This will be the case, to sufficient accuracy, for single stars without any nearby perturbing stellar or planetary companion. These may actually be rare cases, but fortunately the model will be sufficiently good in a range of other situations where the perturbations are small enough. Again, only the Gaia observations themselves are accurate enough to decide whether a specific object agrees with the simple 'single-star' model or not. Robustness is achieved by successively weeding out from the core solution all objects where there is some indication that the single-star model does not hold. The remaining *primary stars* should be sufficiently many, and have a suitable distribution in magnitude, colour, and position, to allow an accurate determination of the geometric calibration and attitude. The aim is to use about 10^8 primary stars in the final astrometric core solution, or about 10% of the total number of objects observed. Once the attitude and geometric calibration have been established by means of the primary stars, they can be used for a more detailed analysis of the remaining objects (the *secondary stars*) as well, using more complex models for their motions when required (e.g., binaries and stars with planetary companions).

The following sections give some details on the currently adopted models for the primary stars, instrument, and attitude. The proposed method for the simultaneous determination of astrometric, instrument and attitude parameters is described in Sect. 5.1.4, and subsequent sections deal with its application to the secondary stars, the need for an outer iteration loop, and the verification of results.

5.1.1 Astrometric Model for single stars

5.1.1.1 Relativistic reference systems

It is widely known that in order to model and successfully interpret Gaia data the models used for data processing must be fully compatible with general theory of relativity. The modelling begins with defining relativistic astronomical reference systems. The International Astronomical Union has adopted a set of relativistic astronomical reference systems to be used for modelling of high-accuracy astronomical observations [IAU01, Ric01, SKP*et al.*03]. Each of the IAU reference systems are defined by the form of its metric tensor within the post-Newtonian approximation of general relativity. Two reference systems have been explicitly adopted by the IAU:

- Barycentric Celestial Reference System (BCRS) is the fundamental reference system covering the solar system and all observed sources. The centre of the BCRS lies in the barycentre of the Solar system. The word "celestial" in the name of BCRS is used to underline that the BCRS does not rotate with the Earth, but is asymptotically Minkowskian. The latter property means that remote sources (quasars) can be assumed to be at rest with respect to the BCRS in some averaged sense. The BCRS will be widely used for the modelling of Gaia observations. This is a reference system underlying the resulting Gaia catalogue. The astrometric parameters of observed sources (coordinates and distances, proper motion etc.) are all defined in the BCRS. The BCRS is also used to model light propagation between the source and the observer. The coordinate time of the BCRS is called Barycentric Coordinate Time (TCB). The TCB will be used to parameterize the Gaia catalogue and orbital solution of solar system bodies. The Parameterized Post-Newtonian (PPN) version of the BCRS valid for certain class of metric theories of gravity can be found in [Wil93, KS00, KV04].

– Geocentric Celestial Reference System (GCRS) is a reference system physically suitable for modelling of physical processes in the vicinity of the Earth. The GCRS is constructed in such a way that the gravitational fields generated by other bodies are reduced to tidal potentials and are thus effaced as much as it is possible according to General Relativity. The coordinate time of the GCRS is called Geocentric Coordinate Time (TCG). Its scaled version called Terrestrial Time (TT) is a physical model of TAI. In Gaia this reference system will only be used to model Gaia tracking data and to relate the on-board time to UTC.

The underlying theory of the local reference systems like GCRS can be applied to any massive or massless bodies of an N-body system. In particular a GCRS-like reference system can be constructed for the Gaia satellite [Kli04]. This reference system is called Centre-of-Masses Reference System (CoMRS). That reference system is physically adequate to model any physical processes occurring within the satellite (for example, the process of registration of incoming photons and the rotational motion of the satellite). From the relativistic point of view CoMRS is a local reference system kinematically non-rotating relative to the BCRS. The coordinate basis of CoMRS coincides with a kinematically non-rotating tetrad co-moving with the Gaia satellite. The Spacecraft Reference System (SCRS) discussed below is a reference system rigidly rotating with respect to the CoMRS. The attitude parameters derived from GIS parameterize the rotational matrix between CoMRS and SCRS and have, thus, very precise meaning in the relativistic framework. A different approach for attitude modelling in the relativistic framework can be found in [BCd03, CdFB^{+}04].

5.1.1.2 Structure of the relativistic model

The relativistic model for Gaia has been described in several papers using different approaches. The Relativity and Reference Frame Working Group (RRFWG), which worked until the end of 2005 as advisory body to the Gaia Science Team, decided to use the model GREM as the baseline for the pipeline data reduction, while the RAMOD model ([dFCV^{+}04, dVC^{+}06] and references therein) will be utilized for an independent verification of the Gaia sphere solution (see section 5.1.8). The GREM model (to which we will refer also as the Gaia *baseline* model) has been described in great detail in a number of publications [Kli03b, Kli04, KP03] and technical reports

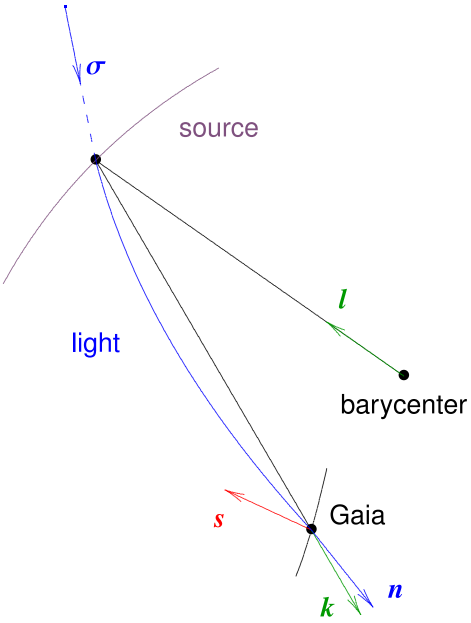


Figure 20: Five principal vectors used in the baseline relativistic model for Gaia (GREM, see text for explanations).

remote sources: $\boldsymbol{s} \overset{(1)}{\longleftrightarrow} \boldsymbol{n} \overset{(2)}{\longleftrightarrow} \boldsymbol{\sigma} \overset{(3)}{\longleftrightarrow} \boldsymbol{k} \overset{(4)}{\longleftrightarrow} \boldsymbol{l}, \pi \overset{(5)}{\longleftrightarrow} \boldsymbol{l}(t_0), \pi(t_0), \boldsymbol{\mu}(t_0), \dots$

solar system objects: $\boldsymbol{s} \overset{(1)}{\longleftrightarrow} \boldsymbol{n} \overset{(2,3)}{\longleftrightarrow} \boldsymbol{k} \overset{(6)}{\longleftrightarrow} \text{orbit}$

Figure 21: Transformation sequences for the baseline relativistic model (GREM, see text for explanations).

[KB03, Kli03d, Kli03c]. It consists essentially in subsequent transformations between 5 following vectors (Figure 20):

a) s is the unit observed direction (the word “unit” means here and below that the formal Euclidean scalar product $s \cdot s = s^i s^i$ is equal to unity),
b) n is the BCRS unit vector tangential to the light ray at the moment of observation,
c) σ is the BCRS unit vector tangential to the light ray at $t = -\infty$,
d) k is the BCRS unit coordinate vector from the source to the observer,
e) l is the BCRS unit vector from the barycentre of the Solar system to the source.

Note that the last four vectors should be interpreted as sets of three numbers characterizing the position of the source with respect to the BCRS. Vector s represents components of the observed direction relative to the local proper reference system of the satellite. All these vectors would change their numerical values if some other relativistic reference system is used instead of the BCRS. The model consists then in a sequence of transformations between these vectors as shown on Figure 21. The physical meaning of each transformation can be summarized as follows (the numbering here coincides with the numbering on Figure 21):

(1) aberration (effects vanishing together with the barycentric velocity of the observer): this step converts the observed direction to the source s into the unit BCRS coordinate velocity of the light ray n at the point of observation;
(2) gravitational light deflection for the source at infinity: this step converts n into the unit direction of propagation σ of the light ray infinitely far from the solar system at $t \to -\infty$;
(3) coupling of finite distance to the source and the gravitational light deflection in the gravitational field of the solar system: this step converts σ into a unit BCRS coordinate direction k going from the source to the observer;
(4) parallax: this step converts k into a unit BCRS direction l going from the barycentre of the Solar system to the source;
(5) proper motion, etc: this step provides a reasonable parameterization of the time dependence of l (and, possibly, of the parallax π) caused by the motion of the source relative to the barycentre of the Solar system;
(6) orbit determination process.

These transformations have already been discussed in full detail [Kli03b, Kli04]. The most complicated part of the model is the light deflection model where the effects of (1) monopole fields of all major solar system bodies, (2) quadrupole fields of the giant planets, and (3) gravitomagnetic fields due to translational motion of all major bodies should be taken into account in order to attain the accuracy of $1\,\mu$as. The magnitude of the two former effects is illustrated in Table 5. Practical consequences of the latter effect has been investigated in [KP03]. Moreover, each body with a mean density ρ and radius $R \ge (\rho/1\ \mathrm{g/cm^3})^{-1/2} \times 650$ km produces a light deflection of at least $1\,\mu$as. Therefore, a few tens of minor bodies (mainly, satellites of the giant planets) should also be taken into account in certain rare cases [Kli03b]. The parametrization of time dependence of l in the GREM framework looks exactly the same as in the Newtonian case. The only difference is that all vectors and parameters here (parallax, proper motion, etc.) are coordinate quantities defined in the BCRS.

Note that the “input” source parameters of the model is the barycentric vector l at the moment of observation and the barycentric distance to the source at the same moment. The relation between these parameters and the standard 5 astrometric parameters (two coordinates, parallax and two components of proper motion) is discussed in [Kli03c].

The baseline relativistic model GREM has been thoroughly tested by itself and compared with particular cases and formulas known from the literature.

Table 5: Principal gravitational effects in the light propagation in μas: δ_{pN} and δ_{ppN} are the post-Newtonian and post-post-Newtonian effects due to the spherically symmetric field of each body, δ_Q are the effects due to the quadrupole gravitational fields, respectively. Symbol "—" means that the corresponding effect is smaller than $0.1\,\mu$as. The angle $\psi_{\max}$ is the maximal angular distance between the body and the source at which the corresponding effect still attains $1\,\mu$as.

body	δ_{pN}	$\psi_{\max}$	δ_Q	$\psi_{\max}$	δ_{ppN}	$\psi_{\max}$
Sun	$1.75 \cdot 10^6$	180°	$\sim$1		11	$53'$
Mercury	83	$9'$	—		—	
Venus	493	4.5°	—		—	
Earth	574	178°	0.6		—	
Moon	26	9°	—		—	
Mars	116	$25'$	0.2		—	
Jupiter	16270	90°	240	$152''$	—	
Saturn	5780	17°	95	$46''$	—	
Uranus	2080	$71'$	8	$4''$	—	
Neptune	2533	$51'$	10	$3''$	—	

5.1.1.3 Implementation of the model

An ANSI C code has been written to implement the relativistic model in its full complexity [KB03, Kli03d]. The model has been implemented in two modes: predictor mode and corrector mode. *Predictor* mode implements the standard way of astrometric reductions when the observed direction to the source is predicted starting from some a priori catalogue parameters (coordinates, proper motion, parallax, etc.) of that source. The catalogue is supposed to be improved later by fitting the parameters to the whole set of data. *Corrector* mode implements the reductions in the opposite direction, that is, the momentary barycentric direction to the source is restored from the observed direction as good as possible. The model and the implementation have been massively tested. The implementation contains several ad hoc optimizations aimed to reduce execution time. An implementation of the full model in the GaiaTools is underway.

5.1.2 Instrument modelling

5.1.2.1 Signal model

Since the astrometric core solution only deals with (effectively) unperturbed, single stars, the stochastic model for the CCD signal becomes particularly simple and reduces to an appropriately scaled and displaced Point Spread Function (PSF), $P(u,v)$. In terms of the observed electron counts N_{km}, i.e. after correction for bias and gain, versus the pixel index along (k) and across (m) scan, the model is:

$$N_{km} = \alpha P(k-\kappa, m-\mu) + \beta + \text{noise} \tag{7}$$

α is the total intensity of the image in electrons, β the background level in electrons per pixel, (κ, μ) the position of the image centroid expressed in continuous 'pixel coordinates' (allowing accurate representation of the position on a sub-pixel scale). The noise term includes Poisson and read-out noise. On-chip or numerical binning of adjacent pixels into samples necessitates a modification of the above model. Of particular importance is that most astrometric observations will effectively be *one-dimensional* through a complete across-scan binning in each window; in this case the signal model can be expressed more concisely in terms of the Line Spread Function (LSF) $L(u) = \int_{-\infty}^{+\infty} P(u,v)\mathrm{d}v$:

$$N_k = \alpha L(k-\kappa) + \beta + \text{noise} \tag{8}$$

The estimation of the signal parameters α, β, κ, μ is done as part of the Initial Data Treatment (Sect. 4.1.6), essentially through a maximum-likelihood fitting of the PSF or LSF to the CCD samples. The estimated intensity parameter α is a main input for the photometric processing (Sect. 5.2), while the positional parameters κ and μ are the main input for the astrometric processing. The along-scan parameter κ can be translated into an effective *transit time* t_{obs} of the object across the CCD, through application of the appropriate timing information and subtraction of the known delays resulting from the TDI integration and electronic processing. The across-scan coordinate μ is retained in units of pixel columns.

The PSF and LSF functions themselves are also determined as part of the Initial Data Treatment. They are in general functions of the object's spectrum, position in the field, and different in the two viewing directions. They also vary slowly with time due to the evolution of various instrumental effects. In principle they are determined by analyzing the normalised counts $(N_k - \beta)/\alpha$ from millions of observations of primary stars, separately considering every relevant combination of spectrum, position in the field, etc.

The value of κ, and hence of the derived transit time t_{obs}, depends directly on what is considered to be the 'centre' of the LSF, i.e., where the origin $u = 0$ is placed in the calibrated function $L(u)$. This is purely a matter of convention, and a definition based on a fixed, odd *weighting function* $w(u)$, will be employed, such that $\int_{-\infty}^{+\infty} L(u)w(u)\,\mathrm{d}u = 0$. A corresponding convention is used for the across-scan origin of the PSF.

5.1.2.2 The Scanning Reference System (SRS)

The Scanning Reference System (SRS) is a rectangular coordinate system $[\mathbf{x}\ \mathbf{y}\ \mathbf{z}]$ fixed with respect to the optical system and serving as an intermediary between the celestial reference frame and the observed data (t_{obs}, μ). On one hand, the attitude specifies the celestial orientation of the vector triad $[\mathbf{x}\ \mathbf{y}\ \mathbf{z}]$ at any instant; on the other, the

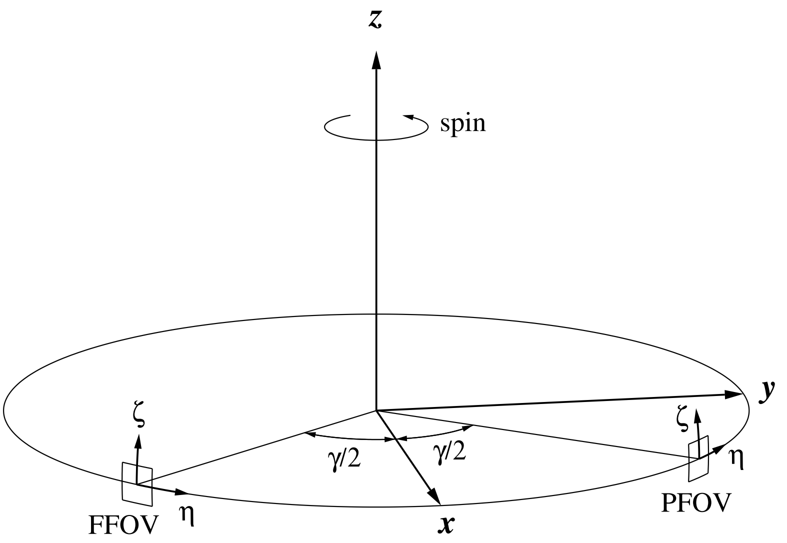


Figure 22: Definition of the Scanning Reference System (SRS) and field angles (η, ζ) in each field of view. The field centres projected onto the sky define the xy plane, the $\mathbf{z}$ axis normal to it, the $\mathbf{x}$ axis bisecting the viewing directions, and $\mathbf{y} = \mathbf{z} \times \mathbf{x}$ completing the right-handed triad of unit vectors $[\mathbf{x}\ \mathbf{y}\ \mathbf{z}]$.

layout of the CCDs are expressed in the SRS by means of the geometrical calibration. The SRS is defined by projecting the centre of the focal plane onto the sky through the preceding and following fields of view; these projections determine the **xy** plane and hence the SRS axes as shown in Fig. 23. Directions (unit vectors) are usually expressed in the SRS by means of *field angles* (η, ζ), where the origin is in the **xy** plane at exactly half the basic angle (γ) on either side of the **x** axis. When using the field angles it is necessary to know which field they refer to; this is given by the field index f (e.g., $f = 1$ in the preceding and 2 in the following field of view). The transformation between a direction in $[\mathbf{x}\ \mathbf{y}\ \mathbf{z}]$ and (f, η, ζ) is exact since the value of γ is fixed by convention.

This leaves just one thing undetermined for the SRS: where exactly is the 'centre of the focal plane'? In fact, no particular feature in the focal plane, such as a specific pixel in one of the CCDs, is singled out for this purpose; instead, the centre is implicitly defined by the adopted model of the geometrical calibration, to be described hereafter.

5.1.2.3 The geometrical calibration model

As explained in Sect. 5.1.2.1 the basic astrometric data resulting from a CCD transit consist of the CCD transit time (t_{obs}) and the across-scan pixel coordinate (μ). These refer to the image centroid crossing a fiducial transit line on the CCD (Fig. 23). The geometrical calibration model is a suitably parameterised representation of the

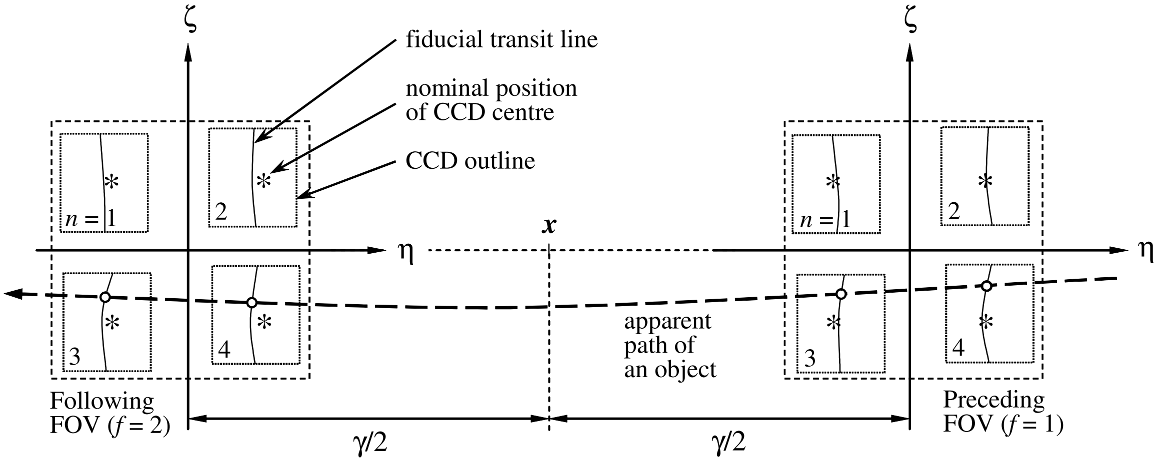


Figure 23: Schematic illustration of the geometrical configuration of CCDs in the astrometric field, as mapped onto the field angles (η, ζ) through a projection onto the sky. The diagram is not to scale, and for simplicity only four CCDs are shown $(n = 1 \dots 4)$. The nominal CCDs centres (η_n^0, ζ_n^0) are indicated by the asterisks. The actual geometry is given by the fiducial transit lines. The long-dashed line shows the apparent path of an object through the two fields of view; its intersection with the fiducial lines represent the CCD transits of the object (small open circles).

fiducial lines for all the astrometric CCDs, taken separately for the two fields of view. Thus, when an image in field f is observed by means of CCD number n, the fiducial line is given by $\left[\eta_{fn}(\mu), \zeta_{fn}(\mu)\right]$ as function of the across-scan pixel coordinate. The functions $\eta_{fn}(\mu)$ and $\zeta_{fn}(\mu)$ depend on a number of calibration parameters, some of which may in turn be functions of time, the brightness of the object, etc., in order to take into account the evolution of the instrument and other effects. The use of gates to avoid saturation when a bright object is observed will result in yet a different set of calibration parameters.

It is neither possible nor necessary to detail the complete calibration model here. Nevertheless, a rudimentary and somewhat simplistic model will be described, in order to illustrate some important concepts. Some details, especially on the motivation for the model, are given in [Bas06].

The equations for fiducial lines are

$$\left.\begin{aligned} \eta_{fn}(\mu) &= \eta_n^0 + \Delta\eta_{fn}(\mu) + \delta\eta_n(\mu) \\ \zeta_{fn}(\mu) &= \zeta_n^0 + \Delta\zeta_{fn}(\mu) + \delta\zeta_n(\mu) \end{aligned}\right\} \qquad (9)$$

where (η_n^0, ζ_n^0) is the nominal position of the centre of CCD number n in the field; $(\Delta\eta_{fn}, \Delta\zeta_{fn})$ express the large-scale geometrical calibration, which is in general different in the two fields (thus explicitly depending on f) due to optical distortion and the deviation of the actual basic angle from the nominal value γ. $(\delta\eta_{fn}, \delta\zeta_{fn})$ express the small-scale geometrical calibration, depending on CCD properties that are generally the same in the two fields. The functional dependence on μ is described by low-order polynomials for $\Delta\eta_{fn}(\mu)$ and $\Delta\zeta_{fn}(\mu)$, and by lookup tables for $\delta\eta_n(\mu)$ and $\delta\zeta_n(\mu)$. A clear separation between the large- and small-scale calibration parameters is achieved by suitable orthogonality constraints; for example, orthogonality between $\delta\eta_n(\mu)$ and the zero-order polynomials for $\Delta\eta_{fn}$ requires that $\langle\delta\eta_n(\mu)\rangle_\mu = 0$ for every n, where $\langle\rangle_\mu$ denotes an average taken over the appropriate range in μ. The origin of the field coordinates is now uniquely defined by the adopted nominal values (η_n^0, ζ_n^0) together with the additional constraints[8] $\langle\Delta\eta_{fn}\rangle_{fn} = 0$ and $\langle\Delta\zeta_{fn}\rangle_n = 0$, where $\Delta\eta_{fn} = \langle\Delta\eta_{fn}(\mu)\rangle_\mu$, etc. This is illustrated in Fig. 23, where the nominal positions (indicated by the asterisks) are symmetric and fixed in (η, ζ), while the actual CCDs (and fiducial lines) are variously displaced, but respecting the constraints given previously.

The calibration model outlined above essentially takes into account the geometrical mapping of pixel coordinates into the Scanning Reference System due to the physical location of the pixels in the focal plane and the properties of the optical system. Unfortunately the observed image centroid will be additionally affected by imperfections of the instrument, which are more difficult to model. Two of these effects are briefly discussed below.

5.1.2.4 Chromaticity

Although the astrometric instrument contains no refractive optics, the images are slightly chromatic because of the wavelength dependence of diffraction. Any wavefront aberration that is an odd function of the along-scan pupil coordinate (apart from a simple wavefront tilt) will produce an asymmetric diffraction image, and the width, shape and position of it will vary with wavelength. As a result, the centroid of the Line Spread Function $L(u)$ will depend on the actual spectral energy distribution of the object, folded with the instrument response function; this effect is known as *chromaticity*. With the kind of aberrations expected for Gaia, the relative displacement between an early-type star and a very red star may be of order 1 mas, or 20 times higher than the photon-statistical centroiding noise for a bright star. It is thus necessary to eliminate chromaticity to a high degree by careful calibration. The exact calibration model remains to be defined, but preliminary studies ([Lin05a],[dBLSea06],[JF06],[BGG$^+$06]) have shown that the centroid position of the polychromatic image is mainly a function of the effective wavenumber

$$\nu_{\text{eff}} = \frac{\int \phi_\lambda \, \lambda^{-1} \, \mathrm{d}\lambda}{\int \phi_\lambda \, \mathrm{d}\lambda} \tag{10}$$

where ϕ_λ is the detected photo-electron flux distribution per unit wavelength (cf. Fig. 24). The effective wavenumber can be estimated from the photometric observa-

---

[8] Note the difference between η and ζ; there is only one constraint for η, but there are two for ζ — one for each field of view

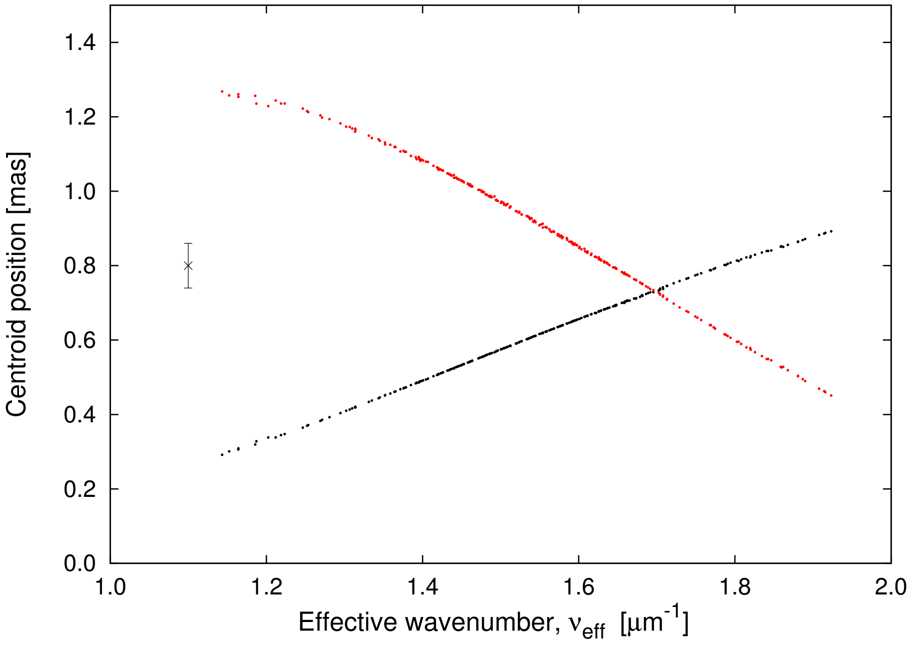


Figure 24: The dots show the computed centroid positions for the polychromatic diffraction images of different spectra, versus effective wavenumber, for two different representative wavefront error maps. Stellar spectra are taken from [Pic98], without and with interstellar extinction ($A_V = 2$ mag). The error bar shows the photon-statistical centroiding error for a single CCD transit of a bright star ($G \simeq 12$ mag).

tions (Sect. 5.2) and the chromaticity can then be modelled by means of a low-order polynomial in ν_{eff}. The coefficients are in general functions of the field index and position within the field, and may also have some variation with time. A reference wavenumber ν_0 must be adopted to fix the zero point, so that the geometric calibration model will include terms like

$$\eta_{fn}(\mu) = \cdots + C_1(\nu_{\mathrm{eff}} - \nu_0) + C_2(\nu_{\mathrm{eff}} - \nu_0)^2 + \cdots \tag{11}$$

where separate coefficients C_i apply for the different combinations of f, n and intervals in μ.

5.1.2.5 CCD radiation damage effects

Radiation damage to the CCDs, mainly caused by the proton flux from solar flares, produces an increasing number of permanent defects in the silicon crystal lattice. The defects may trap electrons as the charge packages are transported across the CCD, and release them at a later time, resulting in a loss of charges and a deformation and shift of the Point Spread Function. There are many varieties of traps with different capture cross-sections and release time constants — the latter strongly dependent on temperature. From the astrometric viewpoint, the systematic shift and deformation of the charge image are most important. Although the design and operation of Gaia aim to mitigate the damage as far as practicable (by shielding, proper choice of CCD operating temperature, use of charge injection and/or a diffuse optical background, etc.), remaining effects must be treated as part of the instrument calibration.

To establish the appropriate calibration model for these effects requires a very significant effort in terms of theoretical modelling, analysis of dedicated laboratory measurements, and exploratory analysis of the Gaia data themselves. It is believed that the underlying physics is sufficiently well understood to set up a quite realistic Monte Carlo model of the charge transfer process, including the detailed electron capture and release events. Such models do however depend on a large number of unknown parameters, which must be constrained by means of experimental data. Using the Monte Carlo model thus constrained, it will be possible to define the functional dependencies that can be expected and which therefore go into the Gaia calibration model. Following the detailed analysis of the actual data, it is however likely that the model can be significant improved.

The deformation of the Point Spread Function will be calibrated as part of the signal model (Sect. 5.1.2.1), while the shift will be included in the geometrical model. Part of the shift will be absorbed by the large- and small-scale functions $\Delta\eta_{fn}(\mu)$ and $\delta\eta_n(\mu)$, but the 'variable' part (depending on other data such as the flux parameter α) require additional terms. For the sake of illustration, assume that the shift has a flux-dependent component that can be modelled as a low-order polynomial of the logarithm of the flux. The appropriate terms are then

$$\eta_{fn}(\mu) = \cdots + D_1 \ln(\alpha/\alpha_0) + D_2 \ln^2(\alpha/\alpha_0) + \cdots \tag{12}$$

where α_0 is a reference flux used to fix the zero point, and where separate coefficients D_i apply for the different combinations of n and intervals in μ and time.

5.1.3 Attitude modelling

The attitude specifies the orientation of the instrument axes $[\mathbf{x}(t)\ \mathbf{y}(t)\ \mathbf{z}(t)]$ of Gaia (i.e., the Scanning Reference System, SRS; see Fig. 22) in the celestial reference frame (in principle the CoMRS defined in Sect. 5.1.1). If the latter is represented by the vector triad $[\mathbf{X}\ \mathbf{Y}\ \mathbf{Z}]$, with $\mathbf{X}$ towards $(\alpha = 0,\ \delta = 0)$ and $\mathbf{Z}$ towards $\delta = +90°$, the attitude at a particular instant may be represented by the 3×3 matrix

$$\mathbf{A}(t) = \begin{bmatrix} \mathbf{x}(t)\cdot\mathbf{X} & \mathbf{x}(t)\cdot\mathbf{Y} & \mathbf{x}(t)\cdot\mathbf{Z} \\ \mathbf{y}(t)\cdot\mathbf{X} & \mathbf{y}(t)\cdot\mathbf{Y} & \mathbf{y}(t)\cdot\mathbf{Z} \\ \mathbf{z}(t)\cdot\mathbf{X} & \mathbf{z}(t)\cdot\mathbf{Y} & \mathbf{z}(t)\cdot\mathbf{Z} \end{bmatrix} \tag{13}$$

where the dot signifies the scalar product of two vectors. Physically, the attitude must be continuous in time, and the elements of $\mathbf{A}(t)$ are therefore also continuous functions of time. However, it is not convenient to model the attitude directly in terms of the nine functions $A_{ij}(t)$, because it would be difficult to ensure that they always satisfy the six orthonormality constraints $\mathbf{A}(t)^T\mathbf{A}(t) = \mathbf{1}$. Alternatively, the attitude can be represented by means of three Euler angles, or by a quaternion. The choice for the Gaia data processing will be to use a quaternion representation, which is free of singularities, computationally efficient, and very well adapted to the attitude determination process. The attitude quaternion $\mathbf{q}(t) = \{q_1(t), q_2(t), q_3(t), q_4(t)\}$ is connected to the attitude matrix through

$$\mathbf{A} = \begin{bmatrix} q_1^2 - q_2^2 - q_3^2 + q_4^2 & 2(q_1q_2 + q_3q_4) & 2(q_1q_3 - q_2q_4) \\ 2(q_1q_2 - q_3q_4) & -q_1^2 + q_2^2 - q_3^2 + q_4^2 & 2(q_2q_3 + q_1q_4) \\ 2(q_1q_3 + q_2q_4) & 2(q_2q_3 - q_1q_4) & -q_1^2 - q_2^2 + q_3^2 + q_4^2 \end{bmatrix} \tag{14}$$

The components of $\mathbf{q}(t)$ must satisfy the constraint $q_1^2(t) + q_2^2(t) + q_3^2(t) + q_4^2(t) = 1$, but this is easily achieved through a simple normalization procedure. The component functions $q_i(t)$ can thus be represented by any suitable set of continuous basis functions.

The angular velocity and angular acceleration of Gaia in either $[\mathbf{x}\ \mathbf{y}\ \mathbf{z}]$ or $[\mathbf{X}\ \mathbf{Y}\ \mathbf{Z}]$ are readily computed in terms of $\mathbf{q}$, $\dot{\mathbf{q}}$ and $\ddot{\mathbf{q}}$. In the rigid-body approximation, they are interconnected via the inertia tensor to the torques acting on the satellite by means of Euler's Equations of Motion. Thus, if the torques are known, or can be accurately modelled, there are additional dynamical constraints on $\mathbf{q}(t)$. However, with the foreseen attitude control system, high-frequency ($\sim$ 0.01–10 Hz) thruster noise effectively limits the 'dynamical memory' of Gaia to a few seconds, which makes it doubtful whether the inclusion of dynamical constraints could improve the attitude determination accuracy in a significant way. Pending further investigation and characterization of the micro-propulsion system, spacecraft dynamics is therefore essentially ignored in the current attitude modelling.

The functions $q_i(t)$ are modelled as cubic splines defined on a semi-regular knot sequence. The typical knot separation, which determines the flexibility of the splines and thus its ability to represent high-frequency attitude excursions, may be in the range 5–15 s. The lower limit is set by the CCD integration time (4.4 s), which effectively filters out higher frequencies, and the rate of primary star measurements in the astrometric fields—in the Galactic polar regions some 50 along-scan measurements are obtained per second in the brightness range relevant for AGIS (G<17). The upper limit is set by the level of thruster noise. The splines are expressed as linear combinations of basic splines (so-called B-splines), the coefficients of which form the set of attitude parameters. In the attitude determination, these parameters are estimated by a weighted least-squares fitting of the predicted (t_{obs}, μ) values to the observed ones. A single, continuous attitude solution will be obtained for each uninterrupted interval of observations, possibly up to several weeks' duration.

Micrometeorite hits causing a sudden, noticeable change in the angular velocity of Gaia will happen several times per day. Their presence in the data can be detected through a dedicated post-fit analysis of the attitude residuals, where they produce a characteristic pattern (see [vLF05], [vL05]). Once detected, the hits can be included in the modelling by inserting multiple knots at the estimated time of impact. In a more refined analysis, the torque impulse might be estimated and folded through the dynamical model in order to reduce the number of free parameters in the additional spline coefficients. Larger impacts, causing a temporary interruption of science observations and hence in the attitude modelling, will happen perhaps once per week.

As previously described, the attitude is determined as part of the astrometric core solution through a least-squares fitting of the quaternion function to the along- and across-scan measurements of primary stars. However, the convergence of the global iterative solution (GIS) and the systematic accuracy of the resulting reference frame critically depends on the solution's ability to detect and even out regional errors through their superposition in the two fields of view. This will not work properly if the weight of the observations are very different in the two fields of view, since the attitude errors would then essentially be determined by the regional errors in the area with the higher weight. This could easily happen if the density and brightness of primary stars were allowed to follow the general distribution of stars in the sky. For this reason it will be necessary to restrict the primary stars used for the attitude determination to a subset with a more even weight distribution. (The instrument calibration, on the other hand, will benefit from the maximum number of primary stars independent of their sky distribution, and especially from their improved coverage in magnitude, colour, etc.) An alternative method will be to equalize the weights between the fields at all times by applying an artificial downweighting of the observations in the dominant field (but only for the attitude determination) [vLF05].

5.1.4 Astrometric solution of the GIS stars

As outlined in Sect. 5.1, the astrometric core solution needs to determine a very large number of astrometric, attitude, calibration and global parameters for a subset of primary stars, using the observed CCD transit times $t_{\rm obs}$ and across-scan coordinates μ for these stars as the main input. Moreover, these parameters must be determined in an entirely self-consistent manner, and the astrometric and attitude parameters must in the end refer to the International Celestial Reference System (ICRS).

5.1.4.1 Principle of solution

Assuming some 10^8 primary stars, the total number of unknowns for the astrometric core solution includes some 5×10^8 astrometric parameters, $\sim 10^8$ attitude parameters, and a few million calibration parameters. The condition equations connecting the unknowns to the observed data are intrinsically non-linear, although they generally linearise well at the sub-arcsec level. Direct solution of the corresponding least-squares problem is infeasible, by many orders of magnitude, simply in view of the large number of unknowns and their strong inter-connectivity, which prevents any useful decomposition of the problem into manageable parts. The proposed method is based on the *Global Iterative Solution* scheme [EFN$^+$97, Vol. 3, Ch. 23], which in the current context is referred to as the *Astrometric GIS* (AGIS) since related methods are adopted for the photometric and spectroscopic processing. It is necessary to have reasonable starting values for all the unknowns, so as to be close to the linear regime of the condition equations. These are generally provided by the Initial Data Treatment.

The idea of AGIS is then quite simple (as already outlined in the Gaia Concept and Technology Study Report, [gai00], Section 9.5.2), and consists of the following steps:

1. Assuming that the attitude and geometric calibration parameters are known, the astrometric parameters can be estimated for all the stars. This can be done for one star at a time, thus comprising a least-squares problem with only 5 unknowns and of order 1000 observations. Moreover, this part of the solution is extremely well suited for distributed processing.

2. Next, assuming that the astrometric parameters and the geometric calibration are known, it is possible to use the same observations to estimate the attitude. This can be done for each uninterrupted observation interval at a time. The length of such an interval may vary considerably, but for the sake of argument let us consider a length of one week. The number of attitude unknowns is then about 500 000 and the number of observations $\sim 2 \times 10^7$. The number of unknowns may seem rather large for a least-squares problem, but the band-diagonal structure of the normal equations resulting from the spline fitting makes the memory consumption and computing time a linear function of the number of unknowns, rather than the cubic scaling for general least-squares solutions. The problem is thus easily manageable, even for much longer time intervals.

3. Assuming then that the astrometric and attitude parameters are known, the geometric calibration parameters can be estimated from the residuals in transit time and across-scan field angles. In the simplest case of estimating (say) a single parameter $\Delta\eta$ for a particular combination of field and CCD index, this essentially amounts to computing a weighted mean value of the residuals for the corresponding subset of observations. The orthogonality constraints discussed in Sect. 5.1.2.3 are imposed afterwards by normalization.

4. It is now necessary to iterate the sequence of steps 1, 2, 3 as many times as it takes to reach convergence. Once the linear regime has been reached, the convergence should be geometric, i.e., the errors (and updates) should decrease roughly by a constant factor in each cycle. Based on simple considerations of redundancy and the geometry of observations, a convergence factor of 0.2–0.4 is expected. If a geometric behaviour is indeed observed, it may be possible to accelerate the convergence by over-relaxation. The iteration must be driven to

a point where the updates are much smaller than the accuracy aimed at in the resulting data.

5. After convergence, the astrometric and attitude parameters refer to an internally consistent celestial reference frame, but this does not necessarily coincide with the ICRS. A subset of the primary stars and quasars, with known positions/proper motions in the ICRS, is therefore analyzed to derive the nine parameters describing a uniform rotation between the two systems, plus the apparent streaming motion of quasars due to the cosmological acceleration of the solar-system barycentre. The astrometric and attitude parameters are then transformed into the ICRS by application of a uniform rotation.

It is envisaged that the whole sequence 1–5 is repeated several times during the processing, initially perhaps every 6 months during the accumulation of more observations. These repeats are called outer AGIS iterations. Optionally, the iteration loop 1–3 may also include an estimation of global parameters. From a processing point of view, the global parameters may be treated exactly as the geometric calibration parameters.

5.1.4.2 Input data

The input to AGIS consists of the elementary Astro observations (each corresponding to the transit across one CCD) for every detection of a non-solar system object, with associated information such as the photometric fluxes used to compute the Line Spread Function. These are provided by the Initial Data Treatment (IDT) per sky region, after cross-matching and including references to the astrometric input catalogue (see below). The IDT and First Look (FL) also provide the initial estimates of attitude and calibration for the AGIS.

Numerous auxiliary data sets are needed in various stages of the AGIS processing. Most important are the accurate ephemerides of Gaia and solar-system bodies in the barycentric reference frame, timing information for interpreting the transit times on the barycentric time scale, and housekeeping and satellite monitoring data that may help to set up the appropriate models for the attitude and geometric calibration.

AGIS will also use an astrometric input catalogue, containing the best ground-based positions and proper motions for a subset of the objects. This serves two purposes. First, by having good starting values for a significant fraction of the objects, the convergence of the solution is considerably faster than if all data have to be determined *ab initio*. Secondly, after convergence of AGIS, the resulting reference frame needs to be transformed into coincidence with the conventional celestial reference system (ICRS), for which the input catalogue will need to include an identified subset of objects that are part of the definition of the ICRS. It is emphasized, however, that the resulting astrometric reference frame is independent of the input catalogue in the sense that random and systematic errors of the latter are not propagated into the final results.

5.1.4.3 Selection of primary stars

The subset of primary stars is iteratively selected as part of the AGIS processing. The selection is based on the observed astrometric, photometric, spectroscopic and imaging characteristics of the stars, as well as the need for an adequate sampling versus position, magnitude and colour. Stars for which there is some indication that they may be non-single or otherwise deviating from the standard astrometric model are excluded from the primary set. Clearly it is an advantage from the point of view of accuracy to include as many primary stars as possible in AGIS. However, this must be balanced against the computing load that increases roughly in proportion to the number of objects, and the risk of introducing modelling errors if a more relaxed selection of primary stars is adopted. From simple arguments based on redundancy factors (number of observations per unknown) it is expected that the accuracy of AGIS will not benefit significantly from having more than about 10^8 primary stars, or 10% of the total number of objects. On the other hand, such a fraction does not seem unreasonable from an astrophysical viewpoint (frequency of multiplicity and variability, etc.). The aim is therefore to handle up to 10^8 primary stars in AGIS.

5.1.4.4 Robustness

The processes described above (updating of the astrometric, attitude, calibration and global parameters) will be based on the weighted least-squares method for computational efficiency. However, the implementations must be extremely robust in the following two senses. First, they should provide sensible (if not useful) results in all situations that can reasonably be encountered during the real mission, e.g., when the problem is under-determined because of a lack of data or too many parameters. This may for example happen with the attitude determination if too few primary stars are encountered in some region. These situations should always be handled gracefully (no floating exception due to a singular matrix, for example) and generate appropriate warning messages.

Secondly, the processes should be able to handle a significant fraction of outliers among the observations and still achieve solutions that are nearly optimal from the accuracy viewpoint (given the quality of the data). Both mild outliers (indicating longer tails than a Gaussian but still what you often expect in a real data set) and the occasional corrupt data item (clearly and utterly wrong) should be appropriately handled. Very often, especially in the initial stages of the processing, outliers are also produced by modelling errors, for example unrecognized astrometric binaries among the primary stars. In practice all of these cases require the consistent use of adaptive and robust solution methods. These will iteratively estimate the normal range of residuals while identifying and taking care of outliers (e.g. by rejection or down weighting). While simple rejection schemes may serve as a first approximation, the

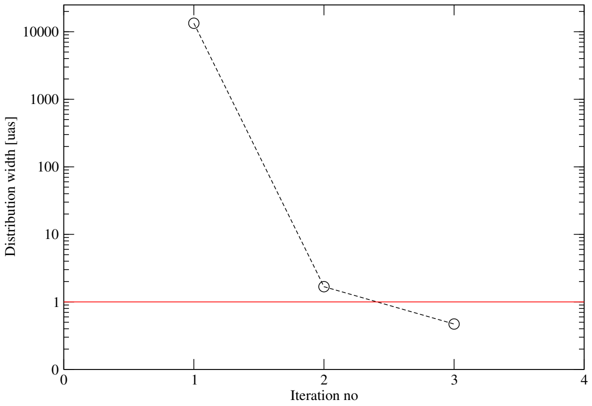


Figure 25: Evolution of the parallax distribution width with iteration number for AGIS cycle 452 involving 1.1 million stars simulated over a 5 year observation period. Convergence was achieved within 3 iterations.

more efficient and comprehensive treatment of outliers is an area in need of intense investigation.

5.1.4.5 Monitoring and analysis tools

The AGIS system will include a number of graphical and interactive tools to monitor, analyze and control the individual processes as well as the overall convergence of the iterations. It will be possible to select subsets of data (parameter values, updates and residuals) in terms of specified ranges for a number of control variables (such as time, position on sky, and position in the FOV), calculate their statistics and make various plots (histogram, scatter/density plots, time evolution, etc).

5.1.4.6 Preliminary Results

A first implementation of the AGIS system incorporating the source, attitude, and calibration update steps (see Sect. 5.1.4.1) has been completed in early 2006 [OHHL05, LHHO06]. Since then a series of test runs have been performed to both demonstrate the correctness of the implementation and to validate the fundamental working principles of AGIS on a large-scale.

In this system the iterative loop is controlled through a stopping criterion that is based on the monitoring of the adjustments to the stellar parallaxes. In each iteration the updates to the parallax values with respect to the previous iteration are binned into a histogram and the width of the resulting distribution determined. The system is considered converged if this width falls below 1 microarcsec. Fig. 25 shows the evolution of the parallax distribution width with iteration number for a completed AGIS validation run involving 1.1 million stars simulated over a 5 year observation period. It can be seen that convergence was achieved within 3 iterations in this (somewhat artificial, see below) test case.

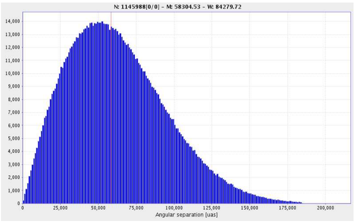
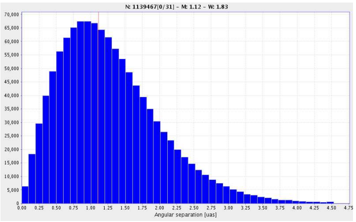
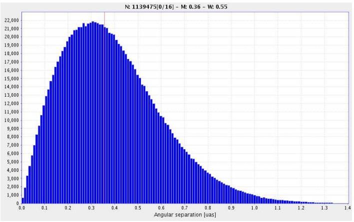
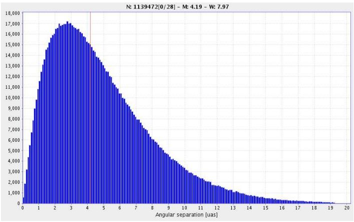
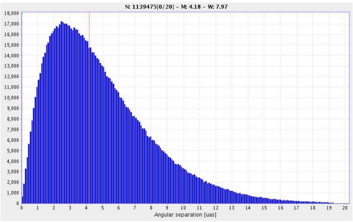
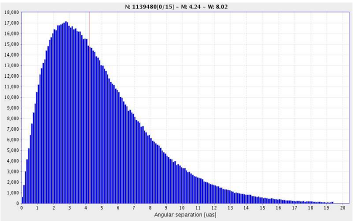
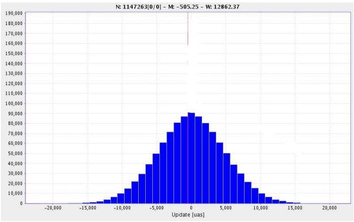
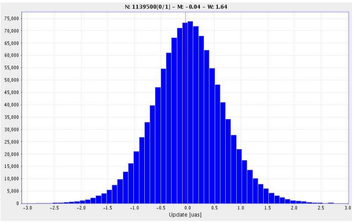
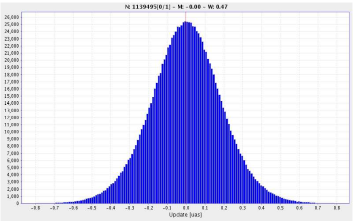
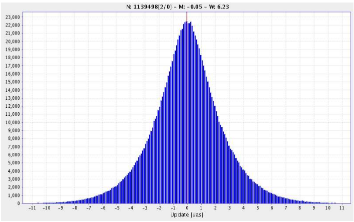
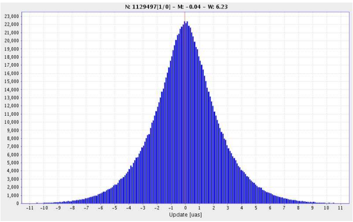
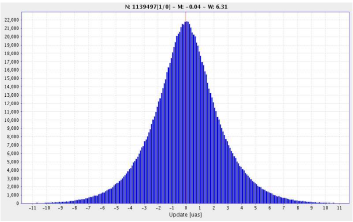
Iteration 1	**Iteration 2**	**Iteration 3**
Position update		
		
Position error		
		
Parallax update		
		
Parallax error		
		

Figure 26: Position and parallax convergence histograms from the three iterations of AGIS cycle 452. The median (M) and width (W) of each distribution is given in the text field above each histogram - see text for further details.

In Fig. 26 four types of convergence histograms are shown for the three executed iterations 1–3:

1. Position update: Angular separation of unit vector to corresponding star in previous iteration

2. Position error: Deviation of calculated direction to star from true direction (known from the simulations)

3. Parallax update: Update of parallax value of star with respect to previous iteration
4. Parallax error: Absolute error of computed parallax with respect to true value (known from the simulations)

All attitude and calibration parameters were initially set to their true values while 50 mas Gaussian noise was added to the initial source positions. This is reflected in the position update histogram for iteration 1 which shows a median update of ≈ 58 mas. The position errors at the end of this first iteration have already settled at μas level, thus, the initial noise was completely eliminated in just one step. A similar pattern is present in the parallax values, i.e., large (mas) updates in the first iteration that rapidly reduce the errors to the expected levels. In the subsequent iterations 2 and 3 updates to parallaxes and positions drop to sub-μas level while the errors remain constant at a few μas. Essentially noise-free observations ('true' transit times and across-scan pixel coordinates) were simulated for this validation run.

The total run time of the system was measured at a few hours per GIS iteration on an adequately sized hardware platform. Extrapolating this result to the volume of data expected for the real mission indicates execution times of a few weeks on a suitable computer system in 2011 [Lam06a].

Over the coming years to launch the algorithmic and technical complexity of the system is expected to grow, reflecting a gradually improving realism and capability to deal with anticipated features in the real mission data. This shall be accompanied by corresponding testing campaigns with simulated data gradually increasing in volume and realism with respect to e.g. noise levels.

5.1.5 Astrometric solution of the secondary stars

At the end of the AGIS (Sect. 5.1.4), one has obtained an accurate and consistent set of parameters describing the attitude of the astrometric instrument and its geometric calibration. At the same time, the astrometric parameters of the $\sim$10% primary stars have also been determined as part of the AGIS. Astrometric parameters for the remaining $\sim$90% of the observed stars — called *secondary stars* — can now be determined in one single processing step by simply applying the known attitude, calibration and global parameters. Thus the astrometric solution for the secondary stars in principle amounts to a single run of the AGIS astrometric parameter adjustment process (step 1 in Sect. 5.1.4.1).

In a given outer AGIS iteration, a star (or other celestial source, e.g. a quasar, compact galaxy, compact planetary nebula, ...) may be treated as a secondary star (i.e. ignored) for several possible reasons. The most important ones are:

1. It may be too faint to effectively contribute to the AGIS (adding noise and computing effort rather than weight and stability to the system).

2. It may be very bright or display extreme colours (necessitating usage of less well-determined calibration parameters for its treatment), or show strong photometric variability (making its set of observations very inhomogeneous).
3. It may be situated in a dense star field in the Milky-Way band and have been de-selected as primary to avoid the danger of unbalanced weights between the two fields of view, as described in the last paragraph of Sect. 5.1.3.
4. It may have consistently shown signs of spatial extension (making centroiding a fuzzy concept) or of barely resolved duplicity (making special centroiding and analysis methods necessary).
5. In addition there will be stars that are apparently single and also 'normal' in all other relevant aspects, but with astrometric observations having shown excessive scatter around the single-star astrometric model in the previous outer AGIS iteration.

Several of these reasons may, of course, hold for the same star simultaneously. All of the reasons may become apparent at any time within the mission, e.g. signs of duplicity may be obvious from the very first images, or they may become significant only after the collection of hundreds of observations. Similarly, time variations of the proper motion due to a binary orbit may be quick and strong, or they may show up only after years of observations, e.g. in the case of long-period binaries.

Conversely, a reason for downgrading a star to secondary status may also disappear in the course of the Gaia mission duration. For instance a few outliers (caused by cosmic-ray events or chance superposition of faint star images from the other field of view) may cloud out the observation history of a star early in the mission, but may become recognizable as such in view of many other consistent observations collected later on.

Secondary stars of all groups above will be subjected to the secondary-star solution process. That is, a single-star model will be fitted to their data, exactly as for the primary stars in AGIS. After this initial step, the treatment of the different groups diverges.

If a star of group 1, 2 or 3 shows an acceptable scatter around the single-star solution, the case is considered closed, and the resulting parameters are simply written to the database. Otherwise the star is handed over to the Object Processing task (see Sect. 5.4.1). Members of group 3, if well-behaved, will be reconsidered for upgrading to primary stars if needed, i.e. if downgrading of neighbouring stars should have created a paucity of primaries in the sky area. Members of group 4 will be handed to the Object Processing task regardless of their success in the single-star model adjustment.

The members of group 5 fall into two broad categories: Stars really deviating from the astrometric single-star model, and stars having an unfortunate observation history. A clear distinction will not in all cases be possible. A member of group 5 may, contrary to the results of the previous outer AGIS iteration, now pass the consistency test with the single-star model. Such a case will be considered as closed, the results will be written to the database, and the star may be reconsidered for upgrading (as with group 3). Otherwise the star is handed to the Object Processing task for further analysis.

The precise sets of criteria and the levels of significance for the observed astrometric disturbances at which

- a star is downgraded to a secondary,
- a star is subjected to more detailed treatment by the Object Processing task
- a star is considered for upgrading to a primary

may be different, and in principle need not even be negotiated between the astrometric Core Processing and the Object Processing groups. The detailed schemes to be applied for these decisions are not yet defined. In any case they will be highly adaptive in view of

- the gradually accumulating information about the individual objects in the course of the mission, and
- the need to keep a suitable set of primary stars for the AGIS.

5.1.6 Update of the intermediate data

The main result of the treatment of raw data by the IDT process is the production of intermediate data, the higher-level image parameters for each observation, and its cross-matching with the existing source catalogue. After a daily run of IDT, the raw data are stored in the raw database. The resulting intermediate data are stored in the MDB , and are ready to enter AGIS as well as other reduction processes like photometry or radial velocity determinations.

To produce intermediate data the IDT makes use, besides of the invariant raw data, of current values for calibration, attitude, satellite orbit, and source colour, the latter being extracted from the photometric data. Thus, every time that the AGIS process produces new calibrations, refines attitude and source astrometry and other reduction processes provide information on the sources themselves (i.e. colours for the stars, radial velocities, etc), the whole set of intermediate data obtained since the beginning of the mission up to the current date can be improved. The process of treatment of the raw data using both the better calibration parameters and the better source parameters, from AGIS and from the general source updating, to improve the intermediate data has been named Intermediate Data Updating (IDU), and will run every six months. In Fig. 27 an overview of IDU is given.

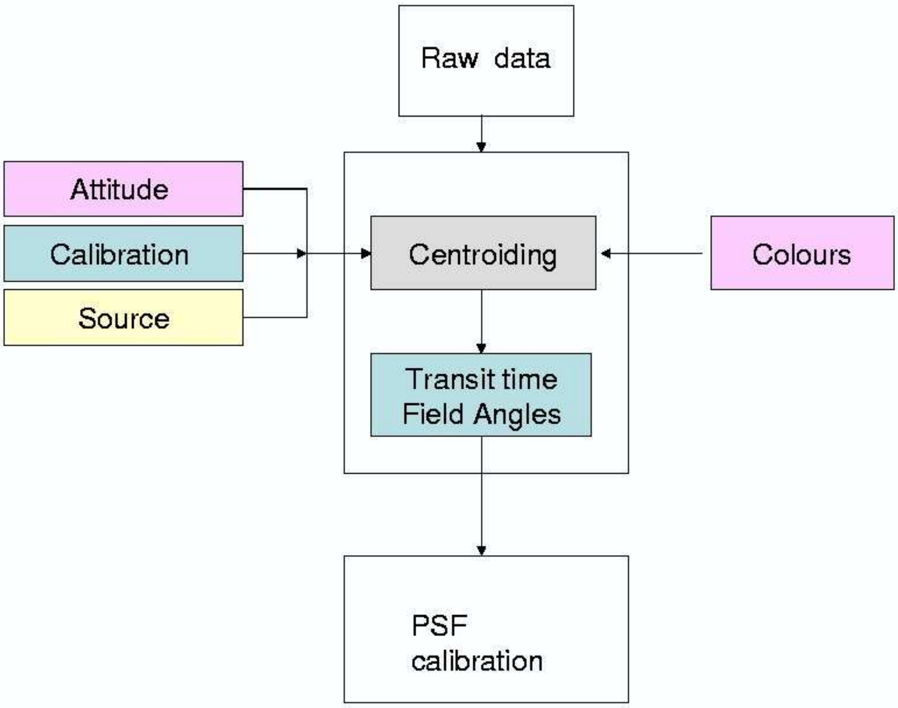


Figure 27: Overview of the IDU interfaces

IDU will repeat the IDT processes and in some respects it will be simpler than IDT, because some of the variables have already been fixed (i.e. the attitude is much better as a result of the AGIS process, and the attitude determination steps of IDT can be avoided), and other variables are better determined (i.e. colours, calibration data). In other respects it will be much more complicated and demanding, if all the information coming from the treatment of objects (photometry, radial velocity, binarity, others), and the results of the imaging process (i.e. disturbing sources in the vicinity and in the other FOV) has to be considered. In addition, some processes like cross-matching can be rerun with the better astrometric and photometric data, and better ephemerides of solar system objects and orbiting stars, resulting in new and more accurate associations of elementaries to sources. As more sources are introduced in the source catalogue, thanks to the imaging, more intermediate objects will become linked to multiple sources, thus adding to the complexity of the cross matching. For more details on the data flow Sect. 7.4.2.

While it remains clear that the IDU proper will redo the centroiding for the SM and AF transits, and will redo the cross matching between intermediate objects and sources, some questions remain open:

- will IDU treat in any way the raw data from BP, RP, or RVS?
- will IDU host other tasks needing access to raw data, like PSF and LSF calibra-

tion?

- will IDU centroiding include subtraction of disturbing signals from sources in the vicinity or other FOV?
- will IDU centroiding run more efficiently by time than by space?
  - if IDU works by time it should not host jobs working by space.
- in which process will the sky background be determined, and how?

For PSF and LSF calibration, stars of all magnitudes are needed, as effects like CTI depends strongly on the signal. For primary stars, high quality astrometric data and colours are available as a result of the source updating process and an improved centroiding is provided by IDU. Attitude and calibration data for the observing time of each transit are too, so these stars will be well suited for the PSF and LSF calibration. Knowledge of disturbing signals from the other FOV is needed for excluding such transits.

As for the questions listed above, there is no apparent reason why IDU would need the BP, RP or RVS samples, which would mean a large increase on the storage requirements. We tentatively assume that IDU will work by time, and could conveniently host the PSF and LSF calibration, but would not be able to host a process like the imaging. A main reason for working by time is to facilitate the process of subtraction of disturbing signals, which appears to be the more demanding part of IDU. The cross matching is decoupled from the raw data processing, and will work by space.

5.1.7 Large-scale data flow for the astrometric core solution

The overall data flow for the astrometric core solution is visualized in a very compact, high-level form in Fig. 28. All details that are not needed to illustrate the basic logic are omitted.

The basic inflow of data into the AGIS logically originates from the IDT and FL treatment in the form of pre-processed observations and provisional attitude and calibration parameters, plus the list of newly-identified sources (large blue cylinder at the top of the figure). Organisationally, however, it enters the AGIS in the form of a Main Database delivery every half year. Other major inputs are the initial Gaia source list, to which most of the observations have been cross-matched by the IDT, and auxiliary data from ESOC and other provenances. There are a lot of additional, minor inputs which for clarity of the figure are not shown.

In the first operational run of the AGIS, these data will cover about 6 months of mission. They are processed to produce refined, high-precision attitude, calibration and primary-star source parameters. The attitude and calibration data are immediately forwarded to the secondary-star solution process to derive astrometric source parameters for the secondary stars as well. The combined dataset is delivered to the MDB Integrator program which creates the first updated version of the Main Database from them (in combination with deliveries from other main data reduction chains).

As soon as all the main data reduction chains have integrated their outputs at the main database, the improved knowledge of the sky and instrument can be used to improve the original IDT processing. This is the IDU process at bottom centre of Fig. 28. It completes the updated version of the Main Database by revising the source list, the cross-matching table and finally the IDT's pre-processing of the observations.

The red arrows in Fig. 28 indicate the major part of the feedback into the iterations of the global Gaia data processing scheme. After about six months, the contents of the updated version of the Main Database flow back to the AGIS, to be processed again. In addition, newly arrived observations from these six months, pre-processed by the IDT and FL, are included. The initial Gaia source list is not used again, but replaced by the refined source list from the IDU.

This cyclic (or rather helical) scheme is repeated until the end of the operational phase of the Gaia spacecraft. Afterwards there will be a small number of additional iterations, without the influx of fresh data indicated at the top of Fig. 28. Technically, the last delivery of core processing products into the Main Database is the same as before. Nevertheless it is indicated separately (green arrows and elements in the lower left part of Fig. 28) to highlight its significance in the global scheme. The "Final Catalogue Integrator" will be a somewhat extended version of the MDB Integrator program used in previous iterations of the global Gaia data processing scheme. Technically, the final Gaia catalogue is another version of the Main Database.

5.1.8 Astrometric verification

The main motivation for a full verification of the astrometric experiment aboard Gaia stems from the very nature of the mission itself, which intends to provide, via self-calibration, absolute astrometric quantities with a targeted accuracy well beyond any possibility of external checks with independent observations[9]. Also, the intrinsic complexity and the iterative nature of Gaia data treatment are other critical factors. Besides, being able to understand the results at the highest level of accuracy possible will be especially important for science topics like evolutionary cosmology, experimental gravitation, extrasolar planets characterization, and materialization of the reference frame, as these are the most sensitive to small systematics. Indeed, those themes will be at the forefront of astrophysical and fundamental physics research for the next decade, and for the Gaia mission to make profound contributions in these

---

[9] Possible exceptions which might become available are: the Reference Grid expected to be established by the NASA mission SIM, and, although very limited in sky coverage, the very high accuracy (differential) measurements by PRIMA at the VLTI.

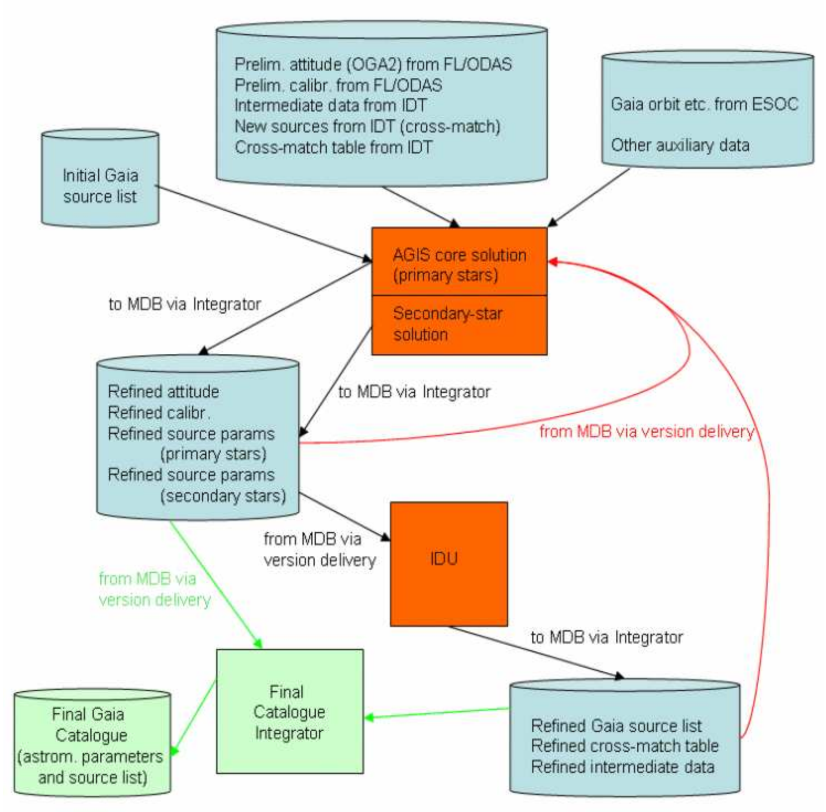


Figure 28: Large-cale data flow for the astrometric core solution. Only the basic logic is shown. For explanations see text.

areas its final astrometry will have to be of the ultimate quality.

Quantitatively, the improvement sought in the Gaia astrometry is a factor of 100 better than that achieved by the Hipparcos mission, thus making the implementation of two complete and independent data reduction consortia, as done for Hipparcos, highly recommended. However, this is not an option, as the scale of the Gaia data processing is just too large for any realistic expectation on material and human resources availability. As stated in [MBJ05], the Gaia data processing structure is "based on one data reduction consortium only, with no independent treatment duplicating the entire data analysis." In this context, the concept of verification of the Gaia astrometry must be a well-structured effort, focused on those data processing and analysis areas critical to mission success, and capable of gauging the degree of success throughout the mission. For these critical areas, independent procedures/models will be designed and implemented, and the results will be compared to the baseline processing pipelines.

The concept of astrometric verification is intended to be devoted primarily to the main[10] experiment on-board Gaia, i.e. verification of the processing of data from the astrometric focal plane and the consequent estimation of astrometric, attitude, instrumental and global parameters. The activities carried out within AV do not replace the validation of algorithms and implementations that must be carried out on all software modules, including those for AV, submitted for integration into the Gaia data reduction infrastructure. With this scope in mind, Astrometric Verification is considered as an intimate part of the Gaia data reduction scheme.

5.1.8.1 Methodology. The first task of the AV concept is a detailed review study of the overall astrometric baseline model. Validation efforts will then be focused on those data processing elements identified as being critical to achieving the absolute errors targeted by the astrometric processing. Next is the development of data reduction methods, independent to the degree deemed necessary, followed by direct comparisons of the output from both the baseline and independent processing, as well as with simulated data when these are available; the differences will be characterized and investigated when significant.

In the following we describe the application of the AV concept to the areas of the IDT and the AGIS data processing, as well as the role it has in monitoring the astrometric instrument, including the basic angle monitoring.

5.1.8.2 Sphere Solution. The sphere solution is the most critical step of the astrometric experiment aboard Gaia or, with a more general perspective, of what is usually referred to as *global space astrometry*. This is when all the measurements of a subset of astrometrically well-behaved, high-signal-to-noise objects are brought into a set of observation equations, linking together the different types of unknown parameters: astrometric, attitude, instrumental, and global. Although the astrometric parameters constitute the most relevant output for astronomy, all of the different parts of the solution of the sphere equations are equally important at this stage (e.g., the importance of γ). It is here that the Gaia rigid network of positions and proper motions is established, which will later become the most accurate materialization of the absolute celestial reference frame. And it is the derivation of attitude, instrumental, and global parameters done at this stage that allows propagating the astrometric solutions to the rest of the objects observed by Gaia [11].

Therefore, the verification of the pipeline sphere solution (AGIS) is of paramount importance. This is done by comparing the AGIS results to similar results from inde-

---

[10] The adjective “main” refers to the fact that the goal accuracy of the astrometric experiment goes way beyond anything currently done from the ground and in space, in contrast to the photometric and spectroscopic processing.

[11] Of particular importance is the propagation of the astrometric solution to the fainter extragalactic objects, which might take an important role in the orientation of the Gaia astrometric system into the absolute reference frame

pendent implementations of the sphere reconstruction, operating on the sample, or a significant sub sample of the same, of well-behaved stars used by AGIS. Initially, the AV global sphere reconstruction (indicated as GSR from now on) might simply be an independent implementation of the baseline model. This first version of GSR (GSR0) will serve as a starting point, providing a first verification of AGIS. Successive versions of GSR will progressively diverge from the baseline model, implementing a different relativistic model, data (signal) model, instrument model, and attitude model, as deemed necessary by the initial requirements study.

For those areas of AGIS where independent procedures are deemed unnecessary, the AV GSR will maintain close correspondence with the model as it is implemented in AGIS. At least one GSR model will be consolidated before launch to be used for verification during the post-launch phases.

A full-blown Schwarzschild model for Gaia-like global astrometry based on a non-perturbative approach was developed in 2001 ([dBLV01]). In 2003 a PPN version of that model was provided for the GDAAS-2 experiment ([VCdFL03]) and modified to demonstrate on realistically simulated mission data the potential of Gaia to determine the PPN deflection parameter γ to $\sim 10^{-7}$ ([VLB+03]). The rigorous relativistic model for the Gaia orbit and attitude was also introduced in 2003 ([BCd03]), and integration with the PPN code is straightforward. Thus, a relatively sophisticated GSR model is already available for the verification of the AGIS solution, which is based on the relativistic approach of Klioner (see for example [Kli03b] and references therein). The only limitation of this particular comparison would be accuracy. In fact, the Schwarzschild metric utilized in this GSR model would hold only if the contributions to the light propagation of solar-system bodies different from the Sun can be neglected, which of course is not the case. However, a comparison to an independent solution good to $10\mu as$ is certainly valuable, especially during the first 3 years into the mission, and this could be accomplished with such a GSR model by including only observations more than 10° away from Jupiter and 2° from Saturn. Comparisons at the $1\mu as$ level are anticipated for the final iterations. This requires a more sophisticated astrometric model for GSR. This has been recently developed ([dP06, Cro03]), following the new model for light propagation in the solar system developed in [dVC+06].

It should be noted that the verification achievable by independent processing at the sphere level can in principle go much deeper than a simple system-level check performed on the final solution (for example, using the “true” values provided by the simulations in the pre-launch phases). For instance, comparisons between GSR and AGIS could be performed after a single iteration, or even after the completion of one of the blocks making up AGIS; independent processing allows a detailed analysis to better identify the source of eventual problems, whereas a final system check can only indicate that a problem exists.

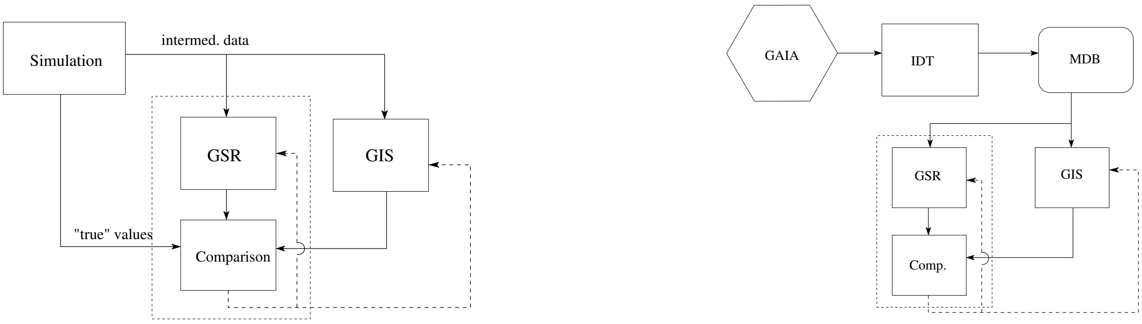


Figure 29: AGIS verification during the pre-launch (left diagram) and post-launch (right diagram) phases. Solid-line arrows represent data flows and dashed-line arrows indicate reporting channels. Simulated intermediate data may come directly from CU2, or from IDT processing of simulated telemetry. Processes in the dashed-line box are AV activities. (MDB = Main Database)

5.1.8.3 IDT processing. AV will consider those aspects of the IDT which are most critical for the astrometric error budget. The initial requirements study will identify the critical IDT steps based on their contribution to the error budget and, differently from the AGIS verification, confirmation will be performed *only* on these critical steps (i.e. a complete alternative IDT processing chain *will not* be developed). While the final list of possible items to be investigated is to be determined, we can anticipate the following to play a critical role: image parameters, PSF/LSF calibration, CCD calibration, and transit-level attitude diagnostics.

For illustration of the methodology, we briefly address one item of particular interest for IDT and for the consequences it has on the astrometric error budget, i.e., the treatment of bright objects, which will constitute the bulk of the well-behaved celestial reference points utilized in the core processing (sphere solution). Operationally the term ”bright” refers to objects of magnitude < 16 (12-pixel windows). The search for the best possible centroiding performance is of course of the utmost importance for these stars. On the other hand, saturation starts at $G \sim 13$ and will become severe for those objects in the brightest magnitude bin of interest to Gaia. One way of dealing with different saturation levels is the use of gates, i.e. to allow the loss of photons to make the brighter stars behave like the fainter ones, thus maintaining similar centroiding performances over the whole bright-magnitudes interval. On the other hand there is evidence that significantly better centroid errors can be achieved by dealing directly with saturated images, i.e. without (or very limited) actuation of the gates, when an accurate calibration of the actual in-flight PSF/LSF is made available ([GBGL05b], [Gar06]). Therefore, comparisons of the different techniques have the potential to benefit IDT processing greatly.

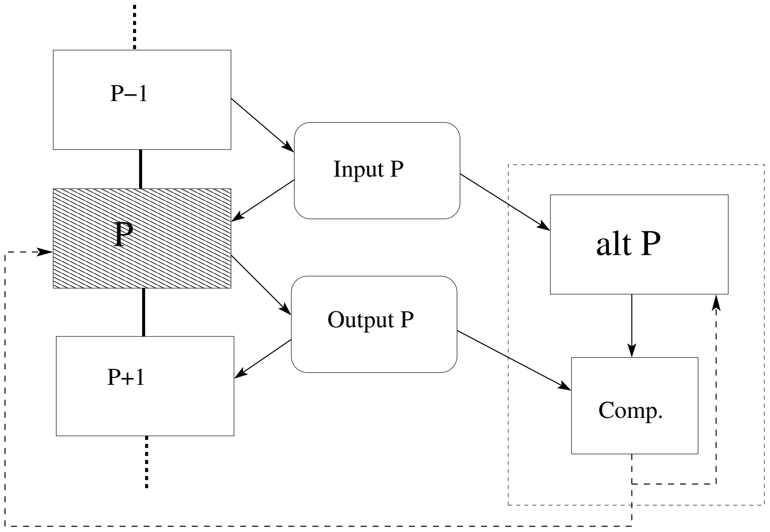


Figure 30: Verification of a single IDT process (labelled P) in the IDT processing chain. AV activities are limited to the processes inside the dashed-line box, namely the alternative process (alt P) and comparisons.

5.1.8.4 Verification. This is the part of the AV concept where the outputs from the baseline and independent processing are actually compared. Before any comparison can take place the independent solutions, e.g. the GSR solutions, will be internally validated through statistical techniques similar to those used to validate the AGIS solutions.
In the pre-launch phases, comparisons will also be performed with the "true" values provided by simulations, while checks with any external data with comparable accuracy, if such exists, will be attempted during the operational phase [12]. Also, an appropriate rotation to a common frame will always precede any direct comparisons of the different solutions.

Figures 29 and 30 show schematic diagrams for the AGIS and IDT verification. Specialized tools will be developed to analyze the differences between the baseline and alternative solutions, assessing their significance and characterizing any significant differences, when such exists. Internal statistical tests of residuals between observed and predicted positions (centroids) will also be developed. When possible, hypotheses as to the causes of any significant differences will be formed based on the aforementioned analysis, and a testing protocol will be outlined to identify the source of the discrepancies. Results of the comparisons will be disseminated to all relevant parties so that improvements to the components being evaluated can be discussed

---

[12] For example, VLTI measurements with the PRIMA instrument (50 μas) plus high accuracy radial velocities (better than 0.5 km/sec) of resolved SB2's can provide dynamical parallaxes of 10% accuracy to 1 kpc.

and implemented. Finally, the results from the independent processing will also be made available.

Ways to investigate results from different processing methods of essentially overlapping observational data (quasi-external comparisons) exist and new ones have been developed in the past ([BTL93]), and applied to the case of ground-based and spaceborn astrometric catalogs, including the case of the FAST and NDAC solutions from the Hipparcos mission ([KLF+95]). The very same methods, appropriately adjusted, can be applied to uncover significant deviations of the AV solutions from the pipeline ones.

5.1.8.5 Astrometric instrument model. This part of AV is devoted to the monitoring and diagnostics of the astrometric instrument response during in-flight operations. The goal is the identification of an efficient set of global (effective) parameters ([GB04])for representation of the instrument signature in the data, in spite of the probable degeneration of the real physical parameters. This should optimize the parameter estimation process within the GSR sphere solution with respect to computation load, accuracy, or both. The method is that of analyzing the impact on the data of perturbations to the instrument and operation parameters (optics [BGG+06], attitude, detection system), including ground-to-orbit variations, ageing and noise. Its implementation will make use of the instrument modeling tools developed for data simulation (section 6.3), and develop appropriate analysis tools. The astrometric instrument model will allow the investigation of optimal calibration and diagnostic procedures for their possible inclusion in the astrometric core data processing([GBGL05a])

5.1.8.6 Basic Angle materialization and evolution. This aspect of the AV concept is different in that there is no comparison of different solutions. Nevertheless it is critical for the verification of the absolute quality of Gaia's astrometry.

Hipparcos-like global astrometry requires a spinning satellite acquiring observations from two viewing directions imaged on a common focal plane. The angle between these two viewing directions is the basic angle (BA). For two telescopes implementing the two different viewing directions, the BA is the angle formed by the lines of sight of the optical systems endowing each individual telescope.
But why is the BA so important? The substance of the answer can be given in simple terms. The instantaneous equation that links the location of two celestial objects (1,2) in the two FoVs is

$$\xi_2 - \xi_1 + \varepsilon + \varepsilon_{BA} = s \cdot (y_2 - y_1) + BA;$$

where ξ_i, y_i are the along-scan position of object i respectively in the sky and on the image, ε_{BA} is the residual error due to BA fluctuations, ε represents the other contributions to the final error, and s is a function representing a general platescale. This equation can be rearranged to express the positions of the individual objects as

$$\xi_i \mp \frac{1}{2}(\varepsilon + \varepsilon_{BA}) = s \cdot y_i \mp \frac{1}{2}BA,$$

which clearly tells that the astrometry is directly affected by errors on the BA. During the study phase of the astrometric payload it has been shown ([GLM$^+$04]) that ε_{BA} cannot always be considered a purely stochastic error. Known and/or unexpected perturbations might make the elements of the optical configuration exceed the stringent limits set for their stability, causing the BA to fluctuate at different time scales. Therefore, the final astrometric accuracy could be severely affected if these variations are not monitored and, if necessary, removed from the data. The auto-calibration property of the Sphere Solution will only apply to time scales longer than the spin period. For higher frequencies we will have to make use of data from the BA monitoring device, which is capable of directly assessing the stability of the two lines of sight with a precision comparable to the final astrometric accuracy. Also, with those data AV will develop a long-term model for the BA fluctuations, which can then be compared with similar information obtained at the level of the sphere reconstruction.

5.2 Photometric processing

Gaia will produce broad-band and low-resolution dispersion photometry for all objects included in the astrometric survey. Those data have dual roles: to support the astrometric data reductions with, for example, essential information on colours to be used in chromaticity corrections, and to provide astrophysical information on the objects observed. The latter covers amongst others the distinction of non-stellar objects and the determination of astrophysical parameters for all objects observed.

Large-scale calibrations The mission will be self-calibrating for both sets of photometric data: as a survey mission this is the more reliable way to establish a well defined and internally distortion-free system of magnitudes. The procedure of self-calibration will involve iterations and differential calibrations between different CCDs. The iterations will be between the calibration of the photometric instrument-model parameters (for intervals of time that will have to be defined on the basis of the actual data), and the selection of standard stars from the calibrated data. This part of the calibration will take care of the large-scale, mainly optical, distortions in the images. A special procedure for initialisation of the iteration is to be designed, but exists already in outline. Specific complications in the iteration procedure are expected for the dispersion photometry, where the dispersion and wavelength scale will have to be calibrated differentially as a function of the ordinate of the transit.

Small-scale calibrations Differential CCD to CCD calibrations will also be part of the modelling. Calibration of small-scale features in the instrument model (typically down to the size of a window: several pixel columns), will be accomplished through examination of accumulated residuals over long (weeks to months) time intervals. This part of the calibration will compensate for local blemishes in individual CCDs.

The absolute calibration of the photometric systems will be obtained through comparisons, for specially selected stars, between the internally calibrated satellite data and spectro-photometric data, obtained through a dedicated ground-based observing program. The exact requirements on this selection, and its relation to final formal accuracies, are currently being assessed.

The photometric calibrations will involve only a small subsection of the entire photometric data stream. It is the intention to access the full data stream for the application of the calibration only when calibrated data are required. This will be done through the local IoA DPC database, which is likely to contain only references to flat files of raw data. These input files will remain unchanged through most of the data processing, they will only be renewed when the IDU is applied to the raw AF and SM data. The calibrated data are subsequently made available for further processing or distribution. One aspect of the further processing is the detection of variability and the related process of selecting constant stars for the calibration.

At least partially-separate calibrations will have to be done for observations using different gates. It will be important to establish sufficient overlap between calibrations for different gates to avoid systematic errors in the photometric catalogues in magnitude ranges. Overlaps should be provided naturally through the uncertainty for the on-board software in establishing the magnitude of a transit from the SM data. This implies uncertainty in the assignment of gates for stars with magnitudes close to one of the gate boundaries.

Distribution of data (the transfer to the central database at ESAC) may take place every 6 to 12 months, and will require the application of the latest models for both the differential and absolute calibration, to provide further users with fully calibrated data as based on our latest results. These releases will include fully calibrated photometry for the broad band data for further analysis. For variability analysis, epoch resolved data (at field-transit level) will also be prepared and released. The dispersion spectra may be released as mean, normalised and calibrated data and/or in the form of pseudo pass bands. Epoch photometry at CCD transit resolution will be preserved and made available at the end of the mission. The possibilities for analysis of these data are too limited to justify the large overhead that would be caused by half-yearly releases.

Special provisions have to be put in place to assess the potential disturbances of images. Maps will be created for all images as based on the transits through the SM CCDs and through the AF CCDs for which images with extended windows will be obtained. These maps will have to be interpreted automatically by a software package, to decide how a disturbance can be treated in the processing. In the ultimate case of the disturbance being too severe the data relevant data will be made available for object analysis.

The photometric analysis will take place on hardware provisions in Cambridge, but at this stage it is still too early to decide whether this will be dedicated hardware or shared, possibly grid-based hardware.

5.2.1 Photometric signal modelling

The white-light photometric signal in the *G*-band will be derived from the astrometric observations collected in the AF field. Hence the signal model for the *G*-band is the same as the astrometric signal model described in Sect. 5.1.2.1 (Eq. 7 and Eq. 8) and will not be repeated here. The rest of this section is focused on the signal modelling for the BP and RP data.

A proper design of the data processing for the BP/RP data requires a detailed understanding of the signal contained in dispersed images acquired with CCDs operated in TDI mode. An important property of data obtained in TDI mode is that the charge in each *data* pixel is produced by all the *physical* pixels along one of the CCD columns. Thus the data pixels are very different from the physical pixels on the CCD. A consequence of the dispersed image formation in TDI mode is that the dispersion curve for the prisms will not be observed as some relation between wavelength and location in the focal plane array. The dispersion curve can only be defined in terms of a relation between wavelength and the offset in the data space with respect to some reference point in the dispersed image. The following mathematical description of the BP/RP image formation process is therefore given completely in terms of the data pixels.

The AL and AC data pixel coordinates (or data pixel running numbers) are indicated as k and m. The corresponding *continuous* coordinates in the data space are indicated with (κ,μ). The location in the data space of some reference point (defined by a reference wavelength λ_0) in the dispersed image is given by (κ_0,μ_0). This location can be computed if the complete geometrical calibration of the instrument, the instantaneous scan rate, and the centroid of the corresponding SM image are all precisely known.

The dispersion curve of the prisms can now be defined with respect to this reference point and is given by:

$$\kappa-\kappa_0=\kappa_p(\lambda,\kappa_0,\mu_0) \quad \text{and} \quad \mu-\mu_0=\mu_p(\lambda,\kappa_0,\mu_0)\,. \tag{15}$$

These equations give the offsets $\kappa-\kappa_0$ and $\mu-\mu_0$ as a function of wavelength. The equation for the 2D dispersed image $\mathscr{I}(\kappa,\mu)$ can now be given:

$$\mathscr{I}(\kappa,\mu)=F\tau\int N(\lambda)T_0(\lambda)T_p(\lambda)T_f(\lambda)Q(\lambda)P_\lambda(\kappa-[\kappa_p+\kappa_0],\mu-[\mu_p+\mu_0])d\lambda\,, \tag{16}$$

where the different quantities in the equation are: the telescope pupil area F, the integration time per CCD τ, the source SED $N(\lambda)$ in units of photons s^{-1} m^{-2} nm^{-1}, the telescope (mirror) transmittance $T_0(\lambda)$, the prism (fused silica) transmittance $T_p(\lambda)$, the prism filter coating transmittance $T_f(\lambda)$, the CCD quantum efficiency $Q(\lambda)$, and the monochromatic PSF P_λ. The latter is centred on the point $(\kappa_p + \kappa_0, \mu_p + \mu_0)$. Implicit in the equation are: 1) the dependence of the monochromatic PSFs on the position in the focal plane; 2) the variation of the dispersion curve across the focal plane; and 3) the wavelength dependent effects of the geometrical calibration. The latter leads to an additional variation of dispersion curve, depending on the across-scan position of the image in the focal plane. The PSFs in Eq. 16 are taken to be monochromatic effective PSFs, which include the optical diffraction and aberrations, pixel size integration, TDI smearing, charge diffusion, etc. $\mathscr{I}(\kappa,\mu)$ thus represents a continuous version of the integrated detected image *before* it is sampled by the detector pixel grid.

The equations above make explicit that the dispersion curves themselves depend on the position in the focal plane where the image was recorded and that the dispersion direction may not be aligned perfectly with the AL direction (i.e., the function $\mu_p(\lambda, \kappa_0, \mu_0)$ is not necessarily constant). This misalignment can occur if the prisms are not perfectly aligned with the CCD pixel grid. Another type of misalignment can occur due to the optical field not being perfectly aligned with the pixel grids (for example an M1 mirror is rotated about the axis perpendicular to its surface). This type of misalignment is included in the PSF and leads to images tilted with respect to the AL direction.

The actual *data* that will be acquired is the sampled (discretized) version I of $\mathscr{I}$ limited to a window containing most of the flux:

$$\begin{aligned} I(k,m) = \int \mathscr{I}(\kappa,\mu)\delta(\kappa-\kappa_0-k, \mu-\mu_0-m)d\kappa d\mu + \varepsilon_I(k,m) \\ k=0,\ldots,K-1; \quad m=0,\ldots,M-1 \end{aligned} \tag{17}$$

The array I consists of $K \times M$ AL×AC pixels and the coordinates κ_0 and μ_0 again indicate the location of some reference wavelength λ_0 in the data space. The function $\delta(x)$ in the equation above represents the impulse or Dirac delta-function. Thus Eq. 17 represents the discrete sampling of the continuous function $\mathscr{I}$ at the points $\kappa = \kappa_0, \kappa_0+1, \kappa_0+2, \ldots, \kappa_0+K-1$ and $\mu = \mu_0, \mu_0+1, \mu_0+2, \ldots, \mu_0+M-1$. The phase of the sampling with respect to the pixel grid is included in κ_0 and μ_0. In general the 2D dispersed image will not be measured. Instead a 1D vector S of photon counts is measured by a summation of I in the AC direction:

$$S(k) = \sum_m \int \mathscr{I}(\kappa,\mu)\delta(\kappa-\kappa_0-k, \mu-\mu_0-m)d\kappa d\mu + \varepsilon_S(k) \qquad k=0,\ldots,K-1\,, \tag{18}$$

where the vector S consists of K samples (each sample consisting of $1 \times M$ pixels). The summation is performed over a sufficient number of AC pixels (12) so that S contains most of the flux in $\mathscr{I}$. Note that the samples in S consist of electronically binned pixels (i.e. summed before read-out), therefore one cannot write $S = \sum_m I$

which would be correct for numerical binning after read-out. The quantities ε indicate the noise added to the data due to the measurement process. Alternatively, one can think of S as representing an incomplete measurement of a one-dimensional dispersed image $\mathscr{S}$ which is built from wavelength dependent LSFs. One can then write for the one-dimensional dispersed image $\mathscr{S}$:

$$\mathscr{S}(\kappa) = F\tau \int N(\lambda)T_0(\lambda)T_p(\lambda)T_f(\lambda)Q(\lambda)L_\lambda(\kappa - [\kappa_p + \kappa_0])d\lambda\,, \tag{19}$$

where L_λ is the LSF. Note however that one cannot view S as a sampled version of $\mathscr{S}$. This is because the LSF is obtained from the PSF by integration from $\mu = -\infty$ to $\mu = +\infty$, whereas the summation over $\mathscr{I}$ in Eq. 18 is only over a limited range in μ.

Although the equations above look straightforward the actual signal modelling will have to include many of the details that were left implicit. These include the following effects: the variations of the dispersion curves due to the prism mountings, wavelength dependent geometrical calibration changes, and the evolution of the prism material in the L2 environment; the evolution of the various response curves in time and their variation across the focal plane (notably the CCD QE curves, and the pixel response variations); variations of the PSF across the focal plane; the effects of charge transfer inefficiency (here included in the PSF); the effects of gates and CCD non-linearity for bright sources; the overlapping of the two fields of view; the presence of cosmic rays. In addition the modelling of the data for crowded fields will be essential for the development of robust flux extraction algorithms. For more details on the BP/RP signal modelling see [Bro06a].

5.2.2 Photometric Calibration

Simplistically, a *Gaia* photometric observation is the result of the overall sensitivity of the instrument to the incident flux energy. Since the sensitivity changes with wavelength and with time and across the focal plane, all observations have to be reduced to a 'common' instrumental system. Complications arise because of the decentering of the sources within the window, both along and across scan, because of the across scan motion during the transit, and because of the CTI effects due to radiation damage. Once the 'common' instrumental system is established, it has to be tied to the physical world by providing the transformations to the absolute fluxes of energy.

Gaia will observe all sources (of all types) brighter than 20th magnitude crossing its field of views. Thus, to continuously monitor the instrument, the set of 'standard sources' should include all kinds of sources covering the whole range of magnitudes and densely distributed in the whole sky. It is not feasible to design a calibration procedure exclusively based on 'standard stars' with on-ground observations. Instead, the calibration can (and must) rely on the *Gaia* observations themselves complemented with 'few' on-ground observations. In this way, the calibration processing can be split into two main parts: an internal calibration using a large set of *Gaia* standard sources, and an external (absolute) calibration using a small set of already known standard stars.

The photometric calibration has to account for the G-band fluxes, the BP and RP spectra, bright and faint sources, and stellar and non-stellar sources. The case of BP and RP spectra includes the calibration of the wavelength scale, both in terms of the zero-point wavelength (geometry) and in terms of spectral dispersion. The proposed photometric model assumes that the calibration can be performed as small corrections to a current set of calibration parameters. This processing is necessarily iterative, as it is also for attitude and astrometric calibrations.

The main input data to the internal photometric calibration stream consists of:

- elementary data (pre-treated "raw" data by the IDT/IDU chain, see Sect. 4.1) with:
    1. identification of the FoV, CCD-row, CCD-strip and pixel column of the observation, as well as the gate used, and time of last injection of charge, if any,
    2. for G-band: flux and time (by fitting a LSF/PSF using colour information) assumed to be sky background & bias subtracted,
    3. for BP/RP: 1D or 2D discrete spectrum (raw data) and time,
    4. flags resulting from IDT/IDU chain, and
    5. flags coming from detection/confirmation on-board, if any.
- position and motion of the source within the window, derived from astrometric parameters, geometric calibration and attitude of the satellite for the time of the observation (to account for the flux loss due to the non-infinite size of the window),
- photometry statistics of every observed source from previous observations, allowing to select the internal standard sources, and
- the current set of geometric and calibration parameters (nominal or the last determined).

The output data consists in a new set of calibration parameters providing the relative sensitivity (for G-band and BP/RP spectra) and the wavelength scale and zero point (for BP/RP spectra) of every unit with respect to the instrumental system.

The ratio of the observed and predicted flux f could be expressed in a functional form as:

$$\frac{f^{\rm obs}}{f^{\rm pred}} = A(t, N(\lambda)) \cdot B(t, x, y) \cdot C(t, \dot{x}) \cdot D(t, \alpha, \beta, \ldots) \cdots \tag{20}$$

where A,B,C,D account for the relative sensitivity, flux loss due to de-centering, flux loss due to across scan motion, flux loss due to CTI effects, and so on. This functional form could be solved per calibration unit (defined as every entity with no variation of the calibration parameters and consisting in combinations of time intervals and sets of pixel columns, CCD rows, FoVs) using all available observations of the internal standard stars in that calibration unit. Large- and small-spatial scale units can be considered combined with short and long-temporal scale units, respectively.

The internal standard sources have to be well distributed on the sky (to ensure having enough observations to account for spatial and temporal effects), well isolated (to avoid contamination by close companions), and representative of the whole magnitude and colour ranges. For relative sensitivity calibration, the internal standard sources have to be non-variable sources (this non-variability needs to be defined for each kind of objects; for instance, cool stars are spoted and show variability and they cannot be all rejected because it would produce biases on the calibration parameters), while for the wavelength calibration they have to show identifiable spectral features (H_β line, H_α line, emission lines of QSOs, emission lines of WR stars, and so on).

During the five effective years of mission, every source will be observed 80 times on average, yielding a total of $6{\cdot}10^6$ observations per day and per CCD row and half of that per each field-of-view. Assuming that only 10% of the observed sources are suitable as internal standard sources, about $3{\cdot}10^5$ transits per day per CCD and per FoV and 320 transits per day per CCD and per pixel column (both FoVs) are available.

To derive the transformation of fluxes in the instrumental system (in counts) into physical flux units, a grid of best suited spectrophotometric standard stars (SPSS) are required. These stars will provide the zero-point for the absolute photometric calibration of Gaia data [BBF$^+$06].

The SPSS should be as featureless as possible (for instance white dwarfs of different kinds, hot subdwarfs and very metal-poor stars). They must have accurately flux-calibrated spectra obtained from ground-based observations, covering the whole Gaia wavelength range (330–1050 nm). They must be sufficiently bright to be observed by Gaia at S/N larger than $\sim$100 per sample over the whole wavelength range. They must lie in uncrowded spots of sky. They should possibly cover a range of spectral types so that the maximum flux output occurs at different wavelengths across the entire range of the Gaia photometers, thus ensuring the best possible S/N, and hence accuracy in the calibration.

For a discussion of the existing libraries of SPSS, the criteria for selecting primary and secondary SPSS, the observing strategies and observing facilities see [BBF$^+$06]. It is being investigated if ground-based broad band photometric observations of stars of different spectral types and a wide range of magnitudes can be used to check the coherence of internal calibration at the faint end and when gates are activated.

5.2.3 Photometric processing

Here we present, in the form of a flow-chart (Fig. 31), the principle elements of the photometric processing. At the most basic level, nearly the same elements apply to both the *G* band and the dispersion-spectra photometry.

The photometric processing is based on the principle of internal calibration. This implies an iterative process between the actual photometric calibration (determining a set of instrument parameters describing the relations between observed intensities anywhere in the focal plane and an internally defined set of reference fluxes), and establishing reference fluxes for the calibration stars used in the modelling.

The basic elements of the data processing are presented here in the form of a flow diagram. In this diagram, all rectangular boxes represent major processes. Each box is accompanied by an acronym, representing the country or institute responsible (UK: Cambridge, Leicester or Edinburgh; UB: Barcelona; INAF-OABO: Bologna). Light-blue boxes concern processes to be implemented at the IoA DPC, orange boxes will be part of IDT or FL processes, to be implemented at ESAC or Barcelona, the red box will be implemented in Bologna.

The skew boxes indicate data, the purple ones concern temporary data files, the blue ones concern permanent data, to be ingested into the IoA DPC data base. The iteration loop for the internal calibrations is indicated by the red arrows.

Variability detection will only be done based on the *G*-band photometry, all other processes apply to *G*, *RP* and *BP*, though exact implementations vary.

5.2.4 Variability detection, selection of standard sources

The main purpose of variability detection in the photometric processing will be the selection of non-variable sources, suitable to be used for calibration in the data processing. Variability detection will be done on the *G*-band photometry only, which will be significantly more accurate than the dispersion photometry, while at the same time, variations in magnitude tend to be much larger and easier to detect than variations in colour.

Variability detection requires accumulation of properties of the distribution of magnitudes or fluxes for each observed object. This information, in its most basic form, consists of the (weighted) sum of the fluxes, the (weighted) sum of the fluxes squared, the sum of the applied weights, and a counter for the number of observations added. This information is accumulated in a database at the IoA DPC as the application of the internal calibrations takes place. These accumulations have to be restarted every time a new set of calibrations is applied or when the intermediate data files have changed as a result of IDU. Updated values will be accumulated in memory before being added to the database.

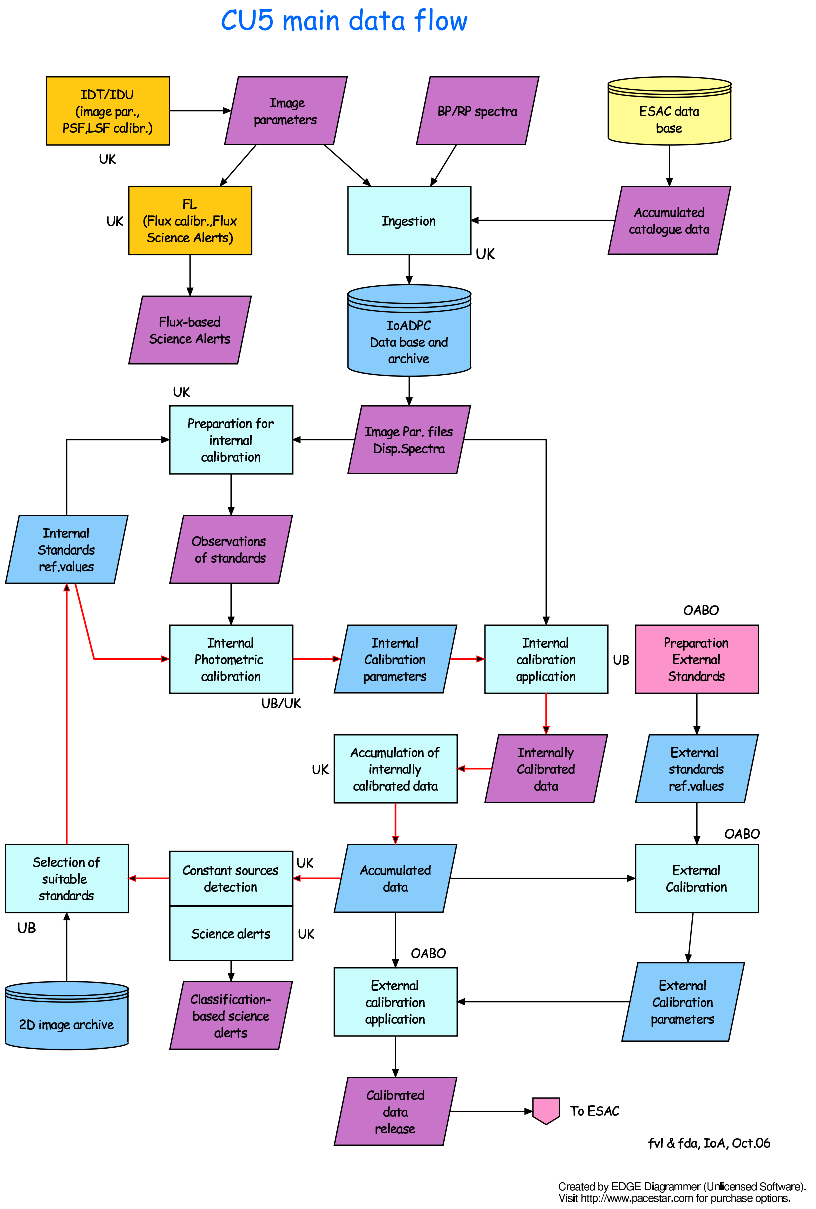


Figure 31: Diagram of the basic data flow in the photometric processing. The flow chart uses the following colour and shape coding: skewed boxes represent data files (purple boxes are temporary data files, blue boxes are persistent data stored in the database or in the data archive), rectangular boxes represent processes (colours and labels refer to the institute/country where each process should be run/developed). Database and data archive are represented using the same cylindric shape. The red line marks the main iteration loop in the photometric processing.

At any time the database can be interrogated for the statistical properties of the accumulated photometry: establishing noise characteristics of the accumulated data, defining the variability detection criteria based on those characteristics. Using those criteria, stars will be flagged as constant, possibly constant, or variable. This information will also be part of the photometric data release to the central database at ESAC.

Selection of standard stars for photometric calibration will require more than a criterion for non-variability: an even distribution over the sky and over spectral types will be required to ensure stable and reliable calibrations for all data. Extension towards very red objects is always a problem in such selections, as there tend to be very few of those objects, and those that exist tend to be variable. Some level of variability may have to be accepted for these stars. A special task will be to first establish requirements on the density of standard stars, and secondly to devise software for the preparation of a suitable selection of these stars from those objects identified as constant.

All of the aspects described above are severely complicated by the use of gates. To ensure connections between calibrations for different gates, potential standards should be selected that have magnitudes close to the gate boundaries, so that on occasion they may be used on either side of the relevant gate, and bridge the calibrations of adjacent gates.

The selection of standards for the dispersion spectra photometry has to incorporate criteria for suitability of stars of different spectral type, metallicity and surface gravity for this purpose. The needs for specific types of stars will also be different from the G-band calibrations and between the BP and RP detectors. The additional problem of the differential calibration of the wavelength scale may be partly solved through using line features in QSOs and other emission line sources.

5.2.5 Variability data processing

Variability processing will make use of photometric, spectro-photometric, spectroscopic and astrometric data. However most sources will be selected for a variability processing through their photometry in G band. Additional variability detection will be done with the spectro-photometry.

The number of sources which will be detected variable at the Gaia photometric precision is still uncertain but can be estimated between 5% to 10% of all the observed sources [Eye05]. The sampling is not very dense, about 80 measurements on 5 years, however it is quite irregular, which produces a spectral window with rather low peaks in comparison with other surveys (cf. [Eye06b]). This is why, for strictly periodic signals, Gaia will be able to recover the period with a very high success rate ([EM05]).

The methods which will be employed to perform the variability processing will range from very classical ones like least-squares fitting, to more novel ones likes Bayesian classification .

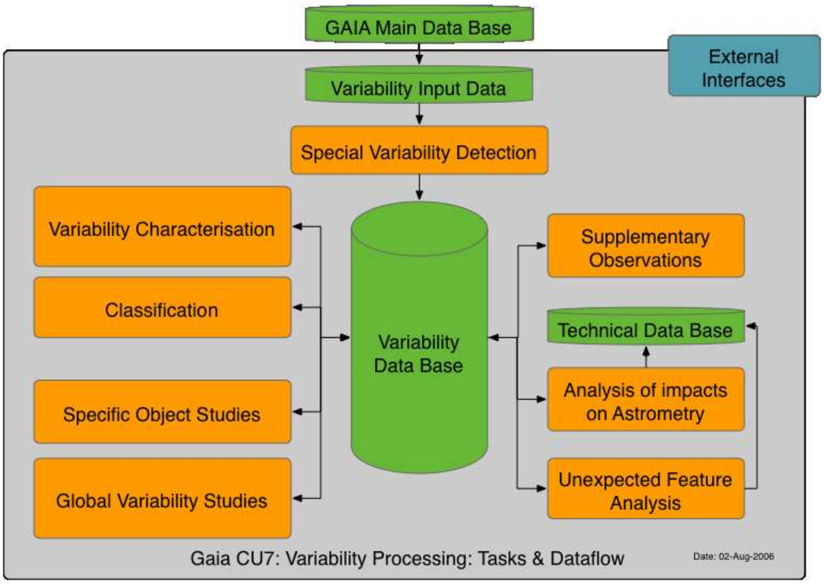


Figure 32: Schematic view of the top-level tasks and main data flow for the variability processing. The first task is Special Variability Detection and the following task order is counter-clockwise.

The purpose of the variability data processing is to populate the Gaia catalogue with information on the variable objects present in the Gaia database. Special care is taken to check the quality of the data and to validate some aspects of the calibrations of the photometry and spectroscopic data. The results from the CU7 variability processing can be used in subsequent cycles by the other CUs to improve the quality of the Gaia data processing.

Figure 32 shows a schematic view of the top-level variability related tasks together with the main data flows. The first task is “Special Variability Detection” and the following task order is counter-clockwise. The result of each task is an input for the next one.

The input data required for variability processing will be extracted from the so-called “Gaia Main Database” residing in the ESA ESAC centre located in Villafranca, Spain. Although, these inputs are a “To Be Defined” subset of the full data set, we expect rather copious amounts, with challenging constraints for both the data transfer and storage. There will be major upgrades of the input database at regular time intervals (currently planned for every 6 months), adding the newly acquired data, but also some older data if they have been re-processed in the meantime. Consequently, part of the data will have to be transferred many times.

The photometric processing is in charge of identifying constant stars and applying general-purpose algorithms for assessing the level of variation in flux, such as some statistical standard tests (e.g. χ^2 test). Special Variability Detection is defined to implement specific algorithms which take advantage of what we know about a particular type of variability and could be applied to all objects. All variable objects are then stored in the Variability Database.

Once variable objects are identified, their variability behaviour is characterized in the Variability Characterization task. The goal is to find the simplest possible description of the observed variations. Once they are identified and characterized, variables go through different Classification methods. The result is a number of groups containing objects with similar variability behaviours, and most probably, similar physical properties.

In the three first tasks, data for all objects are processed in a systematic way. In Specific Object Studies, specific algorithms are applied to objects of one of the groups resulting from the classification. For example, the processing required at this stage for the periodic periodic Cepheid stars is different from the one required for the usually rather erratic distant Active Galactic Nuclei.

After Specific Object Studies, all possible information about the variables have been derived from the Gaia data and are available in the Database. In the next task, Global Variability Studies, tools are needed first to derive a number of results required for the catalogue production. For example a colour-magnitude (or HR) diagram with iso-contour of variability amplitude as done in Hipparcos catalogue. Given the very large number of objects, some tools are also required to evaluate, to visualize in different ways, and to check the Database content.

There are three other tasks which are of very different nature Unexpected Feature Analysis, Analysis of Impacts on Astrometry and Supplementary Observations. They will be covered in more detail later in this document.

5.2.5.1 Special Variability Detection

In some cases, knowledge about a variability type can be used to improve the detection threshold. For example, although the star HD209458 was classified as a constant in the Hipparcos catalogue, a planetary companion was later discovered and soon after a planetary transit (planet passing in front of its companion star) was observed from the ground. Older, so-far unnoticed, transits were then found, looking back into the original Hipparcos data. Planetary transits have very peculiar signatures, the knowledge of which can be used to improve their detection efficiency. Other examples include the detections of low amplitude periodic variables, variables with very short periods, scintillation, solar-like variability etc.

This task is also clearly required to take a coherent approach to variability analysis. As we learn more about Gaia variables in the course of the detailed analyses, we need tools to search for variability throughout the complete input data set with new or improved criteria.

Special Variability Detection can be broken down into the following independent components.

- Planetary Transits
  This task includes the prospect of discovering extra-solar planets. The method usually used in other surveys like OGLE or HAT survey is the box least-square and will be applied for Gaia. Those objects that are detected as planetary transit candidates are flagged in the intermediate catalogue releases and final catalogue.

- Extremely Short Period Variables
  There are phenomena which can be of very short times scales (minutes) and could be detected by Gaia. Two such cases are orbiting binary white dwarfs and pulsating hot subdwarf EC14026 stars. In order to detect and study these types of stars, it is necessary to process the photometric values obtained from each CCD[13]. The integration time on one CCD is 4.4 sec and the same on the 9 successive CCDs. As shown in [Eye06a], a χ^2 test is not powerful enough to detect the slope that may appear in a transit over the 9 CCDs.

- Small amplitude periodic variables
  For periodic signals, as already mentioned, the rate of correct detection of the period is high even for signals with small amplitude with respect to the noise level, thanks to the peculiar sampling of Gaia.

- Star Scintillation
  The idea of detecting star scintillation due to interstellar matter has been proposed by Marc Moniez ([Mon03]) and is very new. This subject is therefore very exploratory. The specific signature left by such an effect is still being modelled, but would be at a very short time-scale. Thus, the time resolution of photometry per CCD is needed[13].

- Solar like-variability (magnetic activity)
  Stars with a magnetic activity detected for example in the RVS will be selected for a further detailed scrutiny of their photometry, and treated for the intermediate and final Gaia catalogues.

- Other Types of Special Variability Detection
  The tasks under special variability detection is not closed and therefore we are leaving open the addition of some new and interesting objects to be identified for the intermediate and final Gaia catalogues.

These tasks are meant to enhance the value of the catalogue. Special variability detection on all 1 billion objects can only be undertaken by DPAC.

---

[13] Transferring all CCD photometric values for all objects to the Geneva DPC represents a challenge. We are however currently studying in further details the Gaia capabilities for these types of objects. It may well be that these studies are only possible for a sub-set of stars (e.g. those brighter than a given magnitude) and thus the data transfer will not be an issue any more. If this is not the case, solutions will be worked out.

5.2.5.2 Variability Characterisation

The goal here is to find the simplest description of the observed variations. Part of the tasks were developed in the analysis of Hipparcos data [Eye98]. This task can be divided into three components:

- Statistical Parameter Determination
  This task is concerned with the determination of a number of statistical parameters, such as means, variances, asymmetry, kurtosis, decile of cumulative distributions, characteristic time-scales, etc.
- Period Search
  The goal is to identify all periodic objects and to derive their periods. Different methods will be developed. The different period search methods have different performance with respect to the type of signal. The methods used may be classified typically in three types: Fourier, string and PDM (Phase Dispersion Minimization) methods. Recently Mignard developed a very powerful procedure that will very probably be used in this task.
- Simple Model Fitting
  This task includes the fitting of (simple) models, such as a linear time drift or a Fourier series, which can be used to fully describe the variability behaviour in some cases.

5.2.5.3 Classification

The Classification task is done through different approaches which can be grouped into the three following categories. Over recent years there has been intensive development in this field. These studies include: with Bayesian classifier[EB02], [EB05b], with Self Organizing Maps [BWW04], [EB05a].

- Extractor
  An extractor processes the complete data set in order to extract all objects of a particular type. All known information about the variability behaviours of the object type is used to design the most efficient search algorithm.
- Supervised Global Classification
  This groups all methods which need a training set to tune some parameters of the algorithm. A training set includes a number of prototype representatives for each class, which are then used as models.
- Unsupervised Global Classification
  This refers to methods which do not use any a priori information. Although they may be less efficient in some respect, they are very important as they are the only methods which can lead to the discovery of new classes of object.

5.2.5.4 Specific Object Studies

The object types listed below may require some specific processing. The purpose of this Specific Object Studies processing is to provide data for the intermediate and final Gaia catalogues. In several cases further studies are needed to evaluate the corresponding amount of software development that will be required.

1. Opacity driven oscillators in main sequence stars (delta Scuti, gamma Dor, SPB, beta Cephei stars, . . .)
2. Rapidly Oscillating Ap stars
3. Cepheids/RR Lyrae stars
4. Long Period Variable stars (LPVs)
5. Solar-like oscillators
6. Compact oscillators
7. Pre-main sequence oscillators
8. Solar-like (magnetic-related) and rotation-induced variable stars
9. Flare stars
10. Eruptive stars
11. Cataclysmic variables
12. Eclipsing binaries
13. Rapid phases of stellar evolution
14. Optical counterparts of high energy-sources
15. Active Galactic Nuclei
16. Microlensing
17. Solar system objects

Because Gaia will achieve such a formidable step forward, new types of variable objects or events are also expected to be discovered.

5.2.5.5 Global Variability Studies

Three categories are identified as listed below.

- Variability Catalogue Visualization
  This task has to provide tools in order to browse quickly through the variability catalogue (e.g. to quickly select and visualise thousands or millions of variable objects either their time series or their variability properties) on one side, and on the other to get a synthetic view of the variability catalogue. Either for internal use or for outside usage. The latter would be subject to agreement from GST. Some or all parts of the software developed here could be delivered to the scientific community at large, in agreement with the GST and the PS. As an example of what was done with Hipparcos we may cite the work of [EG97].

- Variability Catalogue Quality Assessment
  The task aims to control the quality of the variability properties or classification present in the Gaia catalogue with a general approach. An example of such a method would be that of Protopapas et al. [PGF$^+$06]. The quality assessment on a global scale can also bring to light potential photometric and spectroscopic calibration problems.

- Survey comparison
  The purpose of this task is to provide the knowledge necessary to validate the variability results delivered in the Gaia catalogue and to allow an assessment of their quality. This task includes the following activities:
  - Follow up current and future potentially interesting surveys (OGLE , ASAS , Pan-Starrs , LSST , Kepler , Corot , Rave , SEGUE etc., also see `http://obswww.unige.ch/~eyer/VSWG/surveys.html`), copy the relevant data or make sure they are readily available.
  - Using the most appropriate and available multi-epoch survey, derive estimates of the numbers of variables expected in the Gaia data, and this for each of the different types of variables (see preliminary work [EC00] and `http://obswww.unige.ch/~eyer/VSWG/objexp.html`)
  - Compare the different variable type fractions from the above estimates with the results of the Galaxy model developed by CU2.
  - Identify and develop software tools required to compare results from the different surveys with the Gaia expected database results.
  - Operational tasks – perform the comparisons for validation and quality assessment.

5.2.5.6 Unexpected Feature Analysis

It may seem presumptuous to design an Unexpected Feature Analysis task, but many cases of “non-standard” behaviours can be foreseen, and tools can be developed in advance to analyse them. This task will however extend into the operation phase during which the analysis of unexpected features may reveal potential problems in the processing e.g. calibrations. In that case, feedback to the concerned groups will be carried out. It will most probably be required to develop additional tools during the operations to deal with “really unexpected” features. The Unexpected Feature Analysis deals with all measurements which are unexplained by our variability knowledge and statistical expectations. For example, it might be that one value is very far in terms of standard deviation from the mathematical model that describes all other measurements of a periodic variable. This value can only be singled out once the data have been folded around the derived period . Another example is a very slow time drift that may remain undiscovered on smaller time-scales. The detailed analysis of these features may either lead to discoveries of new, interesting properties or objects, or point to problems in the data or in the processing.

5.2.5.7 Analysis of impacts on Astrometry

For some of the objects, the astrometric solutions may be disturbed because of variable properties (e.g variable non uniform light distribution). In the Analysis of Impacts on Astrometry task, tools are developed to extract the relevant cases, derive the magnitude of the impact, and provide feedback to characterize them in the Gaia Main Database (and for the final astrometric catalogue).

5.2.5.8 Supplementary Observations

There are several cases where supplementary observations, either from observatories, or from other satellites, are needed in order

1. to prepare the data processing, or
2. to verify and validate the variability results in the Gaia catalogue.

The ground-based observation needs are evaluated and coordinated at a higher level within Gaia DPAC (cf. section 5.6).

Follow-up observations are left to the scientific community, to be performed after the intermediate or final releases of the Gaia catalogue.

5.3 Spectroscopic processing

The Radial Velocity Spectrometer (RVS) will collect the spectra of about 200-300 10^6 stars down to G_{RVS}=17, at 40 epochs on average (with 3 exposures per epoch, one per CCD along scan) for a total of about 25-35 billion spectra. The numerical processes that will be applied to those spectra can be divided in three broad categories:

- Calibration of the characteristics of the RVS. The RVS possesses no calibration device. Most of its characteristics (wavelength dispersion law, PSF profile, overall throughput, etc.) will be calibrated using its own observations in an iterative self-calibration process referred to as Spectroscopic Global Iterative Solution (SGIS). The RVS general calibration procedure is described in Sect. 5.3.1 and the SGIS is detailed in Sect. 5.3.1.2.
- Extraction, cleaning and calibration of the spectra. These processes are described in Sect. 5.3.2.
- Derivation of the characteristics of the sources, such as radial and rotational velocities, signatures of binarity, temperature, metallicity, etc. These processes are described in Sect. 5.3.3 and 5.5.6. It should be noted that, in term of DPAC organization, the responsibility of the derivation of these parameters is distributed over several groups (see section 8).

Figure 33 presents a schematic view of the overall spectroscopic processing tasks and data flow.

5.3.1 Calibration

The calibration of the RVS will be performed daily, will be refined every 6 months and will be fine-tuned a last time after the completion of the mission.

The daily calibrations of the RVS will be performed with the last 24h of observed data (transmitted from the Science Operation Centre to the Spectroscopic Data Processing Centre-CNES on a daily basis). They will provide a fast and regular monitoring of the variations of the properties of the RVS and will allow for a daily analysis of the data.

The half-yearly calibrations will be performed using the last 6 months of observed data. They will allow for refining the calibrations and for refining the analysis of the last 6-months of data. The half yearly calibrations will be iterative and will rely on the SGIS method (see Sect. 5.3.1.2).

The post-mission calibrations will use all the observations collected during the mission and will allow for a last improvement of the calibrations as well as of the astrophysical quantities derived from RVS data. The post-mission calibrations will also be iterative (based on the SGIS method - see Sect. 5.3.1.2).

5.3.1.1 Ground Calibration Many parameters of the RVS will be calibrated on ground prior to launch. These will be incorporated into calibration files which will form the basis for the initial in-orbit calibrations.

The detailed ground-based calibration plan will be defined through discussions between the DPAC and the project team (acting as interface with the industrial teams).

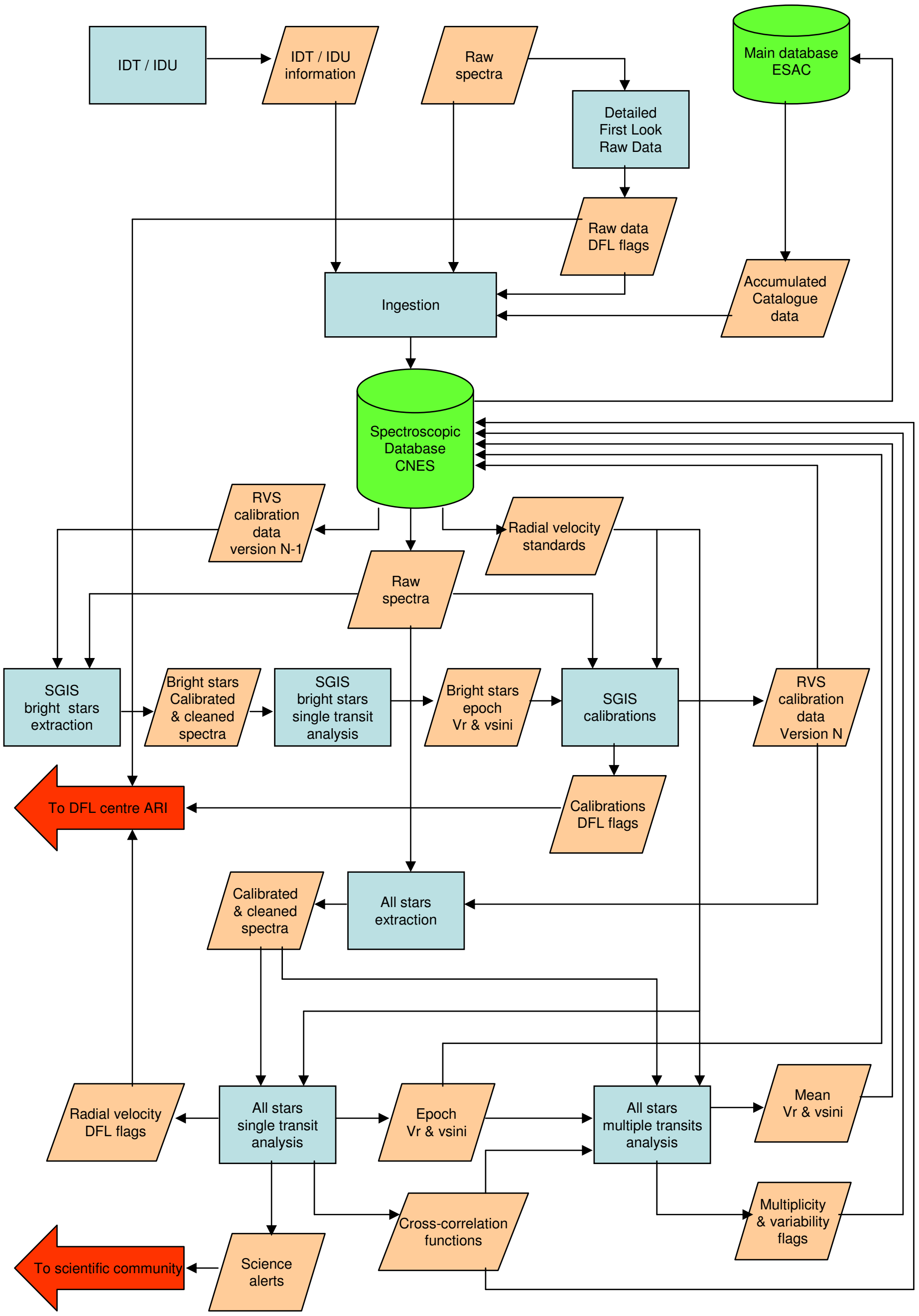


Figure 33: Schematic view of the overall spectroscopic processing tasks and data flow. Blue boxes represent processes. Skewed pink boxes represent data. Green cylinders represent databases. Red arrows represent interfaces with other processing centers or with the scientific community at large.

5.3.1.2 SGIS The RVS does not have internal calibration sources, and the instrument will be calibrated largely using the "Spectroscopic Global Iterative Solution" (SGIS) technique. This method is a transposition to the RVS of the astrometric "Global Iterative Solution" ([Lin01]). The principle of SGIS is to use RVS observations of (mostly) bright and stable sources to self-calibrate the instrument. Each of these stable sources will be observed a large number of times over the 5 years of the mission (40 times on average times three exposures). As the stars are stable, the variations of their measured properties will trace the evolution of the characteristics of the RVS.

The stable sources could be standard sources that have been identified as stable (astrometrically, photometrically and in radial velocity) from ground-based observations. However, most of the stable sources will be identified as stable from the Gaia observations them self. Similarly, the characteristics of the sources could be known from ground-based studies, but in most cases they will be determined from RVS observations.

The SGIS method will be iterative: to derive the characteristics of the sources, one needs "first" to calibrate the characteristics of the RVS and to apply those calibrations to the raw spectra, but to calibrate the RVS one needs to know the characteristics of the sources. Therefore, the 3 steps: (i) characterisation of the sources, (ii) identification of the stable sources and (iii) calibration of the RVS characteristics will be iterated. At iteration N:

1. The spectra of bright sources are calibrated using the (N-1)th calibrations of the characteristics of the RVS. Once the spectra are calibrated, the characteristics (e.g. radial velocity, ...) of the sources are determined.

2. Stable sources are identified by analysing, source by source, the dispersion of the measures of the characteristics of the source, at several epochs of observation.

3. The evolution of the characteristics[14] of the stable sources are analysed to refine the calibration of the characteristics of the RVS and the calibration files updated.

The iterative process is iterated until a satisfactory level of convergence is reached. The process can be initiated by using ground-based RVS calibration data and/or initial in-flight calibration obtained with the (small number) of RVS observations of ground-based standards. Once the SGIS iterative process converges, the measures are translated from the SGIS relative reference frame to an absolute reference frame using the ground-based standards.

The calibration files contain a model of the instrument characteristics, for example the wavelength dispersion law as a function of various payload temperatures, position in the focal plane *etc.*. This multidimensional model will be updated in the iterations 1–3 above. Work at Observatoire de Paris has shown that such an iterative approach converges, but, at the most accurate levels, care is required to ensure systematic biases are eliminated. The sample size used in the SGIS must be sufficient to

---

[14] their measures have been refined at step 1 above

populate the entire multi-dimensional calibration parameter space – sufficient positions on the focal plane at sufficient spread of operating temperatures with sufficient range of source brightness for the two resolution modes and different windowing schemes, and so on. The calibrations produced as a result of the SGIS include:

1. Photometric throughput, linearity, saturation level: these will be derived from photometrically stable stars with different intensities;
2. Across scan line spread function profiles: these can be derived from almost any of the SGIS-selected sources;
3. Along scan line spread function profiles: these will be determined from slowly rotating K-type stars, where possible giants;
4. Wavelength zero point and wavelength dispersion law: these will be derived from a sample of single mainly G and K stars.

The most problematic of these in early mission phases is likely to be the wavelength zero point and wavelength dispersion law, with the main dependence payload temperature. At later phases, radiation damage will affect the photometric throughput and the AL line spread function: these will need to be parameterised as a function of source brightness, background level and perhaps history of illumination prior to the readout of the particular window. These parameters are expected to evolve steadily as a function of time.

5.3.1.3 Other RVS Calibrations

In addition to the SGIS, other calibrations will be required for the RVS. These include the following:

1. CCD bias: this will be available directly from the overscan pixels;
2. CCD readout and dark noises: because there are no dark frames these will be determined indirectly from the noise characteristics of the data;
3. CCD flat fields and blemishes: in TDI mode pixel blemishes manifest themselves as column defects, and these will be mapped by reconstructing the CCD surface from many telemetered windows to generate flat fields (see ”Extraction” below) and isolating statistically low or high columns to identify blemishes;
4. Scattered light and ghosts: these maps will be updated from the ground calibrations as necessary by searching for the contaminating signatures from bright stars.

5.3.2 Extraction of spectra : single and multiple transit

The *Gaia*-RVS will produce spectral traces some 10 pixels AC and 1100 pixels AL. The on-board software will clock the CCD to read out the regions (windows) around the desired spectral information, and will add it in the AC direction to produce a 1-dimensional object which will then be transmitted to the ground.

From this, it would appear that the spectral extraction is a simple procedure, largely performed already on-board. However, this is not the case. Because of the length of the spectral traces, a significant fraction (depending on star density) of spectra will be overlapped. Particular means of dealing with this have been devised, with the nominal window selected for the brighter object, and a more complicated shape of window set up for the fainter object. Figure 34 gives some examples – and even more complex overlapping is possible. The details of how this scheme is implemented on board, and how many layers of overlap can be handled is still under discussion.

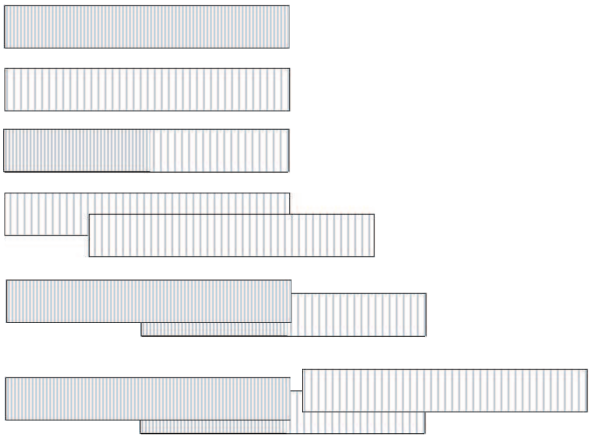


Figure 34: Some window arrangements for RVS. From top, high resolution window, low resolution window, mixed resolution window, overlapping low resolution windows, overlapping mixed resolution windows and multiple-overlap mixed resolution windows. Note that high resolution windows for very bright stars and a small subset of fainter stars used for calibration will not be collapsed in the AC direction, and the high resolution parts of mixed resolution windows will be binned in the AL direction to form uniform low resolution windows. The resolving power in high resolution mode is ~ 11000 while that in low resolution mode is ~ 5000.

Besides the spectra selected for telemetry, there will be fainter spectra from sources not selected contributing to the flux within the window. In some cases, particularly crowded fields, or where complex overlaps prevent the selection of a particular source, or where the source is extended slightly, these other spectra may not be that faint. The effects of these contaminating spectra have to be treated correctly to maintain the level of spurious radial velocities and other parameters derived from RVS to an acceptable level.

Consequently, the data from RVS have to be modelled in some way, most probably requiring re-constitution of the 2-D detector surface where there are overlaps, prior to extraction. Once the data are extracted, this processing stage also envisages the *application* of the (elsewhere derived) calibrations, with the output being calibrated spectra in standard units. The majority of the work will involve the propagation of error/uncertainty information and quality flags.

The extraction sequence will proceed broadly as follows:

1. The CCD bias and any photon flood or charge injection (to fill radiation traps) will be removed. These values will be available from the data itself (details TBD).

2. The cosmic (and if necessary instrumental) diffuse photon background will be subtracted. This background is currently baselined to be derived from the leading and trailing edge of the spectral windows, beyond the RVS bandpass (though other solutions are under discussion). Tests will be needed to select valid background: for example background from bright star windows will be invalid because of filter bandstop leakage, and overlapping windows will need to be treated as a special case, as the background region will contain flux from the overlapping star. The background surface will be interpolated as a function of AC position and time from these individual sample regions.

3. The background resulting from point sources and extended sources not selected by the on-board selection algorithm will need to be modelled, using information from the photometer and Astro, as well as catalogs and other resources. This modelling will involve applying the instrument calibration in reverse, to move from standard units to instrumental ones, as a function of TDI clock and AC position (and all other parameters of significance). At this stage any ghost images or scattered light from brighter sources will need to be introduced. The output of this modelling will be a predicted background surface which will be added to that from the diffuse background. Some reconciliation will be required, for example where ghost images fall in the measured background samples. The details of this still need investigation.

4. The total background will then be reconstructed from adding the diffuse and point sources as appropriate for each window and subtracted.

5. For those windows where there are overlaps, the spectra will need to be deblended. Perhaps a simpler procedure is possible, but at present it is considered that this will most likely require a forward algorithm using model spectra passed backwards through the calibration to predict the spectral and spatial mixing. There will be an iterative adjustment (for example via χ^2 fitting) of parameters to match the observed traces. 2-D images will be produced, from which the telemetered windows and thus predicted 1-D spectral traces can be

> derived. Some parameters, such as the relative positions and (at least approximate) spectral type of the blended spectra will be obtained from Astro and from the photometry, but others, such as the radial velocity will need to be derived from the fitting itself. Once a good match is obtained to the observed data, the degree of mixing is known, and can be subtracted from each of the traces in the overlapped windows to produce the de-blended traces.

Such a sequence will be applied to the individual spectra from each CCD transit. Multiple transits will not be combined in the image plane as this will effectively lead to increased crowding resulting from the different orientations on the sky of the individual scans. However, there will be particular scans which for any particular object are more free from overlap than others are, and a scheme may be envisaged where the information extracted from those scans are used preferentially in the modelling necessary for those scans where spectra are overlapped. This may not be possible for the faintest stars. Further investigation into these possibilities is necessary.

Now the remaining steps are to *apply* the calibration (derived from elsewhere) to the data. In particular this involves transforming from detected photons to fluxes in standard units, the application of the wavelength scale, the removal of cosmetic defects known to occur in each CCD *etc.*

Cosmic rays will be a significant contaminant of the RVS spectra, especially as the data are collapsed in the AC direction. A cosmic ray rejection routine will be applied to the three spectra from each CCD in a focal plane transit. This will identify outliers produced by cosmic rays at some level of significance. The precise nature of the cosmic ray identification and removal scheme will require considerable prototyping as it depends sensitively on the noise characteristics of the data. It should be possible to extend the scheme to lower levels of significance by comparing individual spectra to a mission-averaged spectrum for a particular source.

Finally a spectrum normalised to the local continuum will be generated.

5.3.3 Radial Velocity determination

5.3.3.1 “Classical” Cross-correlation

The principle of the derivation of the radial velocity by cross-correlation is to shift a reference spectrum in radial velocity step by step and at each step to quantify how well the locations of the lines of the reference spectrum match the locations of the lines of the observed spectrum. If the match is “good”, the radial velocity shift applied to the template is close to the radial velocity of the observed source. The quality of the match between the locations of the lines of the two spectra is quantified by the cross-correlation coefficient of the two spectra. The series of cross-correlation coefficients (corresponding to the successive shifts in radial velocity of the reference template) are usually referred to as correlation function or correlation peak.

For the method to work properly, the reference spectrum should display a morphology similar to the one of the observed spectrum: i.e. corresponds to a similar spectral type.

The derivation of the epoch radial velocities is generally split in the successive steps below:

1. (i) Select the synthetic reference spectrum with atmospheric parameters similar to the atmospheric parameters of the observed source.
2. Compute the correlation function: i.e. step by step:
   (a) Shift the synthetic reference spectrum in radial velocity.
   (b) Resample the synthetic reference spectrum to the sampling of the observed spectrum.
   (c) Compute the correlation coefficient CC:

$$CC = \frac{\sum_{s=1}^{Ns}(F_{obj}(s) - \overline{F_{obj}})(F_{ref}(s) - \overline{F_{ref}})}{\sqrt{\sum_{s=1}^{Ns}(F_{obj}(s) - \overline{F_{obj}})^2}\sqrt{\sum_{s=1}^{Ns}(F_{ref}(s) - \overline{F_{ref}})^2} \tag{21}$$

   where s is the running number of the samples in the spectra, Ns is the number of samples per object spectrum, $F_{obj}(s)$ and $F_{ref}(s)$ respectively the fluxes of the s^{th} samples of the object spectrum and of the “resampled” reference spectrum and $\overline{F_{obj}}$ and $\overline{F_{ref}}$ respectively the mean fluxes of the object and reference spectra.

3. Derive the epoch radial velocity of the source by solving for the maximum of the correlation peak.

5.3.3.2 Complementary method: Skew-analysis method

In many cases the the visibility of the spectral features in RVS spectra which can provide the radial velocity measurements are weak relative to the overall spectral flux at that point or are otherwise hard to discern. Reasons include a) faint stars where signal to noise is low, b) Spectral lines which are intrinsically weak (e.g. the ionization state is not highly populated or the elemental abundance is low) and c) unresolved binaries where the spectrum is dominated by one component with the lines from the other being highly diluted.

One technique that can be employed in such circumstances (and of course to cases where the lines are prominent) to determine the radial velocity is skew mapping ([VSHM03]).

Before skew mapping can be applied, the time-separated spectra from individual transit epochs must be individually extracted, calibrated and then normalised based on the local 'line-free' regions. The resulting spectra are cross-correlated (see above section on cross-correlation) against one or more template spectra that are closely representative of the target object, generating a cross-correlation function for each template used. The choice of correlation template can be based on information derived from the photometric or spectral information from the photometric instrument or, in principle, from other knowledge.

The result of the cross-correlations (for a given template) is a series of correlation functions in velocity space, one per transit epoch. These functions are then arranged in a chronological order. In each correlation function there will be one or more peaks at velocity offsets representing potential coincidences of the velocity-shifted template, the higher peaks generally reflecting better matches.

From this 3-dimensional skew map (cross-correlation strength (z axis) against time (y axis) and radial velocity (x axis)), $C(v,t)$, skew mapping attempts to isolate the most likely time-dependent radial velocity pattern by maximizing the line integral, $S(\gamma,K)$, along many model radial velocity paths through the map.

$$S(\gamma,K) = \int C(v,t)dl \tag{22}$$

where l is the model radial velocity path. A radial velocity model of the form

$$V(\gamma,K) = \gamma + K\sin[2\pi(t-t_0)/P] \tag{23}$$

can be adopted, assuming, initially that the spectrum is that of a binary in circular motion. Here γ is the systemic velocity of the system, K is the velocity amplitude, t is the epoch of the spectrum, t_0, the reference time (e.g. the time when the velocity crosses from blue to red-shifted motion) and P is the binary period. In the case where t_0 and P are known (i.e. the binary nature is certain and there is other information on these parameters), computing $S(\gamma,K)$ for various γ and K allows one to construct a map in which the function will peak at the point corresponding to the best model parameters

In the case where binarity is not assumed, one could extend this method to allow both t_0 and P to be parameters, constrained by the observational limitations (i.e. temporal baseline). The problem then becomes 4 dimensional, but still tractable in principle. In the case of single stars, one can expect $K \leq \sigma(\gamma)$, i.e. consistent with zero amplitude. The value of γ reflects the mean radial velocity motion of the system.

The analysis therefore reduces to identifying the peaks in the skew map and extracting the parameter values, according to a likelihood/significance threshold.

Parameter uncertainties can be deduced from Monte Carlo simulations.

There are a number of issues relating to the application of the technique outlined. These include

1. the impact of the analysis and uncertainty estimation on run-time.
2. the model velocity curves to be used in multiple-star systems and in those where significant elliptical motion might be present.
3. identifying the appropriate template spectra to use.
4. Recognition of a weak but true solution amongst low significance peaks in the skew map.

5.4 Other sources: Combined processing

5.4.1 Double and multiple stars and extra-solar planets

Non-single stars (NSS, star+star(s) or star+planet(s)) are important for three reasons:

1. they provide the mass (and sometime radius) of the components, ingredients which remain unknown otherwise and are nevertheless essential in any astrophysical modelling;
2. discovering an extrasolar planet or, even better, a number of them, improves the statistics to infer on the conditions of their formation and still has a real impact among the general public;
3. in the specific case of Gaia, where self calibration is a key issue, binaries will be troublesome and must therefore be removed before the calibration takes place.

From the third item, if follows that the detection of non-single stars will essentially rely on the identification of ill-behaved objects in any reduction pipeline which assumes genuine single stars. This yields three classes of NSS:

- astrometry-based: any object whose successive positions during the mission lifetime cannot be modelled with the basic five astrometric parameters (position, parallax and proper motion);
- photometry-based: any object whose point spread function or colours cannot be fitted with that of a single star;
- RVS-based: any object whose radial velocity variability exceeds the typical measurement scatter of an object of that spectral type and apparent brightness.

These three classes require further dedicated processing in order to make the default information relevant 35. For instance, whereas the average of all radial velocities

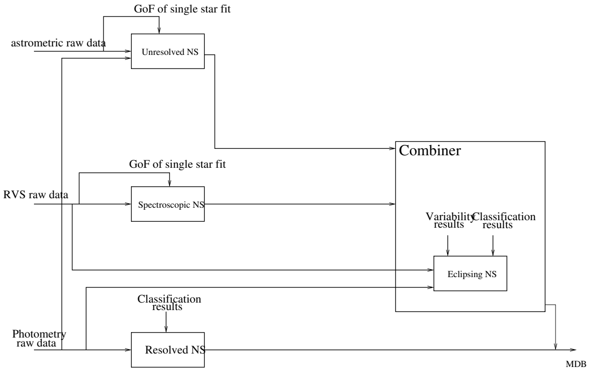


Figure 35: Data reduction pipeline for Non Single Star observations. Each instrument (astro, photo and RVS) yields its own, essentially independent, segment of the pipeline, parallel to the others but some crossover are possible.

will be listed in the source summary for a single star, for a binary, the derived systemic velocity must be listed instead. It may differ considerably from the average, depending on the eccentricity, phase coverage, etc. That value is all that matters to those who will use the Gaia output for galactic dynamics later on.

Except maybe for some Mira-type variables for which the intrinsic variation of the radial velocity can be of the same order of magnitude as the Gaia RVS precision, any object with noticeable velocity variation is a non single star. Similarly, any object whose projected motion departs from that of a single star can safely be flagged as non single (spotted stars might be the exception). The situation is different for photometric variability: at some level, every single star is variable. One nevertheless relies upon that variability to detect a fourth class of non single stars: the eclipsing binaries. That class does not come straight from any of the three instrumental pipelines but relies upon the analysis of the shape of the variability and that is why we consider it as a distinct class rather than including them in the second class (photometry-based detection). However, for a reason of geometry, eclipsing binaries are likely to overlap with those detected by RVS.

Assuming nothing but Gaia data, a fifth class of NSS will emerge as outliers in the statistics based on the overall Gaia results. For example, if in some astrophysical parameter space (e.g. colour–colour diagram), the stellar locus occupies a narrow region, binaries erroneously flagged as single could pop up way outside that locus [SIK$^+$04]. Even if none of the three instrument reduction pipelines notices any problem with the assumed single star model, it might be worth reducing the observations again, accounting for the duplicity (e.g. with the revised individual colour estimates).

All these classes of binaries are closely related to a specific Gaia instrument. We can nevertheless go beyond that by combining data from different instruments. Such a combination will sometimes improve the accuracy of some results while it will always lead to a better understanding of the stellar system. For instance, combining

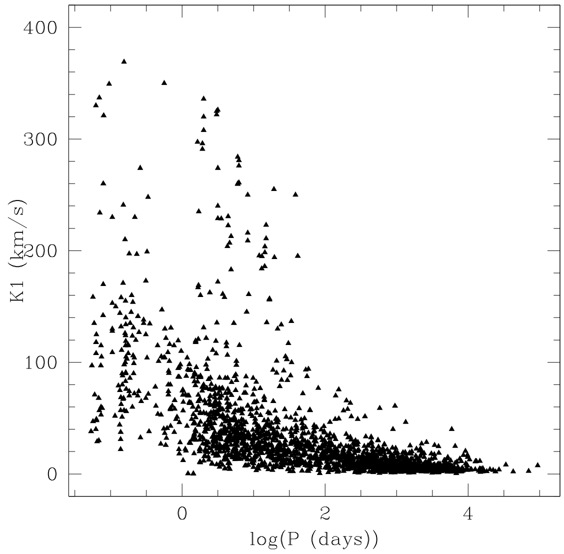


Figure 36: Evolution of the semi-amplitude of radial velocity curve of spectroscopic binaries as a function of the orbital period (source: `http://sb9.astro.ulb.ac.be`)

the photometry and radial velocities of a double-lined spectroscopic eclipsing binary yields the individual masses of the two stars, quantities that neither the spectroscopic nor the eclipsing binary reduction could derive.

On fewer occasions (owing to the low radial velocity precision), astrometry could benefit from such a combination as well. Any phenomenon with a period close to one year (due to a spotted or binary star) will likely complicate the derivation of the parallax. In the absence of independent evidence of the binary nature, the astrometric reduction of such a one-year period binary would fit a single star model with a parallax corresponding essentially to the sum of the true parallax and the appropriately phased component of the absolute photocentric orbit. With the hint from the spectroscopic orbit, it is possible to disentangle the semi-major axis and the parallax, and therefore to improve the accuracy of the latter.

A similar confusion can take place between orbital and proper motions for orbital periods exceeding the mission lifetime. Here again, an indication that the object is a binary with a long period would be useful to the accuracy of the astrometry. Unfortunately, such a hint is less likely to come from RVS than it was for the case of one-year period binaries since those long period binaries have radial velocity curves dominated by low amplitudes (Fig. 36). We are therefore less likely to derive an RVS-based orbit as the orbital period increases.

The methods that will be used to detect them with Gaia are essentially the same as for double and multiple stars. Unfortunately, however, the precision of RVS is too poor to expect any detection with that instrument alone (Fig. 37). Gaia is therefore

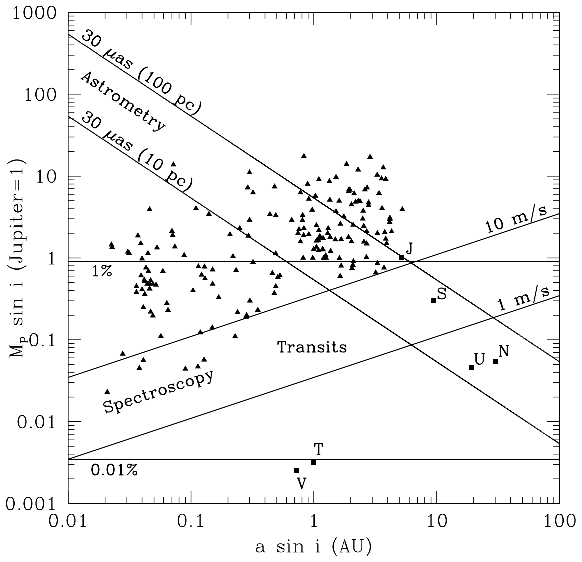


Figure 37: Sensitivity of the photometric or spectroscopic detection methods for planets orbiting a solar-type star. The two lines for the spectroscopy correspond to what is already routinely achievable with ground based observations. The astrometric sensitivity corresponds to the orbit reconstruction limit rather than simple detection. Small triangles denote the known extrasolar planets and squares denote planets of our solar system.

limited to astrometry and photometry (transit).

The probability that, for a given observer, an extrasolar planet passes in front of its host star is quite low, due to the viewing geometry. The Gaia scanning law makes any passage detection even less likely. However, the large number of stars observed partially compensates for this low probability. About 30 000 extrasolar planets are expected to cause three or more tiny (but noticeable by Gaia) flux reductions each. Even though that would make 150 times more planets than known today, these 30 000 are only candidates and will require follow up ground-based observations for their orbits to be derived and their planetary status to be assessed. It is indeed possible for a binary star to mimic a planetary transit, so confirmation will be required.

Astrometric planet detection and even orbit reconstruction are also possible. Unlike transit detection, the capacity of astrometry on that matter scales down with distance. So the astrometric wobble of the star position caused by the planet is likely to exceed Gaia precision only for nearby stars. On the other hand, the size of the wobble is orientation-independent, so the detection does not depend on the inclination. Based on extensive tests carried out by the former Planetary System Working Group, orbit reconstruction is possible for semi-major axes larger than twice the single-observation precision, not end-of-mission [LSSC00, SCLS01, Pou02, LCJ$^+$05, CLM$^+$06]. That limit, for two distances, is plotted in Fig. 37. In the case of astrometric detection. Follow up confirmation of the planetary status will also be required [Sch05].

Even though, on average, the planetary cases are characterised by a lower signal-to-noise ratio than binary or multiple stellar systems, the reduction procedure is essentially the same. The situation is slightly different in case of systems with strictly more than one planet since the considered model has to potentially account for mutual interactions between the planets whereas triple star systems (and higher multiplicities) are almost always hierarchical, thus making mutual interaction rather unlikely.

Whatever instrument the data come from, they will be one-dimensional (1D). However, whereas fluxes and radial velocities are genuine 1D, positions are originally two-dimensional (2D), they are converted to one dimension by projection along the scanning direction.

That distinction between genuine 1D data and de facto 1D data is important and leads to two very different approaches at the orbit fitting stage. On the one hand, the orbital period in authentic 1D observations can be derived using some efficient period-search methods. Even if that period estimate might suffer from aliasing, it is usually good enough for further fitting of the other model parameters.

On the other hand, converted 1D data result from the convolution of the original observations with the scanning law. The only way to derive the orbital period is therefore to simultaneously fit the complete orbital model. This is quite time consuming because it is essentially a blind quest. Once again, on the rare occasions where the system is both spectroscopic and astrometric, RVS data could provide the period, and hence speed up the whole process of orbit fitting.

On top of all scientific justifications for studying the extrasolar planets, one should keep in mind that their quest is very popular among the general public. This is therefore a typical area that can have a great impact in terms of public outreach.

5.4.2 Solar system objects

5.4.2.1 Presentation

Gaia has the potential of producing a real revolution in the field of Solar System studies. Moving Solar System objects brighter than G=20 mainly populate a sky strip centered on the ecliptic and extending on either side by about 10-15 degrees. Most of them are Minor Planets belonging to the main asteroid belt (Main Belt Asteroids, MBA); other less representative categories include, in decreasing number of detectable population, Earth crossers (Near Earth Objects, NEO), asteroids trapped in the co-orbital Lagrangian points of Jupiter (Trojans), minor bodies in the external Solar System (Centaurs, Trans–Neptunian Objects, comets, natural satellites, main planets. An object whose ecliptic latitude is between $\pm 10^o$ will be observed, on average, 65 times.

Since the geometry of observation is constantly changing due to parallax and orbital motion, each Solar System object detection has to be considered as a separate, unique event whose details (source brightness, position, orientation on the sky, etc) are function of time. Intrinsic brightness variations (mainly due to poorly known irregular shapes) and observing geometry changes make rather difficult to estimate with great accuracy the exact number of bodies that Gaia will detect at least once. Simulations and extrapolations of the known population show that this number should be in the range $\sim 3 - 5 \times 10^5$, but this figure depends on the model applied to de-bias the population that is known today and the largely unknown contribution of objects having more exotic orbits.

The role of Solar-System-Object processing is to reconstruct the dynamical and physical properties of the objects by making an appropriate synthesis of those isolated detections. The following final mission products are expected to be produced:

1. accurate orbital elements for all objects (10^2–10^3 times better accuracy than present orbital solutions); accurate orbital solutions for asteroid and planet satellites,
2. masses for some $\sim$100 largest asteroids with a relative accuracy better than 20%,
3. direct size measurements for no less than $\sim$1000 asteroids,
4. parameterised shapes, pole coordinates, spin periods and scattering properties for at least $\sim$10000 asteroids
5. cometary activity indicators,
6. non-gravitational perturbations (cometary jet effect, Yarkosky acceleration, tidal acceleration for planetary satellites..),
7. gravitational PPN β parameter from a set of chosen asteroids with an accuracy better than 10^{-3},
8. a catalogue of Solar System object detections (calibrated positions and accurate photometry based on the derived physical models),
9. a new taxonomic classification based on spectral reflectance properties.

5.4.2.2 Data reduction structure and time scales

The data processing of the Gaia observations of solar system objects is a very well characterized task that relatively easily can be separated from the general processing, once standard calibration for astrometry and photometry are available. This is all the more true because the size of the restricted data set is moderate and could be

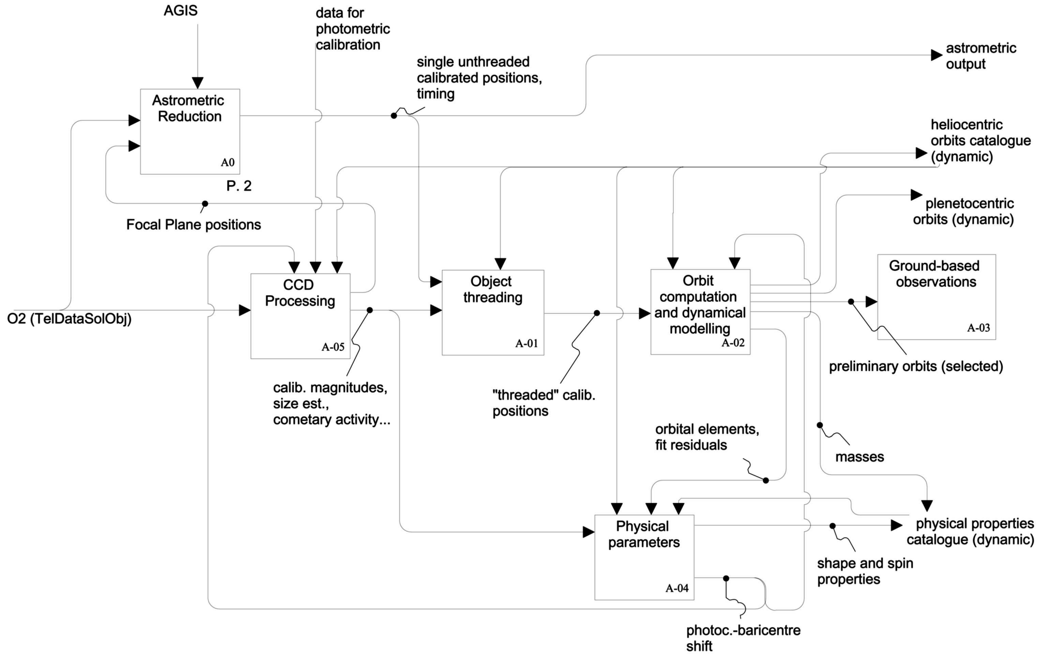


Figure 38: Solar System functional analysis scheme. The processing steps are summarized in the text.

extracted from the database by means of standard procedures. However, this does not imply that the amount of computation is small and that there is no algorithmic complexity. In practice the processing remains a very challenging task and presents specific problems due to (i) the rapid motion of the sources and (ii) their sizeable apparent angular diameter in comparison to the stars. Solar System processing is based on a variety of input data that will be used to extract physical parameters from each detection. These data can be subdivided in two main broad categories:

The first one includes pre-calibrated measurements, useful for any preliminary data reduction:

- astrometric measurements of the object photocentre expressed in the focal plane coordinates, one for each transit;
- calibrated magnitudes/colours for every transit.

The second set will be used for a more refined treatment and it includes:

- telemetry data including raw CCD ”images”;
- raw RB and RP spectra.

In order to understand the structure of the data reduction, it must be underlined that different complexity levels and time scales are involved:

- Short term data processing. These are updated as frequently as possible, and involve the refinement of orbits while the flux of incoming data is active, during the mission. The main aim of this processing is to identify Near Earth objects that due to dynamical or physical properties could deserve rapid attention by Earth-based observers. The dynamical model used for the orbit computation is simplified and positions calibrated with the same preliminary data treatment reserved to stars can be directly used.

- Final data processing. These should use all the available information, both from calibration data and from observations. In this stage photocentre–barycentre corrections will be applied, based on asteroid shapes models (derived from photometric inversion), and all relevant perturbations should be included in the dynamical model. Some typical final computations are specific to this step (for example: masses; derived quantities, such as asteroid density).

- Intermediate data processing. This timescale will implement a level of complexity similar to the final processing (depending upon the fraction of available observations over the total expected). It will be mainly devoted to assess the behaviour of the data processing pipeline on over real data, to check the effectiveness of the reduction algorithms and to its optimisation.

The main stream of the data reduction pipeline is devoted to orbit computation. In fact, it contains the basic processes that allow running all the other tasks. Two main steps compose its core: (1) the identification of observations of each object (object threading); this includes comparison to a catalogue of known objects, and identification of previously unknown Solar System bodies; (2) the orbit computation from the set of observations for each object.

Photometry from threaded observations will be used to derive shapes, pole coordinates and rotation periods. A physical ephemeris can thus be established and used to provide a more complex model to a refined CCD-processing module, performing improved signal fitting. The resulting astrometric positions will be corrected for photocentre-barycentre shift; photometry will also be refined. Iterations with the physical properties and orbit determination will thus become possible, converging to the best possible solution.

5.4.2.3 Dynamical model

The orbit computation will rely on different dynamical models at various complexity levels. Short arc orbit propagation, performed on few observations, and yielding ephemeris valid on a short time span, will be based upon a simple model (Keplerian orbits will be sufficient).

It is considered that new Near Earth Objects falling in some categories of specific interest will deserve an immediate follow-up from the ground. The selection criteria will favour those asteroids that are observable from Earth only for a short period, and whose orbits or physical properties need further data (for example: possible close encounters with Earth in the future). This Earth-based activity requires the determination of the trajectory of the orbit not on the sky plane, but in space (in practice, an orbit) due to the high parallax displacement relatively to the Gaia position in space. Usual extrapolations of the sky-projected position cannot be used in this case, and the trajectory in the three-dimensional space is needed. Using just two transits in the AF, the determination of both positions and velocities of the object allows to constraint the approximate orbital elements of the asteroid in a usable way.

On longer timescales, the orbits of all objects are computed with increasing complex approximations of the dynamical model. The final computation should take into account all perturbations potentially measurable in the data (gravitational and non-gravitational perturbations, relativistic frame).

5.4.2.4 Physical properties

Simulated photometric, disk-integrated Gaia observations of MBAs have been used to test inversion algorithms, using three-axial ellipsoids as models for asteroid shapes (both for simulations and for inversion). The determination of pole coordinates, spin period and ellipsoid axis ratios have been shown to be possible and to provide good accuracy, as long as the number of observations is sufficient. In the plane defined by noise strength, photometric accuracy and observation number, the addition of noises of different amplitudes has shown that there is a well-defined region in which the inversion is successful.

Direct size measurements require fitting the observed photon distributions in the AF CCDs to a model of the object derived from pre-mission data and/or from the disk-integrated photometric inversion. Hypotheses on the scattering properties can also be tested at this level, in order to improve the fit quality.

During the fit, the proper motion of moving objects must be taken into account. This can happen in two different steps. First, the procedure will deal with elongated images coming from each single CCD. Then, it will adopt a specific strategy if the shift of the objects relatively to each transmitted window produces a considerable loss of signal. In fact, in this case, the parameters describing the best-fitting model found for the AF1 signal (or, in general, found for those CCDs where the object is acceptably well-centered in the window) can be kept constant, and a simple re-centering is operated. In this way, a moderate loss of signal due to image drift during the transit can be tolerated, and the best possible exploitation of the available detections is made.

5.5 Astrophysical Parameters

5.5.1 Context

The goal of this area of the data processing is to determine astrophysical classes and astrophysical parameters of all objects (stars, QSOs, galaxies, solar system objects etc.) observed by Gaia, based on photometry, spectroscopy and astrometry. As indicated in Sect. 2.1.3, there is no sharp boundary between what is *data processing* and what is post-mission *analysis*. Some sophisticated analyses will be left to the community after publication of the Gaia catalogue. Nonetheless, we must carefully define which tasks we undertake, guided by the unique contributions of Gaia and its primary scientific goals, even if the choices may seem slightly arbitrary in places. Section Sect. 2.1 and Sect. 8.9 put these choices in the context of the Gaia mission and intended science. Sect. 8.9 also discusses more specifically the boundaries of these tasks.

5.5.1.1 Background

The main goal of Gaia is to study the composition, origin and evolution of our Galaxy through very accurate stellar distances and 2D (and 3D) space motions. However, this information will be of limited use if it cannot be associated with the intrinsic physical properties of the target stars.

An important difference between Hipparcos and Gaia is that Gaia goes much fainter, to about $G = 20$ rather than $H_p = 12.4$. Partly for this reason, Gaia will perform real-time on-board detection of targets and will not have an input catalogue. As a consequence, we generally have no prior astrophysical information on the targets, not even knowing whether it is a star. The main purpose of the low resolution spectrograph on Gaia (RP/BP) is to enable us to classify all the sources and to estimate their astrophysical parameters (APs). Data from the RVS spectrograph will also be used to characterize brighter sources, and parallaxes, proper motions and variability information can be exploited where appropriate. The on-board detection system selects only point sources for observation. “Classification” in this section therefore refers to the use of one-dimensional spectral information (plus astrometry) rather than 2D spatial (morphological) information. The challenge is to design a classification and AP estimation system which can optimally use these heterogeneous data and be robust enough to cope with numerous complicating issues.

The most fundamental properties of a star are its mass, age and chemical composition. Of course, age is not directly observable and masses can only be determined directly (i.e. dynamically) in select binary systems. Thus we must rely on indirect atmospheric indicators via the spectral energy distribution, which is RP/BP and RVS in Gaia. From these we can estimate the effective temperature, surface gravity, abundances and the line-of-sight interstellar extinction. Combined with the parallaxes we can – at least in principle – derive the intrinsic luminosity and then estimate the

Table 6: Outline of how stellar parameters may in principle be derived from the Gaia data. The "n" prefix to RP/BP indicates normalized data, i.e. neglecting the absolute flux level. BC refers to a bolometric correction (which may be implicitly present in the model rather than explicitly calculated), ϖ is the parallax and A_G the interstellar extinction.

measured quantity		derived quantity	model
without astrometry:			
nRP/BP, (RVS)	$\Rightarrow$	T_{eff}, $\log g$, [Fe/H], A_G, BC, [α/Fe]	atmospheric model
additional use of astrometry gives:			
G, ϖ, A_G, BC	$\Rightarrow$	L	$2.5 \log L - f(G, BC)$ $= A_G - 5 \log \varpi$
L, T_{eff}	$\Rightarrow$	R	$L = 4\pi R^2 \sigma T_{eff}^4$
$\log g$, R	$\Rightarrow$	M	$g = GM/R^2$
L, T_{eff}, [Fe/H]	$\Rightarrow$	age	evolutionary model

mass and age (albeit with a degeneracy due to the Helium content), as summarized in Tab. 6.

5.5.2 Objectives

The main objectives of the classification are as follows [BJ02] [BJ03] [BJ05]

Discrete Source Classification Determination of whether an object is a single star, unresolved binary, galaxy, quasar, asteroid etc.

Identification of a clean set of QSOs This is used by the GIS to define the extragalactic astrometric reference frame.

Estimation of Astrophysical Parameters (APs) For those objects identified as stars, determine their intrinsic physical properties. The relevant (and obtainable) ones are effective temperature, T_{eff}, surface gravity, $\log g$, overall metallicity, [Fe/H], and line-of-sight interstellar extinction, A_G. (The last of these is effectively intrinsic because we determine it on a star-by-star basis, i.e. without reference to a spatially smooth Galactic dust model.) Other APs of interest (which should be derivable to a greater or lesser extent from either RP/BP or RVS for brighter stars) include: alpha-process elements [α/Fe]; some peculiar abundances anomalies; emission line characteristics.

Provide suitable AP estimates to aid the extraction of the RVS spectra At the minimum a spectral type estimate (based on RP/BP) is required to select a cross-correlation template for the RVS spectrum. (This requirement is essentially covered by the previous one.)

Identification of unresolved binaries Most stars are in multiple systems. Some of these can be recognised from the astrometry, and a few will be visual binaries, but many will go undetected via these approaches. Nonetheless, with favourable brightness ratios, a binary could be detected from the shape of its composite spectral energy distribution. This is important for determining the stellar mass function (as opposed to the system mass function) and for investigating the evolution of stellar clusters.

Identification of new types of objects An extensive survey such as Gaia will inevitably observe new types of objects, such as new types of variable stars, rare stars (e.g. brief phases of stellar evolution), abnormal abundance patterns or rare multiple systems. The classification system must be designed to detect such objects, e.g. via the identification of "outliers". Virtually by definition this will be carried out via unsupervised classification methods.

5.5.3 System design

RVS data are only available for bright sources ($V<17.0$) and only for yet brighter stars will the SNR be high enough to render these data useful for AP estimation. Consequently the majority of Gaia sources will only have RP/BP data, and the classification system is therefore centred on these data.

During the mission the entire Gaia DP system, including the classification pipeline, is operated in six month cycles on all available data. Thus every six months we get (new) classifications and APs of all sources observed so far based on all data available at that point. Early in the mission "all available data" means RP/BP data (robustly averaged if we already have multiple epochs) and RVS data (for brighter sources). Later in the mission it is extended to parallaxes, proper motions, variability information and information on binarity from astrometry and spectroscopy.

This continual processing is considered essential, because it enables us to (a) learn about the nature of the real data which Gaia gathers, from which we will improve our algorithms and models as the mission progresses, and (b) provide early data releases as appropriate. Furthermore, AP estimations based on RP/BP are required by other parts of the data processing, for example the extraction of the RVS spectra.

Fig. 39 shows a sketch of the classification data processing pipeline for the "late mission" phase (i.e. multi-epoch data and astrometric solutions are available). The ellipses on the left denote the data sources and the rectangular boxes the algorithms. First, a classification of the sources is performed with the Discrete Source Classifier

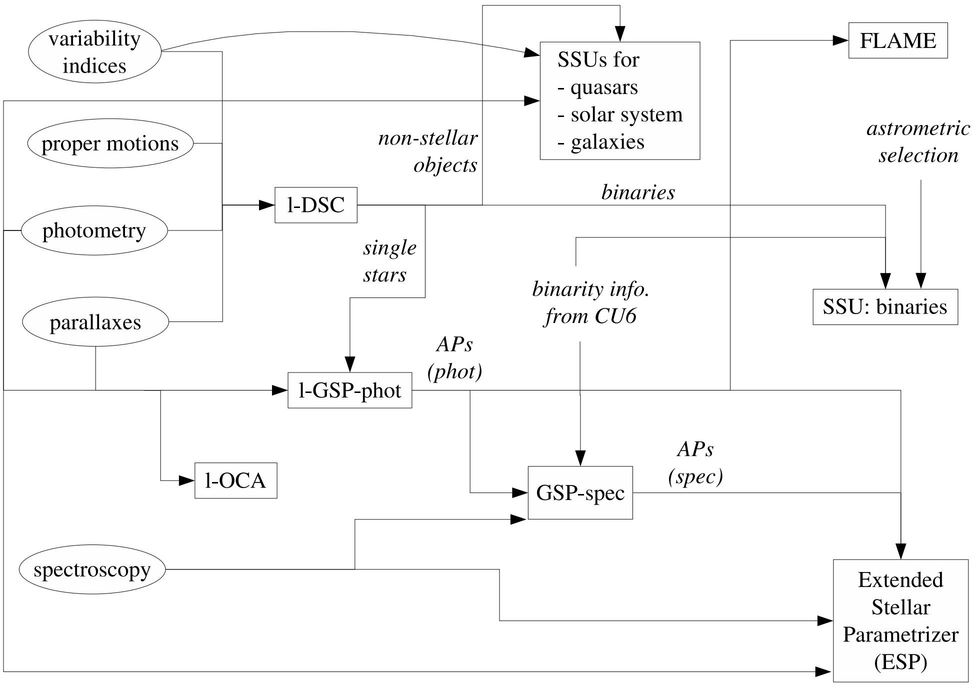


Figure 39: A sketch of the data flow for the classification and astrophysical parameter estimation algorithms in the Gaia data processing pipeline. This assumes a "late mission" phase, i.e. once parallaxes and proper motions are available. The acronyms are explained in the text. Lines which cross (four-way) do not connect to each other, whereas lines which fork (three-way) do. To avoid crowding, the outputs from the right-most algorithms are not shown, neither are the inputs to SSU: binaries.

(l-DSC; the "l" indicating here and elsewhere that it is for the "late mission" phase). This assigns a probability to each source that it is a single star, binary star, QSO, asteroid etc. For those sources with a sufficiently high probability of being single stars, they are passed to the "Generalized Stellar Parametrizer" (l-GSP-phot; the "phot" denoting that it operates on "photometric", viz. RP/BP, data). This assigns APs and their estimated uncertainties (more generally, covariances) to stars[15]. Non-stellar objects are sent to the appropriate "Specific Source Unit" (SSU), which determine intrinsic APs appropriate for that class of object (e.g. photometric redshifts for QSOs, taxonomic classes for asteroids).

---

[15] Stellar parallaxes provide an additional constraint on atmospheric parameters, e.g. $\log g$, and GSP-phot uses these in the late mission phase. But proper motions are not used to classify stars. To do this would be to bias the classifications with our present model of the relation of Galactic kinematics to stellar populations, precisely the topic which the Gaia mission is designed to address. On the other hand, proper motions are used by DSC, because this is a valid method of identifying extra-galactic objects (which will have individual proper motions consistent with zero).

Stellar APs are also estimated using the RVS data with l-GSP-spec (the "spec" indicating that it operates on spectroscopic data). Because the RVS wavelength range is quite narrow, the AP estimates from GSP-phot are used as priors here. Spectral binarity information (processed by CU6) may also be used here.

GSP-phot and -spec are the two main "workhorses" of stellar AP estimation. However, because they will operate over very wide AP ranges, they may not produce the optimal estimates for a small number of peculiar stars. For example, very cool stars, stars with strong emission lines, or rapidly rotating stars require particular attention (and extended model atmospheres) to estimate reliable APs. For this reason, a set of algorithms under the name "Extended Stellar Parametrizer" (ESP) re-examines a restricted subset of "extreme" stars in order to (1) test alternative assumptions (stellar models), (2) provide additional parameters suitable to a specific type of object (e.g. Carbon abundance), (3) explore using additional data (perhaps external to Gaia) or alternative combinations of data. Stars are identified as extreme (and therefore sent to ESP) based on their APs from GSP-phot and GSP-spec.

The FLAME algorithm is the "Final Luminosity, Age and Mass Estimation". This is applied in the post-mission data processing on the final data set. It estimates derived global stellar parameters from the atmospheric parameters (e.g. from GSP-phot) using evolutionary models.

Finally, the "OCA" algorithm in Fig. 39 to "Object Cluster Analysis". This performs an unsupervised classification (cluster analysis, outlier detection etc.) of the RP/BP data. Such analyses do not take into account any class information ("labels"), but rather look at the similarities between sources based only on the data themselves. They are, therefore, independent of any physical models. From this we can extract natural groups and relationships in the Gaia data set which should help the identification of new types of object. OCA therefore complements the model-based (or "supervised") approaches (used in the rest of the classification pipeline), which can only correctly classify sources which are properly represented in their training grids (see below). Note that OCA is the only part of the classification system where sources are treated non-independently.

The system briefly sketched here reflects the current thinking in the DPAC concerning the Gaia classification issue. It may well evolve during the pre-launch development and, more importantly, in the face of real data during the mission. For example, iterations between DSC and GSP-phot could be considered, in which GSP-phot "returns" a source to DSC if it is unable to find any reasonable AP solution.

5.5.3.1 Data products

Every source which Gaia observes will, as a minimum, be assigned a set of class probabilities (from DSC). This will include an "unknown" class. For most sources a single class will dominate (i.e. its probability is close to 1.0). In addition, APs will be estimated for all classes with a probability above some threshold.

A consequence of the system design is that any one star may end up with multiple sets of APs, assigned from GSP-phot, GSP-spec and ESP. This reflects the fact that our knowledge of a system depends on the data and models we use in our inference: Any AP estimate is ultimately model-dependent (in this case on stellar evolution and atmosphere models) so if we use different models then it is quite reasonable that we may get different answers. This is even desirable, because discrepancies between AP assignments may provide insight into peculiar objects and/or problems with our physical or data models. We will attempt to reconcile any differences in AP estimates during the mission and use these to improve our models. Where disagreement remains, multiple solutions will be reported. In order to provide a homogeneous system, APs from GSP-phot will always be reported in the final catalogue. Additional solutions will also be reported as appropriate, and the "best" indicated.

5.5.3.2 Algorithms

As indicated above, all the classification algorithms (except OCA) will be based on "supervised" models. These classify sources or estimate their APs source-by-source based on their similarity to a set of predefined templates. The templates, e.g. synthetic spectra, embody our current understanding of the sources based on known physics. The methods may involve direct comparison of templates or an inference of the data → APs mapping; they range from simple linear regression and minimum distance methods to more complex kernel methods and multidimensional nonlinear regression.

The training data will be synthetic model spectra, modified by real observations. These modifications – or calibrations – are necessary to correct deficiencies in the models. Ground-based high resolution spectroscopy of a grid of stars which Gaia will observe are used to derive accurate APs (via conventional line analysis methods). When combined with the Gaia observations of the same objects we have a set of labelled data which are used to calibrate the algorithms. Ground-based (spectro)photometry of the same sources will also be used to modify the synthetic SEDs directly, e.g. to correct for missing opacities, incorrect oscillator strengths or incomplete input physics. (This could be done following the principle used to construct the BaSeL libraries [LCB97].) The modified synthetic data are processed by an instrument simulator to simulate Gaia data, from which we build our classification models.

Whatever the specific algorithm, the derivation of *physical* parameters always depends on *physical* models, in this case via synthetic spectra. Thus the Gaia APs will, of course, be tied to a specific set of physical models. However, AP estimates based on alternative models can easily be derived by retraining the algorithms on those models and applying them to the (final) Gaia data. It is perfectly plausible to consider even now (and certainly in 2019), a data product from the mission being not just a catalogue but also the classification pipeline itself, which can be rerun using physical models of the user's choice.

The following sections describe the main algorithms in more details. The underlying principles are all quite similar and the different algorithms mentioned should therefore be considered as baselines or examples. Up to launch the DPAC will invest considerable effort into testing existing algorithms and developing new ones for each problem-specific domain, with the final choice being made at the appropriate time.

It must be stressed that the classification system for Gaia does not simply involve implementing existing algorithms. While there has been much research into classification and AP estimation with astronomy in recent years, these have either had relatively modest goals or have worked with a restricted set of data, e.g. in which contaminants have already been cleaned out. Gaia represents a number of challenges which must be addressed in order to produce a robust system. The system must

- cope with heterogeneous data (RP/BP, RVS, astrometry, variability indices)
- accommodate degenerate solutions and report multiple solutions where necessary
- provide covariance estimates on the derived APs
- handle the fact that some APs have a very weak impact on the spectral data (e.g. $logg$) compared to other APs ($T_{\rm eff}$, extinction)
- accommodate prior AP estimates
- be robust to missing and censored data
- be fast enough to classify 10^9 sources in a matter of weeks or less. (This time scale is chosen to match the 6 month intervals of the main database versions.)

5.5.4 Discrete source classification

The *Discrete Source Classifier* algorithm (DSC), determines the astrophysical class of every source which Gaia observes. The relevant classes from the point of view of the Gaia science and the subsequent parametrization are

- single star
- physical binary star
- optical (i.e. non-physical) double system
- galaxy
- quasar

- asteroid
- unknown / other

The main objective is to correctly identify single stars (for the Gaia science) and quasars (for the external reference frame). A secondary objective is to identify the other types of objects and to isolate physical binary stars (which will be used along with binaries detected astrometrically and spectroscopically to correct the stellar *system* luminosity function to the single star mass function.) A separation between single stars and binary stellar systems in DSC is in principle possible based on the composite spectral energy distributions. Of course, for brightness ratios much less than unity this becomes increasingly difficult, so there will be an unavoidable contamination between the two classes.

The DSC algorithm operating only on the RP/BP data can be implemented using a standard classification algorithm. There are many examples in the literature, including linear or quadratic discriminant analysis, logistic regression, classification trees, minimum distance methods, support vector machines and neural networks. The general approach with all of these is to use a representative set of labelled data to either (1) model the (density) distribution of each class in the data space, or (2) to fit boundaries between the classes in the data space. In the latter case, once the boundaries have been defined, a newly measured object is classified simply by seeing in which class region (on which sides of the boundaries) it falls. In the former case, because the distribution is modelled the probability of a new object being a member of any one of the classes can be derived. This is generally preferable, because probabilities allow us to express the natural uncertainty arising in classifying noisy data, as well as to accommodate prior probabilities in a Bayesian framework. It has been decided that DSC will give (normalized) probabilities for all of the above classes (and new classes may be added as we progress). Other parts of the data processing, such as subsequent parametrization algorithms or the selection of quasars, may then set their own probability threshold to produce a sample with the desired level of confidence or trade-off between completeness and contamination.

Of course, the data space may have a high dimensionality, of order 35–200 if using the RP/BP spectra directly (the range uncertainty depends on how the spectra from different epochs are combined). The key issue is how well we can separate the different classes in the data space. This depends on both (1) how well the classes are separated, and (2) how flexible our algorithm is at modelling the nonlinearities. Some classification algorithms, including support vector machines, implicitly project the data into a higher dimensional space and fit simpler (e.g. linear) models in that space. These correspond to more flexible nonlinear models in the original space. Nonetheless, an important consideration is the preprocessing of the data. This can take the form of dimensionality reduction (combining highly correlated dimensions; removing irrelevant ones), transforming the data (to linearize or just increase numerical stability) and weighting. All of these are being considered in the development of

DSC.

As described in Sect. 5.5, later in the mission other data in addition to RP/BP will be available. Particularly useful for DSC are the parallax and proper motion. Extragalactic objects will have these parameters consistent with zero, which will help in their classification. (Of course, we must avoid a circular identification of the quasars!) Variability information may also help in the identification of quasars. Naively, these additional parameters may simply be included as additional dimensions in the data space where we model the class distributions. In practice, it may be better to handle such parameters in a more physical way.

5.5.5 General Stellar Parametrizer : photometry

The algorithm *General Stellar Parametrizer – Photometry*, or GSP-phot for short, is the main algorithm in the DP pipeline which estimates stellar astrophysical parameters (APs). All sources classified by DSC as being single stars above some probability threshold are sent to GSP-phot. This is a supervised machine learning algorithm which infers a mapping function data → AP. Broadly speaking there are two ways to achieve this.

First, one may explicitly fit a global function via (nonlinear) multidimensional regression on a set of labelled ("training") data. This fitting must typically be done with a numerical optimizer, either a gradient-based method (e.g. conjugate gradients or BFGS) or a function evaluator (Nelder-Mead, or stochastic methods such as simulated annealing and genetic algorithms). Examples of such algorithms include projection pursuit regression, additive models (e.g. Multivariate Adaptive Regression Splines, or MARS), feedforward neural networks, support vector machines, and many other variants on these. Once trained, these methods yield a function of the form $\phi = f(\mathbf{x};\mathbf{w})$ which predicts the AP ϕ for a given input spectrum, $\mathbf{x}$. The "astrophysical knowledge" of the model (e.g. which features are relevant for each AP) is encapsulated in the model's internal parameters, $\mathbf{w}$, which were learned from the data during the training process.

Alternatively, one may use a local method. This is essentially the same as the previous method, but now the function is learned locally. An extreme case of this is the simple nearest neighbours algorithm, in which the APs of a measured source are set to those of the "nearest" template in a training grid. In this case the function (f in the above equation) is just a constant.[16] More sophisticated variants of this involve using multiple neighbours and averaging their APs with some weighting function, such as a user-defined (and thus problem-specific) kernel function. One of the main challenges in this work is defining a suitable kernel. For example, work is in progress within CU8

---

[16] The popular χ-squared minimization algorithm is a particular case of this nearest neighbour algorithm in which the distances in each data dimension (spectral bin) are weighted by the expectation of the standard deviation of the noise.

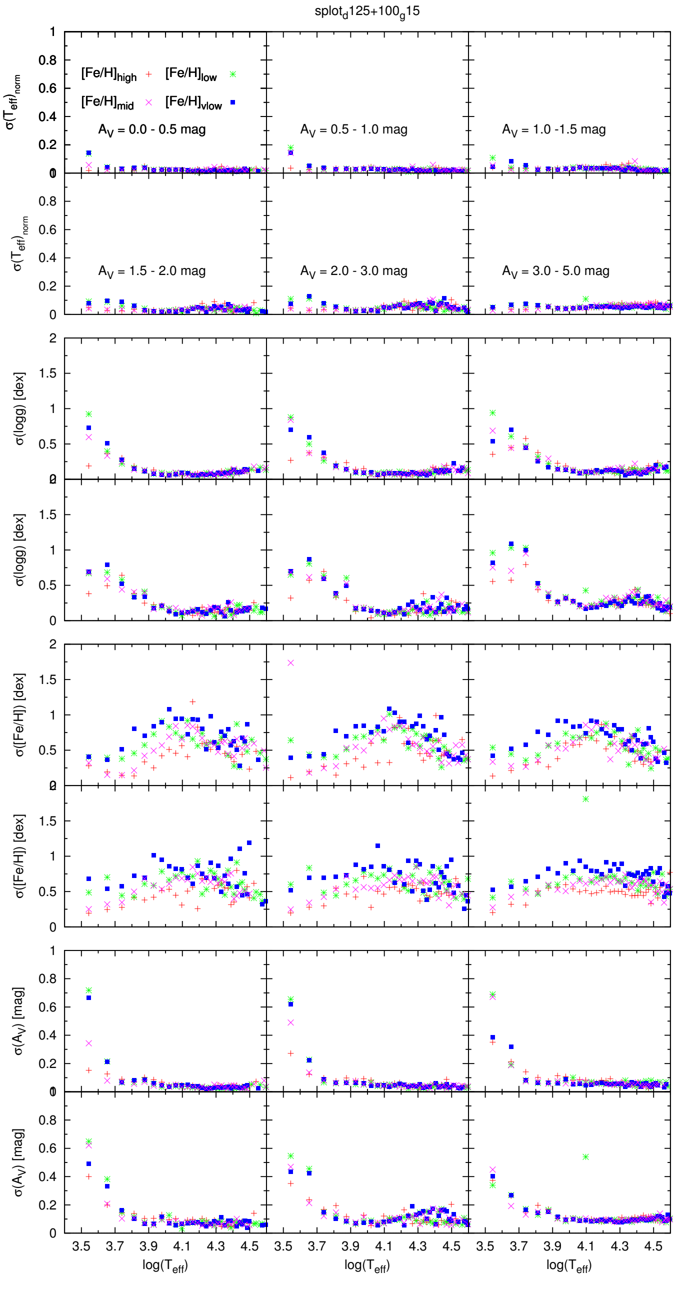


Figure 40: Precision of AP estimates from RP/BP at G=15 (end-of-mission noise level). Each panel plots the standard deviation of residuals versus the logarithm of the (true) temperature. Results are shows for four groups with metallicities of $[Fe/H]_{high} \in [0,1]$ dex, $[Fe/H]_{mid} \in [-1.5, 0]$ dex, $[Fe/H]_{low} \in [-3, -1.5]$ dex and $[Fe/H]_{vlow} \in [-5, -3]$ dex and six different ranges for the interstellar extinction, A_V.

on developing a local regression method which learns the optimal weighting scheme (sensitivity matrix) locally from the training data.

The main difference between global and local methods is that the former define a single mapping function over the whole data space whereas the latter either define no explicit function or fit a function locally for every observed instance. Early in the mission the data space comprise just the RP/BP fluxes. In the late mission phase (see Sect. 5.5.3), the product of the parallax and the G-band flux (to give a measure of intrinsic flux) will be used as an additional input dimensions. Due to the complexity of the relationship between the measured data and the APs, one will typically want to use a nonlinear function for f.

A significant part of the early development work for GSP-phot involves the testing of existing algorithms, the development of new, problem-specific variants, and their evaluation. This has already been started during the Gaia phase A and B1 studies by the working group "Identification, Classification and Astrophysical Parametrization" (ICAP). Most of this work focused on the estimation of stellar APs based just on the photometric data and some results are summarized in [Bro03]. This work also helped to optimize the photometric system [JHBea06] (which was the baseline before RP/BP low resolution spectrophotometry was adopted), specifically via the Heuristic Filter Design (HFD) approach [BJ04a] [BJ04c] [BJ04b].

We have preliminary estimates of the precision with which we can estimate the four main stellar APs from noisy RP/BP data. The following estimates were obtained using a flexible nonlinear, multidimensional regression algorithm[17] and are reported in more detail in [WKBJ06]. This model is trained on a discrete grid with a wide range of APs: A_V [0 mag:5 mag]; [Fe/H] [-5.0 dex: +1.0 dex]; $\log g$ [-1.0 dex: +5.5 dex]; $T_{\rm eff}$ [3000 K:40 000 K]. Its performance is then evaluated on a separate grid of stars with APs drawn from a continuous distribution over these ranges. The precision depends both on the SNR (i.e. the G-band magnitude and the number of co-added transits) plus the AP range itself (e.g. [Fe/H] estimates are a priori poorer for hot stars than cool stars due to the relative weakness of metallicity-sensitive features). Fig. 40 shows results with end-of-mission data on stars with G=15. Summary results are given in Tab. 7.

These results must, however, be treated with some caution. On the one hand they could be considered optimistic because they do not take into account all noise sources (in particular, the effects of radiation damage to the CCDs or calibration errors) and they are obtained from purely synthetic spectra (whereas real stars show variance due to other APs). On the other hand, the algorithm has not been particularly optimized for the specific problem domain and the model is a single regression on the entire range of APs (which may be too inflexible). Therefore, while these are the best estimates available at this time (and are unlikely to be very wrong), it is difficult to accurately predict AP precision at this time. Note also that with end-of-mission data

---

[17] `http://www.mpia-hd.mpg.de/homes/calj/statnet.html`

we will take advantage of the parallaxes which should improve estimates.

Table 7: Estimates of the precision with which stellar astrophysical parameters (APs) can be derived from end-of-mission RP/BP spectrophotometry with noise levels simulating G=15 and G=18. The error statistic is the mean absolute value of the residuals in the units of the AP (except for T_{eff}, which is expressed as a fractional error). Results are reported at three T_{eff} ranges: HOT (12 000 K $\leq T_{eff} \leq$ 40 000 K), MEDIUM (8000 K $\leq T_{eff} \leq$ 12 000 K) and COOL (3000 K $\leq T_{eff} \leq$ 6000 K) for all values of the other APs. Note that there is significant variation in performance as a function of these APs, i.e. better results can be obtained when looking at narrower ranges of APs, e.g. low extinction (see Fig. 40).

T_{eff} range	E(A_V) [mag]	E([Fe/H]) [dex]	E($\log g$) [dex]	fracE(T_{eff})
		G=15		
HOT	0.09	1.31	0.25	0.060
MEDIUM	0.08	1.03	0.20	0.051
COOL	0.24	0.42	0.65	0.043
		G=18		
HOT	0.28	1.55	0.48	0.198
MEDIUM	0.17	1.45	0.38	0.097
COOL	0.32	0.63	0.92	0.062

5.5.6 General Stellar Parametrisation: spectroscopy

The General Stellar Parametrisation-spectroscopy (GSP-spec) algorithm has to furnish the values of T_{eff}, $\log g$, [M/H], [α/Fe], individual chemical abundances and A_V, plus their uncertainties, from calibrated RVS spectra (with robust time average and known flux covariances) on single stars.

Although the final algorithm for GSP-spec has not been definitely defined, we describe here the MATrix Inversion for Spectral SythEsis (MATISSE) algorithm ([RBBdL06]), specifically developed for the RVS data. The parametrization problem, applicable to tens of millions of stellar spectra, is approached in a way that tries to tackle some disadvantages of other automated classification techniques, such as excessive computing times. This method uses the inversion of the covariance matrix of a grid of synthetic spectra to determine a basis set, from which a particular stellar parameter is determined by projection of an observed spectrum.

Let us consider an observed spectrum $O(\lambda)$, corrupted by a Gaussian noise, independent of λ, of standard deviation σ. A grid of theoretical spectra $S(\lambda, \theta)$, where θ corresponds to the stellar parameters, is implemented for its analysis. In this framework and mainly due to the non-linear behaviour of stellar spectra as a function of the physical parameters, the parametrization problem consists of i) computing all the distances between $O(\lambda)$ and the models and ii) interpolating between the distances corresponding to the neighbour parameter values (limiting the computations to a restricted cell of models). In this situation, the determination of the set of parameters, corresponding to the minimum distance, leads to a solution with the accuracy of the grid resolution. The problem is to increase the precision without computing new models.

In order to solve this interpolation problem an algorithm, MATISSE, based on the projection of the observed spectrum on specific basic vectors, has been developed. In other words, the implemented algorithm determines a vector, $B_\theta(\lambda)$, allowing to derive a particular stellar parameter θ by projection of an input spectrum on to it. This θ parameter could be effective temperature, the gravity, the global metallicity, the $[\alpha/\mathrm{Fe}]$ content, individual chemical abundances, v sin$i$, etc. The $B_\theta(\lambda)$ function is derived from an optimal linear combination of theoretical spectra (the neighbour models) and it relates, in a quantitative way, the variations in the spectrum flux with the θ variations.

First of all, the data on a particular θ variable and the spectra of the grid are subtracted of their mean value. The $B_\theta(\lambda)$ basis is then constructed from a linear combination of spectra, with α_i being the weight associated to the spectrum $\mathrm{S}_i(\lambda)$:

$$B_\theta(\lambda) = \sum_i \alpha_i S_i(\lambda) \tag{24}$$

The parameter θ_i is estimated by the projection of a spectrum into the corresponding basis vector:

$$\hat{\theta}_i = \sum_\lambda B_\theta(\lambda) S_i(\lambda) \tag{25}$$

with $\hat{\theta}_i$ being the recovered value.

Combining (24) and (25) we obtain :

$$\hat{\theta}_i = \sum_j c_{ij}\alpha_j \tag{26}$$

where c_{ij} is the correlation value between the spectra S_i and S_j. Taking into account that the spectra have been subtracted of their mean value, c_{ij} can be interpreted as the covariance between the S_i and S_j, if the spectral values are considered as random variables. The α_i are obtained from the maximum correlation between θ_i and $\hat{\theta}_i$.

In the general case of a covariance matrix empirically found to be non invertible, we determine a linear relation between the values of a parameter θ_i for a given set of spectra, $\mathrm{S}_i(\lambda)$, and the product $\mathrm{S}_i(\lambda)\ B_\theta(\lambda)$. Then, to determine the parameters of an object spectrum not belonging to the learning set, it is only necessary to multiply

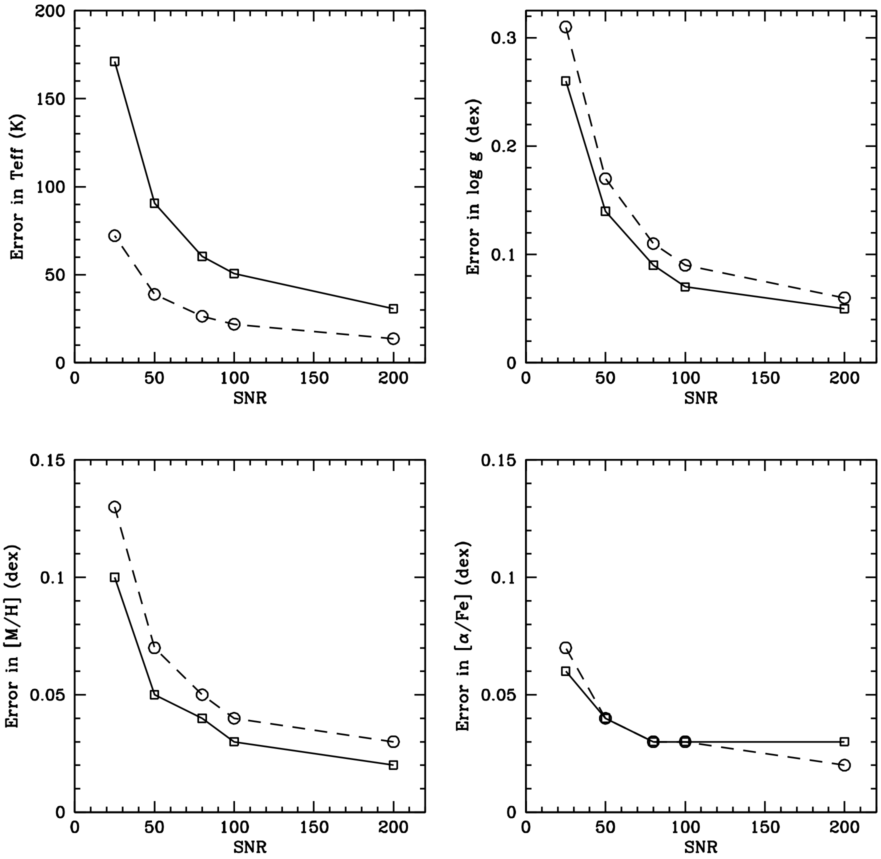


Figure 41: Maximum errors in the recovered parameters as a function of the S/N, for synthetic flux calibrated spectra in the Gaia/RVS domain. The two different lines on each panel correspond to metal rich cool dwarfs (solid line and squares) and intermediate metallicity cool giants (short dashed line and circles)

it by the corresponding $B_\theta(\lambda)$ and to transform the result using the derived linear regression between $\hat{\theta}$ and θ. This procedure is therefore extremely rapid and ideal for the analysis of huge quantities of data.

In order to avoid the effects of important non-linear variations in the spectra, it is advisable to restrict the working domain to a subregion of the spectra grid. To this purpose, a two level procedure can be followed, by deriving initial $B^0_\theta(\lambda)$ functions to make a preliminary guess in the parameters and then to refine the result using local $B^l_\theta(\lambda)$ functions. This initial guess can also be done through an input from the photometric classification (see Sect. 5.5.5).

MATISSE is a form of linear regression in which the weighting vector has been determined from the covariances in a set of synthetic spectra. This has the advantage over some other spectral parameter derivation procedures that the weighting function, $B_\theta(\lambda)$, gives direct insight into the relevance of the spectral inputs in determining the parameter. Basically, $B_\theta(\lambda)$ deviates more from zero at those wavelengths mostly affected by a change in θ, that is, at the spectral regions containing the highest quantity of information on a given parameter for the stellar types considered in the spectra grid. In this way, we are informed on which lines are been used for each parameter under consideration, in a more physical approach close to the traditional spectral synthesis.

The capabilities of MATISSE to accurately derive stellar atmospheric parameters and chemical abundances (as the [α/Fe] ratio) are described by [RBBdL06]. The method gives rapid, compelling and stable results, with negligible biases, even for moderate to low signal-to-noise spectra and flux normalized data (see also Fig. 41). The stable performances for stars in different regions of the Hertzsprung-Russel diagram and the applications to chemical abundance determinations make of MATISSE a powerful tool for the study of stellar populations. In particular, the accuracies attained on chemical abundances are better than about 0.1 dex (except for very low-metallicity hot stars with too few available lines), as required to constrain Galactic formation and evolution models.

Concerning the computational time needed to derive the atmospheric parameters of an unknown star, the efficiency of the MATISSE algorithm is very high. Indeed, once the $B_\theta(\lambda)$ have been derived for different locations of the HR diagram (or subgrids), the stellar parameters are almost instantaneously derived from (26). For each parameter, only a multiplication of two vectors with dimension equal to the number of sampling elements in the spectra has to be carried out. As a consequence, the atmospheric parameters and the [α/Fe] content of the whole Gaia/RVS spectra database could be evaluated in a few hours, with only one present day processor.

5.5.7 QSO and galaxy classification

5.5.7.1 QSOs

Gaia will provide astrometric and photometric (colours + variability) information for about 500 000 QSOs distributed over the whole sky. Besides their own interest and their use in various cosmological applications, QSOs are crucial targets to fix the Gaia Celestial Referential Frame (GCRF).

There are three objectives for the QSO classification: i- getting the *cleanest* QSO sample to determine the GCRF; ii- deriving the *most complete* QSO sample based on the full Gaia data; iii- determining astrophysical parameters (APs) for each QSO. The two first tasks are associated with the DSC (see Sect. 5.5.4).

Regarding the first classification task and using the Gaia-2 design, synthetic QSO and stellar spectral libraries and the Besançon galactic model, the study by [CSVS06] shows that, based on the end-of-mission colour information, supervised Artificial Neural Networks (ANNs) can virtually reject all contaminating stars (including white dwarfs), although the completeness drops to about 20% at V = 20.

The second classification task (i.e. the *most complete* QSO sample) will be reached by relaxing some photometric criteria once the colour information is complemented with proper motion and variability indices obtained at the end of the mission. Adding observed QSO spectra to the synthetic spectral library will also improve the efficiency of the algorithms (e.g. finding Broad Absorption Line QSOs).

Variability is a useful and important criterion in searching for QSO candidates. QSOs vary on time scales of several months at least, and their light curves cannot be reduced to single flares. This type of variability, especially when taking into account the properties of the temporal sampling that will be provided by Gaia, is best studied in the time domain, through methods like structure-function analysis (which is closely related to auto-correlation analysis, see [Eye02]). The colour properties of the light curves will also be used to discard spurious events (QSO variability is essentially achromatic).

An alternative approach is to use the Gaia data themselves to classify QSOs, such as the Naive Bayesian Classifier (NBC; see [Ric04]). The latter has been successfully used by the Sloan Digital Sky Survey (SDSS) to photometrically classify quasars (with 95% completeness and stellar contamination less than 5%). This technique relies on the distribution of *known* QSOs and stars in the colour space to derive the probability of an *unknown* object being a star or a QSO. The plan would be to provide incrementally better NBC quasar catalogues for Gaia as we learn more about quasar classifications from the model-based methods (which can start working immediately). An input list of known “secure” QSOs based on the SDSS catalogue, the Véron-Cetty and Véron catalogue and VLBI/VLA calibrators list could be used. Later on, proper motion and variability indices could be included as new priors (besides the magnitude and galactic latitude).

Finally, a first study of the QSO AP determination has been made with the Gaia-2 design by [CSVS06]. It showed that the nearest neighbour algorithm is able to retrieve the QSO *photometric redshifts* with a median absolute error varying from ~ 0.2 to less than 0.01, depending on the strength of the emission lines and of the true redshift z_{spec}. The largest errors are expected in the range $0.5 < z_{\text{spec}} < 2$, where known degeneracies between emission lines arise. A method based on the *Spectral Principal Components* has also been tested and found promising to recover the redshift and even the spectral shape of objects with high S/N. Other QSO APs will also be estimated, although probably with lower accuracy: the emission line strength, the continuum slope (provided the extinction is known), and perhaps some indices describing the presence of Broad Absorption Lines (to be checked in future studies).

These techniques will first be tested with the final Gaia design and then explored further. This could include incorporating external UV data, e.g. from the GALEX satellite.

5.5.7.2 Galaxies

During the five year Gaia mission, we will be able to observe nearby unresolved galaxies all over the sky. Although the primary goal is the stellar content of our galaxy, there remains a lot of important science to be extracted from the large number (several million) of unresolved galaxies which Gaia will observe.

Currently, although there are several surveys of galaxies, even SDSS – one of the largest photometric and spectroscopic galaxy surveys – does not cover the whole sky. This gives the Gaia mission great importance:

- Gaia will be able to detect about 10^7 unresolved galaxies down to G=20.
- Gaia will provide the first homogeneous survey of galaxies covering the whole sky since the photographic ones (UK, ESO, Palomar Schmidt surveys, 3500 to 6500Å) from 30 years ago.
- Photometry at a more extensive spectral range than the previous ones (3300 to about 10000Å) will be obtained providing magnitudes at about 62 points at this range.
- The mission will permit us to investigate all sorts of variabilities in all galaxy types. Apart from the well known classes of AGNs and QSOs, this presents a unique opportunity to observe galaxy variability in such detail.
- Definition of the galaxy density of the Local Universe and the scale of its variations will be possible.

The objective is to study, develop and test algorithms which provide optimal parameter estimates for unresolved galaxies, based on the assumption that the object is restricted to this class (based on probabilities provided by the Discrete Source Classifier described in Sect. 5.5.4. For this we perform the following steps:

- Provide libraries of galaxy spectra
  An extended grid of synthetic galaxy spectra has been created, using the code PEGASE.2 (www2.iap.fr/users/fioc/PEGASE.html), which is based on models of galaxy evolutionary synthesis and the BaSeL library of stellar spectra. The library contains a random grid of 2700 spectra at redshift zero, in the wavelength range 250 to 1050 nm and at one nm or less resolution, and covers the main Hubble types of galaxies. It is computed on a regular grid of four key astrophysical parameters for each type and for intermediate random values of the same parameters. In addition, a regular grid of 888 spectra has been produced for various redshifts. This synthetic library has been compared with real spectra obtained from SDSS. From two-colour diagrams computed from both the synthetic PEGASE.2 and the real SDSS spectra we see good agreement over the full range of galaxy types [TKK+06] [TKBJ+07].

- Simulations
  Gaia instrument simulations of the selected galaxy synthetic and real spectra
- Parametrization – Classification
  Use of known statistical methods (e.g. support vector machines, neural networks etc) to identify and classify the unresolved galaxies into groups (i.e. according to morphological type) and to determine the key astrophysical parameters. The first tests performed with the library of spectra described above are very promising [TKBJ^{+}07].

5.5.8 Minor planets classification

5.5.8.1 Introduction

The utilized method for taxonomic classification of minor planets will be based an evaluation of asteroid RP/BP spectrophotometry [WL06]. Asteroid colours from the Eight-Color Asteroid Survey [ZTT85] and CCD spectra from the Small Main belt Asteroid Spectral Survey II of [BB02b] are used as input. The success of the Gaia-2 photometric systems C1B and C1M ([JHBea06]) for taxonomic classification has been evaluated using supervised classification techniques and mean taxonomic class spectra in the Tholen [Tho84] and Bus& Binzel [BB02a] taxonomic systems. These results using C1M and C1B are described here, but these, and the newly derived taxonomic spectra should be equally applicable to the new RP/BP spectra.

The tests showed that the Gaia photometric systems are able to discriminate between all of the twelve Tholen asteroid classes for noise-free data. With the most successful supervised classification parametrization, both the Tholen [Tho84] and Bus & Binzel [BB02a] taxonomies are found to be internally inconsistent in a robust best-fit sense such that about 25% of the asteroids are more similar to another taxonomic class. New consistent classifications for 531 ECAS and 1328 SMASS asteroids, with corresponding probabilities for the three most likely classes, have been produced for simulation purposes.

5.5.8.2 The need for a new Gaia taxonomy

The Gaia photometric system is adequate for taxonomic classification in the Bus & Binzel taxonomy. The classification method employed here results in homogeneous taxonomic class domains in principal component space and mean taxonomic class spectra that are derived from less divergent class member spectra, and naturally reflect the fact that asteroid spectra form a continuum in principal component space.

One of the weaknesses in the Bus & Binzel (as well as the Tholen) taxonomy is that it assigns a specific taxonomic class to an asteroid without recognition of the probability of membership in the given class. This is a result of the method of classification, based on well defined but more or less arbitrary cluster domains in (mainly) the

Slope–PCA2′ space ([BB02a]). Asteroids having spectra that do not fall within the defined cluster domains remain unclassified, with no implicit indication of the possible relationship to any of the defined taxonomic classes. The resulting taxonomic class for any individual asteroid thus does not reflect the probability with which the asteroid is assigned membership, as the probability is set implicitly to 0 or 1. These obstacles are circumvented in the new Gaia taxonomic classification method.

5.5.8.3 Method

A number of spectral parameterizations have been tested to identify a supervised classification method which produces classification results most similar to those obtained by Bus & Binzel [BB02a], and which naturally produces a probability estimate for each individual class assignment. This is possible by utilizing an approach based entirely on the mean taxonomic class spectra established by Bus & Binzel, rather than assigning individual memberships from the location relative to PCA cluster space domains. As such, the method improves the accuracy of the classification and assures self-consistency in the class assignment. It also reflects the natural continuity between the different taxonomic classes, expressed in terms of the probabilities of membership which may be high for more than one class.

The discriminative potential of a number of spectral parameterizations quantified with parameters P_n, $n = 1, \ldots, 8$, derived from mean taxonomic and individual asteroid spectra transformed to the Gaia systems have been tested. The most successful method utilizes the total RMS difference parameter P_1 based on relative spectral intensity:

$$P_1(j,k) = ((1/N)\sum_{i=1}^{N}(I_{\text{tax}}(i,k) - I_{\text{ast}}(i,j))^2)^{1/2}$$

where I_{tax} and I_{ast} are, respectively, the intensities of the mean taxonomic and the asteroid spectra in each of the spectral bands $i = 1, 2, \ldots, N$ (5for C1B and 9 for C1M), k is the taxonomic class, and j is the asteroid.

For each difference parameter P_1, the ratio $R(j,k)$ of the parameter value of the asteroid j relative to the standard deviation of the parameter values for all classes k was calculated,

$$R(j,k) = P_1(j,k)/\sigma([P_1(j,1), P_1(j,2), ..., P_1(j,k)]),$$

and used as input to the normal cumulative distribution function

$$p(j,k) = F(R(j,k)|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{R(j,k)} e^{\frac{-(t-\mu)^2}{2\sigma^2}}\,dt,$$

which gives the probability that asteroid j is a member of class k. Here, p is the probability that a single observation from a normal distribution with mean μ and standard deviation σ will fall in the interval $(-\infty R]$. In the present case, $\mu = 0$.

The success rate $\rho(k)$ for each class k is then evaluated with respect to the classification of Tholen or Bus & Binzel and calculated as

$$\rho(k) = N_{\text{tax}}(k)/N$$

where N_{tax} is the number of asteroids for which the most probable taxonomic class k coincides with the Tholen [Tho84] or Bus & Binzel [BB02a] class assignment for any given asteroid, and N is the total number of evaluated ECAS or SMASSII asteroids.

5.5.8.4 Results

Of the 26 original classes in the Bus & Binzel taxonomic system, it was possible to calculate 23 revised mean taxonomic spectra for the full Gaia spectral range by employing spectral information from ECAS objects present in the SMASSII survey. The 23 mean taxonomic spectra are the basis for assigning photometric observations of individual asteroids to taxonomic classes, based on the P_1 parameter.

The cumulative success rates for the three most probable taxonomic classes evaluated with the P_1 difference parameter have been evaluated in the C1B and C1M systems. The success rates for the most probable classification are low, 70–83% for the Tholen taxonomy and only 45–62% for the Bus & Binzel taxonomy. For the C1M system, the most probable classification is different from the nominal Tholen and B&B assignments for 17% and 41% of the asteroids, respectively. However, as many as 97% and 87%, respectively, of any of the three most probable classifications is the same as the nominal classification. This is a dramatic indication of the small spectral differences between many of the taxonomic classes as represented in the C1M system, particularly in the B&B taxonomy.

5.5.8.5 Observations

Low-resolution spectroscopic observations of a selected sample of asteroids are required in order to define all mean taxonomic spectra over the full RP/BP wavelength range at a photometric precision comparable to that of Gaia. 15–20 asteroids in each class should be enough to improve the photometric quality of the present mean spectra and to define good taxonomic spectra for presently poorly observed classes. We estimate that 20–30 nights are required with a 2.5m class telescope. This work will be initiated in 2007.

5.6 Ground based observations, external calibrations

5.6.1 Statement of the problem

Although Gaia is usually qualified as a self calibrating instrument, this is not fully true in photometry and spectroscopy where reference data are needed to calibrate the wavelengths or to fix the zero-point of the magnitude system. Even in astrometry an initial attitude catalogue is required to compute a zero-order attitude to improve quickly on the low accuracy on-board attitude. Solar system ephemeris must be available to model the aberration or scale the parallactic factor and a good database of minor planet orbits will also make the cross-matching of moving bodies much easier. Finally the Gaia orbit, which we cannot dispense with, will be obtained in near real time from dedicated ground based tracking by ESOC. All these additional data, so important for the data processing, are termed auxiliary data and should be available at the start of the Gaia processing. In addition a monitoring of the most important sources will be performed during the mission to check for possible long term instability or drift. While many of these data do exist already and must only be compiled from existing archives and put into the appropriate form, this is not true for several pieces of photometric and spectroscopic reference data, at least with the accuracy required for Gaia. These data must be acquired from new observations.

To this aim a coordinated programme of ground-based observations is being organized in order to obtain the necessary auxiliary data to be used primarily for (i) the absolute flux calibration of BP/RP and G band, (ii) the definition of the wavelength scale and zero point of RVS radial velocities, (iii) the training and calibration of the classification / parametrisation algorithms.

The coordinated effort within DPAC described in this section does not cover all the Gaia related ground based observations, but only those related to the calibrations needs and which must be undertaken quickly. Additional observations will be taken up at the CU level outside this coordination.

5.6.2 The requirements

They have been set in consultation with the relevant CUs and are now in a fairly advanced state.

5.6.2.1 Photometry

For the BP/RP and G flux calibration in CU5, the general requirements have already been studied in details [BBF$^+$06], [CJF$^+$06]. Briefly, about 100 spectrophotometric standard stars (SPSS) are needed, mainly white dwarfs and subdwarfs with weak spectral features, in the magnitude range $10 < V < 15$, have to be observed in absolute spectrophotometry covering the range 330-1050nm with a resolution R=500-

1000 at S/N=100. This is a very challenging demand hard to fulfill in its entirety. Fortunately it can be complemented with an alternative method, including low resolution spectroscopy + photometry with 1% accuracy in flux covering UVBRI. A pilot program has been submitted to test the procedure and methods of observations and the quality of sites [FBD$^+$06]. Brighter stars may be needed to test the gating, as well as fainter stars.

The wavelength zero-point will be derived from sources with identifiable narrow spectral features, like H_β, H_α lines, emission lines of WR stars, QSOs and so on. While the wavelengths of the emission line stars are well known, the lines of QSOs depend on their redshift and so wavelength calibrated spectra of a set of QSOs are required. The number of required QSOs and brightness is still to be assessed as well as the existing data in the literature. The need of additional observations cannot be ruled out.

5.6.2.2 Spectroscopy

The RVS being an integral field spectrograph with no entrance slit and no on-board wavelength calibration lamp, the wavelength scale and radial velocity (RV) zero-point have to be derived from reference sources, stars and asteroids. A stellar grid is being defined, consisting of 1000 to 2500 FGK stars in the magnitude range $6<V<10$. These primary stars are to be stable in radial velocity at the 300 m/s level with no drift. A preliminary list of candidates have been established from recent RV catalogues. Each will have to be observed few times in the interval from now to the late mission phase in order to verify their long term stability. High resolution echelle spectroscopy with a 2 meter telescope is well adapted for this task which does not require a high signal to noise ratio. Secondary standards have also to be considered to cover fainter stars and other spectral types. Asteroids are excellent sources to calibrate the zero-point because their radial velocity can be computed with uncertainties below 1 m/s, but there are very few bright ones and they have a low sky coverage. They will be observed together with stars to study possible systematic effects.

5.6.2.3 Parametrisation

The external calibration of the classification / parametrisation algorithms like (i) Discrete Source Classifier(DSC), (ii) General Stellar Parametriser in photometry (GSP-phot) or spectroscopy (GSP-spec), (iii) Extended Stellar Parametrise (ESP) requires to build an extended grid of reference objects that will be observed in good conditions by Gaia, and that will have their properties well determined in advance. This grid must be representative of all objects that will be observed by Gaia and classified through the DSC. GSP-spec will deal with bright stars ($6 < V < 10$), GSP-phot with fainter stars ($10 < V < 18$) and ESP with both, so that a bright and a faint stellar grid have to be built. If bright reference stars can be selected from existing archives and catalogues, a large observing programme has to be organised to build the faint grid from stars in open clusters, globular clusters or test fields. The precise determination of their astrophysical parameters (AP), like effective temperature, gravity, rotation, chemical composition and interstellar extinction, requires a sufficient spectral resolution, while the magnitude of the targets requires a large telescope. The grids will also be used before launch as training data to make simulations and to test algorithms from realistic data. Finally this CU8 task needs high resolution spectroscopy to characterize AP reference stars, with also medium resolution in RVS range and spectrophotometry of reference stars as training data.

5.6.3 The observing programmes

5.6.3.1 What to observe

An observing plan has been drawn that could meet all those requirements, in the most efficient way, without duplication of efforts. This plan comprised three parts :

1. spectrophotometry or low resolution spectroscopy + photometry or combination of both in order to obtain SEDs (Spectral Energy Distribution) of SPSS (Spectrophotometric Standard Stars) and AP (Astrophysical Parameters) reference stars ($10 < V < 15$). Several well suited instruments have been identified at ESO, Canary Islands and Calar Alto (ALFOSC, IDS, ISIS, DOLORES, EMMI, DFOSC, EFOSC2, FORS2, TWIN, MOSCA, CAFOS).

2. high resolution spectroscopy, covering the RVS range, of RV and AP reference stars ($6 < V < 10$). Well suited instruments are NARVAL at Pic du Midi Observatory and FEROS at ESO.

3. high resolution spectroscopy of faint AP reference stars ($10 < V < 18$). The only well adapted instrument is UVES-FLAMES.

This is clearly a long term programme. SPSS must not vary photometrically over the 5 years of the mission so that a few observations of each must be obtained over the years of preparation. Similarly the radial velocity reference stars must be stable over the same period and variations higher than 300 m/s must be tracked. Long term observations on AP reference stars correspond to the highest number that has to be obtained to cover the AP space with a density corresponding to the final precision of the parameters that is expected from GSP-phot, GSP-spec and ESP.

5.6.3.2 How to coordinate the observing programmes

There are many archives, databases and catalogues providing high quality data that will be useful as a starting point for the calibrations. However these data must be completed to fit the specific magnitude ranges of RVS and BP/RP, to correspond to the variety of objects that Gaia will observe, to guarantee the homogeneity of the calibrations. A way to meet this requirement is to limit the number of different instruments used to carry out the observations.

Obtaining telescope time to get calibration data from instruments that have been built and paid for to produce (quick) science returns is not an easy task. The group involved in the calibrations is aware that allocation committees and/or directors of observatories may be reluctant to allocate time on a programme that will not produce immediate science. For this reason an agreement should be negotiated between the institutes or the authorities responsible for the facilities and the DPAC. Serious arguments may be put on the table by the Gaia community in these discussions. These observations are absolutely needed so that all the Gaia potentialities are exploited to produce a completely new science that will benefit to a wide community. Moreover the availability of these high-quality and validated calibration data in a format that is easy to use, through the VO for instance, will have an immediate impact on other projects, in space or on ground.

Other ground-based observations, which are not described here, will have to be made before launch or during the mission. Such observations are not directly related to external calibrations but they will either facilitate the data processing or give a higher scientific value to Gaia data. As part of the object analysis, a network will be organized to obtain CCD astrometry and photometry with two main objectives : follow-up and alerts on fast moving solar system objects and mass determination of asteroids. Classification includes some projects that concern the preparation of data processing of colour variations before launch, the quality assurance of interesting objects before the release of the catalogue, the follow-up on Gaia observations during and after mission, variability monitoring.

6 Simulations

This section describes the preparation of the simulations needed for supporting the data processing effort.

6.1 Introduction to the data simulation

6.1.1 Overview

An essential part of the preparation of the data reduction for Gaia is the availability of realistic simulations of the mission data. Ensuring that reliable data simulations are available for the various stages of the data processing development is essential to guarantee that the algorithms are fully developed and properly tested in time for the mission launch.

The approach chosen for the Gaia DPAC is that the provision of simulated data will be organised in a centralised way (that is, the development of "home-grown" simulations for specific purposes will be avoided) in order to ensure the consistency of all the results and tests. A dedicated team will assume the responsibility of this provision in close coordination with the rest of the DPAC (Fig. 42): requests for simulated data will be sent by the different teams developing the data processing system and from them an agenda for the development of the different modules of the simulator in each cycle will be jointly agreed, taking into account all the needs, their global priorities within the overall consortium, the availability of the simulation models (which can depend on industrial tests or available algorithm accuracy models) and the available manpower.

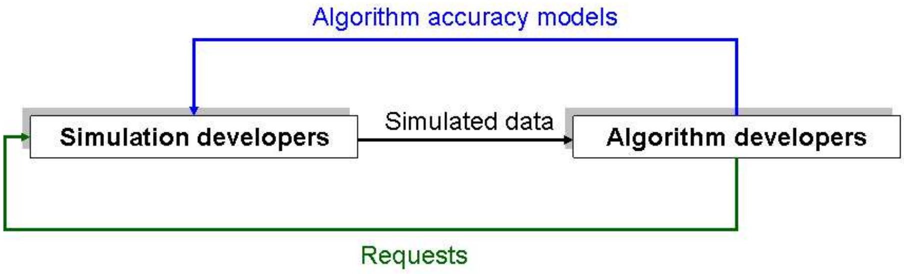


Figure 42: Development cycle of the Gaia simulations

It is important to note that the preparation of the reduction algorithms will not only require raw data (pixel data or telemetry data) but also several levels of intermediate data, e.g. data that have already received some processing. To avoid going through the whole chain (raw data telemetry, database ingestion, core processing, etc.) intermediate data simulation will be provided using models of the intermediate data accuracy. For the algorithms developers should provide accuracy models based on their reduction results (see also Fig. 42).

6.1.2 Organisation

To cover the needs described above, a software system capable of generating the varied and complex simulated data required for DPAC has to be developed. For this, the team in charge of simulations will need to be constituted with a strong software engineering base, able to properly handling the development of such a complex system in a professional way. However, this base alone is not sufficient for the task ahead; a strong scientific component is also needed to ensure that the system fulfills the scientific needs of the DPC, which has to carry on an essentially scientific task, the reduction of the Gaia data and the production of the Gaia Catalogue.

Therefore, the team structure should reflect this dual nature and integrate:

- A core software engineering team able to properly and professionally manage the development of a complex software system, the Gaia simulator. This team should be mainly constituted by software engineers but should also include the appropriate scientific expertise to ensure the proper scientific management of the development and the coordination with other DPAC teams.
- A scientific team able to gather and integrate the expertise for the development of a Universe Model to be used in the Gaia simulator, working in close cooperation with the core team of software engineers and acting as an interface with the wide Gaia scientific community.
- A scientific team providing the expertise to develop models of the Gaia spacecraft and its instruments, working in close cooperation with the core team of software engineers and the industrial teams in charge of building the satellite.
- A Quality Assurance and Validation team (QA&V) ensuring the quality of the simulated data and its fitness for the intended purposes.

The simulation of data for the DPAC will therefore be organised around these lines, as depicted in Fig. 43.

6.1.3 Requirements for the data processing

The generation of simulated data is an activity closely tied to the needs of the rest of the DPAC teams. As the simulation needs in a given development cycle will substantially depend on the schedule and development status of the rest of the consortia,

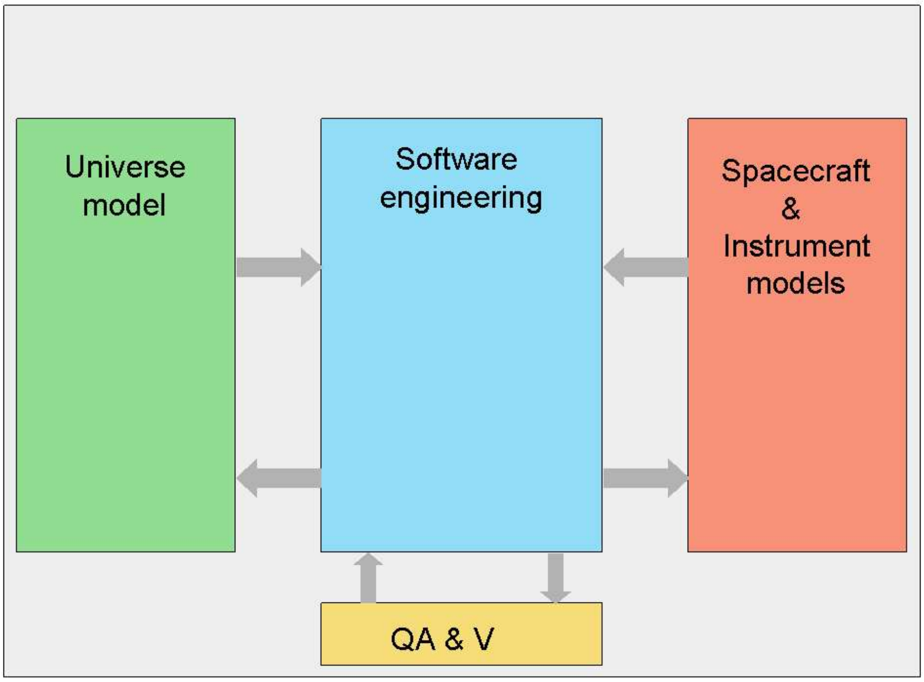


Figure 43: Organisation of the simulation development

fixing the priorities and schedule for simulations from the start would be exceedingly rigid and probably unrealistic. But, at the same time, clearly some long-term planning for the development of the simulations is required.

In order to try to conjugate these two conflicting needs, the planning of the simulation development and the release of the simulated data will be adapted to the DPAC development cycles; in each cycle the short term planning for the next cycle will be fixed and the long term planning for the overall simulation activities will be reviewed, in both cases from the inputs provided by the rest of the teams in the DPAC and the past experience of the team in charge of the simulations. Essentially, the steps to follow will be:

A Initial planning

B Send request to DPAC teams to provide inputs on simulation needs for next cycle *and* review of long term needs for following cycles.

C Review inputs and prepare consolidated list of needs. Prepare prioritized proposal of developments for next cycle, according to available resources.

D Submit proposal to DPACE for review and acceptation

E Implementation of simulations from priorities for current cycle and delivery of simulated data

F Review long term planning from inputs and past experience

G Iterate from B

Another important factor to take into account when organising the design and delivery of simulations is the coordination of the availability of simulated data with the development cycles of the rest of the consortia. Obviously the simulated data for a given cycle can not be delivered at the end of the cycle, but rather in time for the testing phase of the algorithm development. This implies that the simulation development cycle should be out of phase with respect to the cycles of the rest of the consortia.

A final factor to take into account in the design of the simulations for the DPAC is that they should cover the needs of a large and varied group of teams working on very different problems. The needs of these teams will be coordinated through the DPACE with the process described above, but in such a situation one should avoid developing many unconnected simulation tools covering specific needs because this approach would be inefficient and may lead to inconsistencies between the different tools. Instead, a more centralised but flexible approach will be used, based on four components:

1. A core library implementing all the models and tools needed for the simulations, including models of the observable objects (Sect. 6.2) and the spacecraft and instruments (Sect. 6.3).
2. A telemetry simulator (Sect. 6.4) to cover the need for simulated flight-realistic Gaia telemetry data
3. A pixel level simulator (Sect. 6.5) to cover the need for very detailed simulations of Gaia observations
4. An intermediate data simulator (Sect. 6.6) to cover the needs for simulated data mimicking the one that will be exchanged between the data processing centers (intermediate data)

6.2 Universe modelling

6.2.1 Introduction

The building of a Universe Model is necessary to perform the simulations for the Gaia preparation. This Universe Model is a set of algorithms for computing the positions and characteristics of any objects expected to be observed by the instruments.

The distributions of these objects and the statistics of observables should be as realistic as possible for simulations to be usable for estimating telemetry, testing software, simulating images, etc. The algorithms have to be optimised in order that the simulations can be performed in reasonable time and can be redone when necessary. The complexity of the model is expected to increase during the preparation of Gaia.

Objects which are to be simulated are : solar system objects (planets, satellites, asteroids, comets), galactic objects (stars, nebulae, stellar clusters, diffuse light), extragalactic objects (galaxies resolved in stars, resolved galaxies, quasars and active galactic nuclei, supernovae). For each of these simulated objects one needs to have their full 3D spatial distribution together with their spectral characteristics (to be able to compute photometry and spectroscopy), and their motions (for spectral corrections and for astrometric computations).

Two strategies are envisaged for object generation, depending on the specific needs:

- Computation of static catalogues : The simulated sources representing those to be observed by Gaia are generated once and stored onto a disk.
- On-line computation : Sources are generated on demand using random generators. Selection of the seeds are controlled in order that subsequent simulations generate the same sources.

In order to simulate the sky content, the sky is subdivided in smaller regions with a Hierarchical Triangular Mesh (HTM). In the following we describe each type of objects with their relevant characteristics to be computed.

6.2.2 Solar system objects

Solar system objects model would include both real objects with ephemeris (sun, moon, planets and their satellites, minor planets and known big asteroids, known periodic comets) and simulated objects (smaller asteroids, new comets, Kuiper belt objects and centaurs).

It is envisaged that both ephemeris of real objects and simulated ones will be stored on disc, as the number of objects to be simulated is relatively small (of the order of 20000). With a database containing those objects that cross the instruments field of view during the mission, the simulator will search the data base and return those objects which transit during the time interval simulated.

The simulated objects will have the spatial distribution on the sky which mimics our present knowledge of the Solar System. Their spectral, astrometric and photometric characteristics will also be modelled according to it. Realistic simulations of orbits have already been performed [Mig01a, Mig01b] for 20000 NEOs, with magnitudes including rotation-phase effects.

Spectra will be prepared in the framework of the spectral library developed for all kinds of objects by CU8.

6.2.3 The Galaxy

Gaia will detect and measure about 1 to 1.5 billion objects that are part of the Galaxy, providing the main contribution to the object database. A reliable model of the stellar distribution over the sky up to magnitude 20 is most important, with reliable spectral distribution and motions for these objects. These are the first objects to be simulated, as they are simple (point sources) and the first on which reduction algorithms are to be tested. A preliminary version of a Galaxy model has already been implemented in java, based on the Besançon Galaxy Model [RRDP03], here after BGM), and the 3D extinction model from [Dri02]. Here we recall the main ingredients in these two models.

The stellar population synthesis model of the Galaxy constructed in Besançon since the 80's is able to simulate the stellar content of the Galaxy by modelling four distinct stellar populations: the thin disc, the thick disc, the outer bulge and the spheroid. It can be used to generate stellar catalogues for any given direction, and returns information on each star such as magnitude, colour, and distance as well as kinematics and other stellar parameters.

The approach of the Galactic model is semi-empirical as it is based on theoretical considerations (stellar evolution, galactic evolution and galactic dynamics) but is constrained by observations (the local luminosity function, the age-velocity dispersion relation, the age-metallicity relation). The Galactic potential is calculated in order to self-consistently constrain the disc scale height, the thin disc being subdivided into 7 isothermal components of ages varying from 0-0.15 Gyr for the youngest to 7-10 Gyr for the oldest. For computing the scale height as a function of age, the Boltzmann equation (first moment at the first order with the plane parallel approximation) is used assuming an age-velocity dispersion relation deduced from Hipparcos observations [GGU^{+}97].

The distribution in the Hess diagram split into several age bins is obtained from an evolutionary model which starts with a mass of gas, generates stars of different mass assuming an Initial Mass Function and a star formation rate history, and makes these stars evolve along evolutionary tracks. The evolution model is described in [HRC97]. The evolutionary model produces a file describing the distribution of stars per element volume in the space (M_V, $\log \mathrm{T}_{\mathrm{eff}}$, Age). Similar Hess diagrams are also produced for the bulge, the thick disc and the spheroid populations, assuming a single burst of star formation and ages of 10 Gyr, 11 Gyr and 14 Gyr respectively.

In order to compute the N number of stars at any point in the Galaxy of a given population we make use of the equation of stellar statistics :

$$N = \rho(r) \times \Phi(\mathrm{M}_V,\ \log \mathrm{T}_{\mathrm{eff}},\ \mathrm{Age})\,\omega r^2 \mathrm{d}r$$

where ρ is the density law of the population. N is the theoretical number of stars in a volume element with the intrinsic parameters M_V, $\log \mathrm{T}_{\mathrm{eff}}$, and Age. In order to simulate catalogues, from this number a random drawing is performed to produce an integer number of stars, which number might deviate from N due to Poisson noise but which expectation is N.

Metallicity [Fe/H] is computed through an empirical age-metallicity relation. Alpha elements can also be modelled assuming that for a given population, at a given distance from the Galactic centre one can define a probable α/Fe ratio as explained in [JFC^{+}02].

In the framework of the Gaia simulator, only the distribution of the stars and their intrinsic parameters are obtained from the BGM. Transformations to apparent magnitude, colours and spectra for the Gaia instruments are done using the Gaia spectral library and the Gaia instrument model. However some problems not addressed inside the BGM scheme have to be solved in the Gaia Simulator framework :

- Binaries : the BGM model produces only single stars. Simulations of binaries have been introduced in the Gaia simulator in the following way (see [BAC05]). For each single star, a companion is created with a probability depending on the spectral type of the primary. The distribution of secondaries in separation and mass ratio is taken from [Sod04]. The orbits are computed, the positions of both components are modified and astrometric and photometric effects taken into account. Eclipses are also produced. This treatment has an effect on the total mass of the Galaxy : the secondaries being added to the single star populations the total mass in stars is larger by several tens of per cents. This will be corrected in the near future by normalising the evolutionary model no longer on the single star luminosity function of the solar neighbourhood, but on the luminosity function of primaries (from the CNS3).
- Variability : variability is taken into account by introducing several variability types (about 15) for major known regular types (Cepheids, RR Lyrae, delta Scuti, etc.) and semi-regulars. The process has been started and 1 type has been introduced in the simulator [ERER05]. The variability probabilities depend on the position in the HR diagram and variable characteristics are computed randomly using a light curve modelled by a Fourier decomposition which allows asymmetric light curves. Period and amplitude are taken randomly from a 2D distribution defined for each variability type. A unique seed is used for the generation in order to ensure that at each simulation the same stars are found to be variable with the same characteristics. No colour effect is introduced yet but it will be included if necessary. Multi-periodicity, semi-regular and irregular variability can be modelled as well, but the computational effort may not be feasible for whole sky simulations, and would be limited to smaller-size simulations. Microlensing effects will be also introduced in the simulator in the future.

Extra-solar planets will also be simulated in the same framework as binaries. Exoplanets can be considered as small secondaries with specific spectra assigned to them, included in the Gaia spectral library. One may also consider to simulate field exoplanets if the need is raised by the data processing developments.

The extinction model used at present in the Gaia Simulator is described in [Dri02]. It is based on the dust distribution model of [DS01] with improvement with respect to the correction based on FIR data and on spiral arm geometry. This full 3D extinction model is a strong improvements over previous generations of extinction models as it includes both a smooth diffuse absorption distribution for a disk and the spiral structure and smaller scale corrections based on the integrated dust emission measured from the FIR. Some uncertainties remain in the way the correction is applied, the choice of which component should be corrected not being trivial. In the future alternate models will be evaluated and tested to improve the quality of the simulations specially in the galactic plane.

6.2.4 Reliability of simulated star counts

At bright magnitudes ($G < 12$) star counts from real observations are more reliable than model predictions, because they are more realistic by nature and do not suffer from extrapolations and model defects. Therefore, real counts have been used in the prediction of telemetry rates for mission designs, based on GSC-II data and partly from space data (Hipparcos and Tycho2). However, for general simulations purposes the star counts generated by the Besançon Galaxy Model are used because they provide supplementary (simulated) information not available from observed data.

For fainter stars, several tests have been performed in order to ensure the reliability of model predictions. First, a detailed comparison of predicted star counts in the G band has been done between the Fortran version (original code) and the java version (Gaia simulator implementation) to ensure the robustness of the coding. Second, the star counts have been compared with GSC-II star counts transformed into the G band over all sky [DBL^{+}03]. From this comparison the reliability of the predicted counts is assessed to be at the level of 15% at medium and high galactic latitudes. In the galactic plane uncertainties are larger due to the incomplete knowledge of the extinction distribution and the fact that photographic surveys suffer from high crowding at faint magnitudes. Similar tests have also been used to compare with star counts in the near infrared at low latitudes on which the impact of assumptions on the extinction is lower. However the reliability of the counts at the small scale are still uncertain at a level of 50% in the galactic central regions, but at larger scale (degree) the counts can be considered reliable enough for estimating the telemetry. In the future we shall consider to increase the reliability of the counts in the galactic plane by improving the extinction model.

6.2.5 Open and globular clusters

Simulation of open and globular clusters is done with a similar approach as the Galaxy model: using stellar models of given age and metallicity, the luminosity function is computed for a given IMF and evolutionary tracks. Spatial densities can be computed by a King model for the globulars and ad hoc models for open clusters. The space distribution of these clusters can be estimated by analysis of the real cluster distribution and will probably follow the spatial distribution of the field stars of similar ages. We might envisage to introduce also mass segregation effects inside clusters if necessary. A catalogue of known clusters could also be introduced in the simulator if necessary.

6.2.6 Extragalactic objects

Gaia will observe more than a million Galaxies [Vac02]). Nearby galaxies resolved in stars can be simulated either following the simulation approach for of star clusters or by obtaining real data. The latter alternative can be specially appropriate for the Magellanic Clouds, where the number of stars will be large and the background will severely perturb the measurements (particularly in numerous diffuse nebulae). This can be introduced as a test for reduction algorithms.

For galaxies not resolved in stars, planned simulations will account for extension, which can be simulated most of the time by the sum of a disc and a bulge. The scheme is primarily based on the STUFF code from E. Bertin adapted to Gaia by C. Dollet, IRAF based galaxy profiles and SKYMAKER. Spectral type will be defined and the spectral library developed for classification will be used to compute photometry and spectral distribution. Distribution in space and red-shift will be simulated to mimic the real sky, accounting for the most up to date cosmological parameters, and spectra corrected accordingly. Gravitational lensing effects and clustering must also be taken into account.

Special attention will be put onto QSOs simulations , as they will be used for determining the reference frame. QSOs distribution must include the clustering and lensing effects which render the spatial distribution far from uniform. Spectra will also be taken from the Common Spectral Library prepared for classification.

Supernovae events will also have to be taken into account. They will be considered as random events with realistic probabilities according to the type of the underlying galaxy. They can be simulated on top of their host galaxy or directly in the field, for simulating cases where the host is not detected by Gaia.

6.2.7 Relativity model

The simulations must fully take into account the general theory of relativity. Description of the relativistic reference systems and the structure of the Gaia Relativity Model are given in section 5.1.1.

6.2.8 Backgrounds

Backgrounds will play a significant role in the reduction process, perturbing the detection of objects and biasing their measurements. Three types of backgrounds should be simulated in order that reduction algorithms can be optimised for such cases: the zodiacal light, galactic backgrounds (diffuse nebulae, planetary nebulae, HII regions) and extragalactic backgrounds. The latter should be small and is a rather secondary effect. The simulator implementation of backgrounds will probably allow for small scale specific simulations to be performed upon request with variable backgrounds (for galactic and extragalactic) while it should account for the zodiacal light as a non-optional feature.

6.2.9 Radiation environment

The radiation environment will have effects on the CCD aging. The instrument model (DU4) will simulate the damage caused by the radiation environment. However a model for the expected radiation environment of the satellite during the mission has still to be provided. This task will be conducted by the Gaia Prime Contractor and the Space Environment Section at ESTEC. An interface between the project team and simulations will ensure that galactic cosmic ray and solar particle fluxes and spectra adopted by simulations are consistent with best knowledge at any given time.

6.3 Instrument modelling

6.3.1 Motivations

The instrument model has the goal of implementing a set of tools for the simulation of Gaia astrometric, photometric and spectroscopic data, supporting the development of the data reduction software and its subsequent usage.
Instrument response variation is unavoidable and significant, at the level of sensitivity targeted by Gaia. The instrument model goal is to include in the measurement model a detailed description of the instrument response, as a function of the hardware and operation parameters (nominal values in the Gaia Parameter Database), allowing the generation of a realistic representation of the science data, including all known contributions which might affect the performance, in terms of both noise

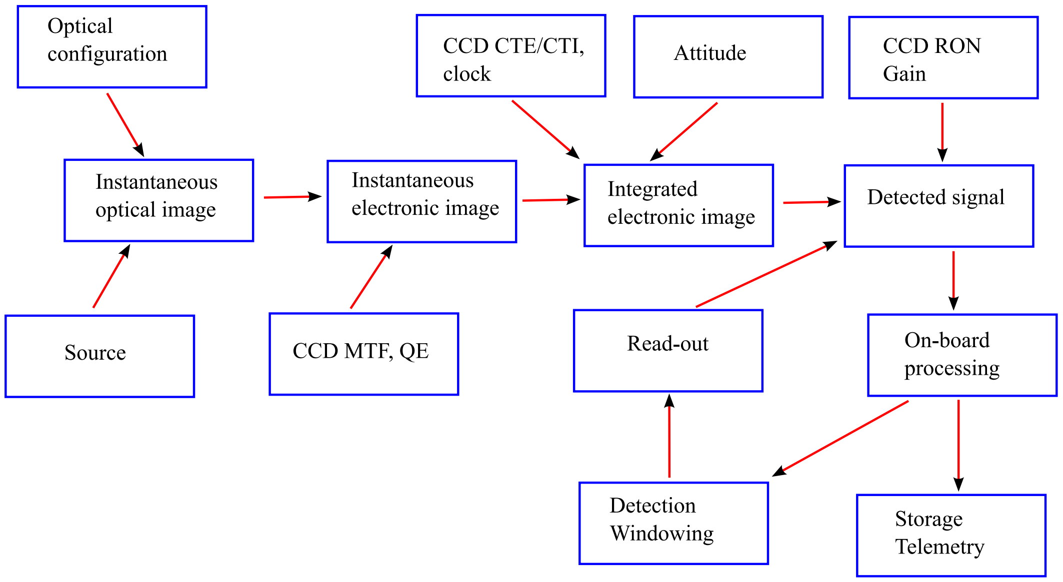


Figure 44: Schematic chart of the measurement process .

level and systematic error.

During development, the instrument model helps in the generation of realistic data sets for implementation, optimisation and validation of the data reduction algorithms. During operation and data reduction, it supports the sanity check of the instrument by comparison of the data with the measurement model, contributing to identification of effects induced by variation of critical parameters.

6.3.2 Functional description of the main contributions

The instrument model structure reflects the measurement process, briefly recalled below for clarity; a schematic diagram is shown in Fig. 44.

The radiation from a source in one of the Gaia FoVs, at a given time and wavelength, enters the corresponding telescope as a planar wavefront, and is propagated accordingly to the local optical response (which may be represented e.g. in terms of wavefront error) to the focal plane, where it generates the instantaneous optical image. This is translated into a charge distribution on the detector, depending on its geometry (alignment) and electro-optical response (pixel size, charge diffusion, quantum efficiency); the charge distribution can then be integrated over the detected spectrum, generating the instantaneous electronic image. During the elementary exposure, the source image travels over the CCD in step with the pixel clocking in TDI mode, with a mismatch associated to the differential distortion, high frequency components of attitude, vibrations, clocking error, and eventually clock discretisation, producing the integrated electronic image which is fed to the CCD readout register.
In the SM region, all pixels are read, and the detection process identifies the targets

to be followed on subsequent chips of AF, BP/RP and RVS, where only the appropriate regions of interest are digitised, stored and transmitted to ground. On-board operation has a crucial impact in the definition of the actual astrophysical sample measured by Gaia, i.e. the selection function. Close to the faint limit (and possibly in critical regions, e.g. with high stellar density), sources may be detected only in a fraction of the scans, and the readout region in a given transit may be defined with larger errors, inducing a degradation of some elementary exposures.
The output samples are recorded in terms of intensity vs. time, so that each exposure ideally provides, through the photo-centre estimate, the equivalent time of transit of each target on the detector reference position. Detector response and its variation is a key factor to the actual Gaia performance.

The μas level of precision targeted by Gaia will be achieved by the data processing (5), in which both astrophysical and instrumental parameters are estimated, and progressively adjusted, to best fit the whole set of data. Notably, such precision corresponds to knowledge of the instrument down to values (e.g. nanometres, micro Kelvin) rarely reached on ground and in just a few laboratories. Besides, measurement of the on-ground parameters is not sufficient to the Gaia purposes, because the environment does not allow such precision, and above all because several on them are likely to suffer modifications due to launch and ageing.
In-orbit variation is supposed to be fairly slow and small, to allow good estimate of parameters from the data. This is a crucial aspect to be verified in the data processing implementation, in particular by ensuring a very realistic and flexible simulation framework.

6.3.3 Basic aspects of the instrument model

Several among the instrument and operation parameters are entangled in the science data; it is not a simple task to demonstrate if, and at which level, some of them could be effectively and uniquely separated from the others in the science data reduction. A demonstration to the contrary is lacking as well. The instrument model implementation within the Gaia simulator must include all the known physical contributions to the signal, as described above and shown in Fig. 44, but, since some degeneration among parameters may be expected, a convenient formulation may be based on a set of *effective* parameters taking into account several similar contributions. For example, a description in terms of "effective" aberrations (including also non-optical contributions to the signal shape) may be flexible, because a large family of image profiles can be built. However, this approach may prove impractical because of estimate complexity, due to the non-linear relationship between aberration coefficients and the resulting image.

In principle, the instrument model, in *forward analysis*, can be used to build the science data from a realistic description of the current configuration; besides, the configuration must be retrieved from the measurements, in *backward analysis* (in particular in the instrument parameter estimate, as in 4.2, and in 5.1), and the instrument model must be able to cope with the estimate of a smaller set of relevant global parameters, since the individual component characteristics will no longer be available. This shows the need for a sufficiently flexible implementation of a common underlying physical model, to be developed with adequate care for compatibility with both development and operation stages of the mission.

The complete set of requirements for the instrument model is assumed to be progressively defined, in agreement with other needs of data reduction, and implemented in the Gaia Simulator throughout the development of the whole data reduction software, in compliance with the selected procedures, interfaces and standards.

6.3.4 Implementation

Implementation of a detailed and realistic instrument model requires a correspondingly detailed technical description of the satellite "as built", in order to include all aspects relevant to science performance. Thus, the development must retain strict connection with the project management, to ensure timely update of the most complete engineering information from industry, i.e. not only the detailed design of the satellite, but also the technical data on device characterization, on-ground instrument alignment and system integration.

A high level definition of the tasks maps the signal evolution shown in Fig. 44. Therefore, there is an optics package, further divided between non-dispersive optics (feeding the SM and AF from both telescopes); dispersive optics (feeding the BP/RP and RVS from both telescopes); a detector task (on the individual and collective properties of the CCDs); an attitude / orbit task (introducing the effects induced by the satellite real position and motion); an on-board processing task (contribution of focal plane operation). The consistency between the two optical packages, and with the other tasks, shall be ensured by appropriate definition of the interfaces and of reference cases for internal verification and cross-check.

The instrument model must also provide the description of the relationship among the main and auxiliary instruments: therefore, the optical model must not only describe the individual telescope response, but above all the response of *both* instrument arms with respect to the common reference frame. Besides, the Basic-Angle Monitoring (BAM) device is an auxiliary instrument foreseen for keeping track throughout the mission of the most crucial aspect of the astrometric instrument, in particular over short time scales. The BAM model may allow introduction of astrometric corrections in the data reduction.

Individual telescope response is optimised at the beginning of life of the mission by alignment, also based on the data from a Wave-Front Sensor (WFS) pair. Availability of the WFS data at different moments during the mission lifetime may allow better assessment of the optical configuration status, solving part of the degeneration among instrument degrees of freedom, and thus consolidating the instrumental knowledge.

The instrument model provides basic data, and where required algorithms, for the Gaia Simulator, even if many analysis tools may remain external for practical reasons: e.g. ray tracing software packages (Zemax, CodeV) cannot easily be ported to Java, and some investigations (as on parameter perturbations) may not be of general need. This appears to be the best trade-off between simplicity and flexibility.
The internal representation of tabulated data for each parameter can be interpolated for intermediate values (e.g. time, wavelengths and FoV positions). The nominal and perturbed configurations are described by different sets of parameters, i.e. a sequence of tables or specifications on their evolution. Investigations on convenient sets of global parameters for system description during operation are planned, to account for the requirements generated by the optimisation of the data reduction algorithms.
The instrument model implementation must define, during each development stage, a set of reference test cases to allow internal assessment of the model quality (accuracy, reliability, sensitivity to parameter variations), whenever not already specified elsewhere. Generation of intermediate data is expected to be necessary for development and independent test of each section of the model. The need for progressive model improvement may be evidenced by investigations on crucial aspects, like detector electro-optical response variation.
Detailed physical modelling will impact not only on the set of parameter values describing the instrumental configuration, but on the algorithms included in the software development as well. The relevance of several aspects of the instrument transfer function may depend on the actual range of values taken at any given stage; e.g., it be marginal in the nominal case, which may be representative of the instrument at the beginning of operation, but may increase to relevant or critical level with the evolution of the equipment (ageing, radiation damage, failures).
It is assumed that the code complexity of the instrument model will increase significantly during the development, thanks to analysis of realistic effects and data, but provisions are made to face the probable need for significant update efforts after launch, after detailed assessment of the consistency of the model with actual data, with respect to both parameter values and algorithm improvements.

6.3.5 Detector

The effects of CCD characteristics (at individual and assembly level) are introduced into the signal model by parameters related to geometry (position, orientation) and electro-optical response (MTF, QE, gain, RON). MTF and QE are wavelength dependent and must be introduced in the composition of monochromatic PSFs, whereas geometry, RON and gain only need to be taken into account in the definition of sampling and detection of the polychromatic PSF. The implementation is based on progressive improvement of the corresponding algorithms and update of the relevant parameters, also based on the results from device characterisation and consequent evolution of the CCD physical model. Detailed description of the local defects impact and of TDI operation will similarly be implemented, likely at a later stage, by means of algorithms applied at the appropriate stage of construction of the detected signal, also taking into account possible magnitude-dependent factors.

On-ground characterisation of individual CCD response, in particular with respect to noise and charge transfer efficiency (CTE) degradation as a function of the radiation damage expected in the Gaia operating environment, is crucial to define a faithful algorithm representation of the detector model, suited to the diagnostics implemented in the data processing (in particular in First Look, 4.2).

6.3.6 Optics

The construction of monochromatic PSFs at a set of wavelengths is performed by numerical implementation of the diffraction integrals; the encoding strategy (e.g. spatial and spectral resolution requirements) and algorithms shall be reviewed and possible improvements identified. Optical analysis shall support the model encoding by computation of the relevant contributions (e.g. distortion and other image parameters), and by evaluation of feasibility and performance of alternative descriptions more suited to mission operation, in which detailed instrument parameters are not directly available and shall be inferred by the data.

The optical response can be internally represented by a limited set of parameters, provisionally identified in the Zernike coefficients up to order 21, for a discrete set of positions over the focal plane, and for each arm of the instrument. Additional intermediate optical information (e.g. tables of distortion, straylight model, etc.) will be provided.

6.3.7 Attitude

The satellite is expected to follow a scanning law affected by disturbances. The dynamical modelling requires knowledge of the satellite mass distribution, and of the geometry and characteristics of the actuators. Since the attitude is controlled in closed loop, details of on-board operation are needed to estimate the effective disturbance. This is a critical part of the attitude reconstruction activity (5.1.3), and its results are required for encoding in the instrument model, e.g. as low frequency errors plus high frequency noise, which can be described in terms of Power Spectral Density (PSD) and time series.

The orientation of the satellite at any moment defines the position on the sky of the Line Of Sight (LOS) of each viewing arm, and thus the observed regions. The attitude input (model of scanning law) is included in the instrument model through a coordinate transformation between the Universe Model convention and the on-board reference system. The conversion algorithm identifies the sources in the FoV and defines their FoV positions, for subsequent generation of the signals associated to each instrument section during the transit. The model will progressively include the improvements to attitude modelling.

6.3.8 Auxiliary instruments: BAM and WFS

The instrument model must describe the relationship between the BAM measurements and the LOS separation, including the realistic effects associated to detector, optics, structure variations. It is therefore a module with one input and one output variable, but internal dependence on several parameters, although the simplest model is linear. The output supports the analysis of the BAM data for verification of the astrometric stability of the instrument, by cross-check with the astrometric solution, and potentially for estimate of corrections.

The WFS measurement defines the local WFE map for the FoV position corresponding to the associated detector, allowing generation of the local PSF and therefore a quite direct check by comparison of the consistency between the current instrument model and the data. The WFS is modelled by optical analysis and inclusion of the detector response, to provide an assessment of the sensitivity (noise performance), and of the expected systematics (i.e. discrepancy between real and measured WFE).

6.4 The telemetry simulator

6.4.1 Introduction

The GAia System Simulator (GASS) is one of the tools developed to provide realistic simulations of the observation data. In particular, GASS will generate telemetry data as it will be sent to ground during the actual mission. As part of a more general Gaia Simulator, which includes a pixel data and an intermediate data generator, GASS implements a set of common packages and libraries that ensure the use of the same parameters, instrument models and universe models for all the data generators.

The generation of simulated data in general, and of GASS data in particular, is closely related to the needs of the other CUs. So, as explained in the introduction of this section, a close interaction between the preparations for the data processing (specifying simulations needs), the DPACE (accepting and assigning priorities to the data processing needs), and the GASS development team of Simulations will be needed.

6.4.2 Goals

The main goal of GASS is to simulate the telemetry stream of the satellite, using models to generate the observable objects on the sky (Universe Model) and instrument models. These, in particular the instrument models, assume some simplifications in order to allow the generation of a large amounts of simulation data with current computer resources. The goal of GASS is not the generation of very detailed data at the pixel level (this will be made by GIBIS), but rather the generation of *realistic* observation-level data streams for the 5-year mission.

The data generated by GASS is used for several purposes as:

- to provide basic input data to test the all data-reduction chains
- to test the data handling algorithms (e.g. cross-matching, telemetry compression, first look)
- to estimate the amount of telemetry to be down linked

GASS is developed in Java and uses common tools such as the Gaia Parameter Database and the GaiaTools library in order to ensure coherence with other simulation modules as GIBIS or GOG, and more generally, with the data processing environment.

The current version of GASS (GASS 3.0) follows the latest industrial design. However, some simplifications have been made in order to be able to generate the volume of data corresponding to 5 years of observations with reasonable computer resources. The simulated data volume can reach a few hundreds of GB of telemetry data just for simulations limited to stars brighter than G=12 mag.

6.4.3 Inputs

GASS uses a configuration file to set several options of the simulations. Apart from some internal configuration parameters, there are two main categories in the configuration file:

- Universe Model: specifies the kind of objects to be simulated. The current possibilities are: *galaxy*, *solar system* and *extragalactic*. For each category, it is possible to specify the actual physical model used in the simulations (e.g. Besançon galaxy model). The extinction model can be also chosen.

  Finally, the G limit magnitude of the generated objects has to be set.
- Instrument Model: through the configuration file it is possible to choose the instruments (*Astro / BP-RP / RVS*) to be used in the simulations.

Other specific options of the simulation have to be hard-coded in the GASS code or implemented through the use of the appropriate files (for instance the use of CCD readout noise, the use of a specific LSF, the time variation of the CCD positions, and more). In the same way, some properties of the Galaxy Model, like the inclusion of binaries and/or variables stars, the star density or even the star distribution over the sky can be specified in the code. So, it is mandatory to define a set of assumptions for each run of simulations that specify every relevant option, in particular the amount of sources to be generated and their distribution (uniform or not). Those assumptions will be used by the GASS development team to prepare the next version of the simulator and therefore the simulated data itself.

6.4.4 Deliverables

GASS generates several ASCII files containing different kinds of data:

- Attitude file: contains the attitude data in quaternion form. The four components of the quaternion together with the time are given for every second. An appropriate noise simulating the error in the on-board attitude determination can be added.

- Telemetry file: contains the TM stream according to the adopted TM model. The observations (*star packets* (see Sect. 3.6.1) are packed in *star sets*. Each star set contains a header with the priority level (the same for all the observations in the set) and the number of stars in the set (between 1 and 128).

  Although the output of GASS is an ASCII file, it can be translated into a binary file using appropriate codices. This transformation, however, is outside of the GASS scope for the moment.

- Source file: the source file is a *catalogue* of the sources observed by Gaia. It contains mainly astrometric and photometric information as well as other physical parameters such as effective temperature, gravity, and population.

- Auxiliary file: this file contains the *true* values (i.e. the actual values assuming the nominal position of the CCDs) of transit times and field angles for each CCD transit of each observation. The source and transit identifiers allows to cross-match each entry with the corresponding sources and telemetry entries, respectively.

  The file contains also the flux of the sources in the G band.

As a common rule, one telemetry and one auxiliary file are provided for each day of simulation. The sources are provided in a single file containing all the objects observed during the simulated time interval. In the case of the attitude, a single file covering an interval of time slightly bigger than the telemetry simulation interval is provided.

6.4.5 Access

As a general rule, GASS will be run under request of the CUs needing telemetry data. The large amount of data involved in even short simulations, the need of significant computational resources and the peculiarities of the output (TM model, assumptions, etc.) makes the execution of GASS difficult for non-expert users. Therefore, Simulations has proposed a protocol to provide simulation data to the data processing developments (see above). In short, every 6 months the simulation managers will send a request to the data processing units asking for their short term simulation needs. After approval of the requirements and definition of the next simulation run, GASS will be prepared to fit these requirements. Finally the data will be generated and send to the data processing units, at the same time as a new release of GASS is published. A more complete description of this protocol can be found in the introductory section of this chapter.

6.5 Pixel level simulator

6.5.1 Goals

To design and test the performances of the algorithms developed by the DPAC, simulations as realistic as possible of the sky observed by Gaia and of the instruments themselves are needed. The Gaia Instrument and Basic Image Simulator — GIBIS — has the task to generate simulated data down to the maximum level of detail possible. To be able to obtain those data in a reasonable amount of CPU time, GIBIS should allow to concentrate on specific aspects of the mission. While GASS, described in the previous section Sect. 6.4, simulates a huge amount of realistic raw telemetry stream, GIBIS has been created to simulate a smaller amount of Gaia observations to a greater level of detail, down to the pixel-level.

Pixel-level simulations will be needed in all studies of specific complexities that the data processing will have to handle. In particular GIBIS will be needed to study the calibration of the instrument, develop imaging capabilities, reduce crowded fields, analyse the sky background, handle multiple star contamination of the same observed data, reduce extended objects, etc.

Those detailed GIBIS simulations will be used to provide statistical results to be applied by GASS for large scale telemetry simulations.

6.5.2 Functionality

The main structure of GIBIS is presented in figure Fig. 45.

The user can specify the characteristics of the sky portion for which Gaia observations are requested, the characteristics of the Gaia instruments, the simulation methods to

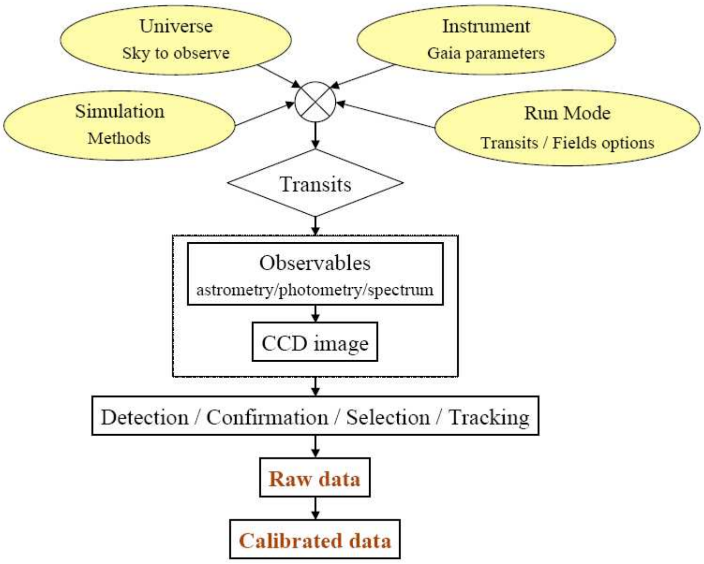


Figure 45: GIBIS process main structure

be used and the selection of the transits and fields to simulate.

Then for each transit, GIBIS transforms the astrophysical characteristics of the sources into observables in the fields requested, creates the CCD images and runs the on-board algorithm prototypes to deliver to the user the resulting raw Gaia observations per transit.

As most algorithm development will not need directly raw data but calibrated ones with a number of pre-processing activities already applied (cross-matching, spectrum wavelength calibration, multiple transit combination, etc.), a calibration step is planned.

Inputs

GIBIS universe GIBIS aims to be able to simulate all the different kind of sky configurations that Gaia will observe. The objects can be point sources (stars, quasars), extended (unresolved galaxies), moving within the integration time (asteroid) or a combination of those (resolved galaxies are both extended and contain stars, near-earth objects can be both fast moving and extended for the high Gaia angular resolution). Not only average sky properties through statistical distributions are simulated but also extreme cases such as high stellar density and substantial background variations.
To specify the exact characteristics of the sky to be observed, the user can select either statistical models implemented in the universe model (c.f. section Sect. 6.2) or special configurations using source catalogues (e.g. a globular cluster) and background

images (e.g. an HST image of a nebulae). Those configuration files can be taken from the GIBIS web site or provided by the user himself.

GIBIS instrument The main characteristics of the Gaia satellite and payload are simulated in a modular way, allowing a progressive improvement of the detail level of the simulations. The instruments models are described in section Sect. 6.3. The nominal configuration parameters of the models are initialised according to the Gaia Parameter Database [gai]. The user will be able to choose to add variations compared to the nominal instrument model, including defaults to the CCDs, ageing of the optics, radiation damage, perturbations of the attitude, etc.

GIBIS simulation methods Considering that highly detailed simulations can be very CPU-time consuming and that not all simulations need the same level of precision, options are available to switch on and off different simulation methods. For example, the generation of the PSFs on the fly can be replaced by the use of typical PSFs stored on disk.

GIBIS run-mode Finally GIBIS allows to specify the fields (astrometric, photometric or spectroscopic) for which data should be generated and the transits requested. Either all transits observing the sky region defined can be queried or specific transits which characteristics can be chosen.

Outputs

The main output of GIBIS is raw data.

Information can also be provided for checking purposes, such as the observables catalogue, the full CCD images (as observed by the CCD after the time-delay integration but before the sample read-out), as presented in figure Fig. 46, or on-board algorithm data process information not transmitted to ground.

Considering that most algorithm developers will work on main database data rather than raw data, some calibration and pre-reduction shortcuts on the GIBIS raw data will be made available. This can be done either in form of other data outputs with different formats, or by providing access to a library with methods implementing those reduction shortcuts. The generation of those calibrated data will be based on the models developed for the intermediate data simulator GOG described in the next section Sect. 6.6.

GIBIS output formats, for both raw and calibrated data, will be standardised with the other data generators GASS and GOG, based on the telemetry and main database interface format definitions.

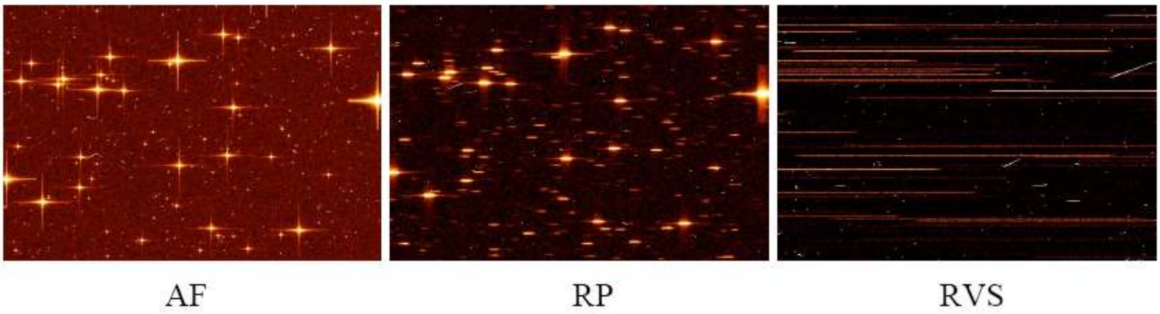


Figure 46: GIBIS 3.0 simulation of a globular cluster observed in the astrometric, photometric and spectroscopic fields.

6.5.3 Deployment

Access to the GIBIS simulator is provided to the DPAC community through a web interface using Java Servlets. It is currently deployed at `http://gibispc.obspm.fr:8080/gibis`, but the migration to a deployment at CNES is under finalisation.

CNES deployment will also allow to run simulations automatically in command line, by-passing the web interface.

Deployment of mirrors of GIBIS could be investigated if needed.

A database of simulated GIBIS data could be created to provide access to ”classical” but CPU demanding simulations.

6.5.4 Development and Status

GIBIS has been under development since 2001 and used by the Gaia community through the web interface since 2002.

The design of GIBIS has been made in UML (Unified Modeling Language) and its core developed in Java. It uses the core library of the simulator, used also by GASS and GOG, containing the universe and instrument models. It uses the Gaia ToolBox [Ang05] and the Gaia Parameter Database [gai]. Some specific simulations can be written in any other language and are called from the Java core. This allows a quick integration of the programs provided by the astronomical community which uses a wide range of programming languages.

The current release of GIBIS is version 3.0, described in [BCS06]. History, contributors and other GIBIS documentations are available at `http://gibispc.obspm.fr:8080/gibis/documentation.html`. The next step in the GIBIS development will be to produce calibration data. It will then continue to increase step by step towards more realistic simulated data. Its development plan will follow the general planning of the simulation described in section Sect. 6.7.

6.6 Intermediate data simulator

6.6.1 Concept

In the previous sections the pixel level data generator (GIBIS) and the telemetry generator (GASS) have been introduced. Although these two elements will cover a broad range of DPAC needs, a third data generator will be necessary to the development teams. This third data generator will be named GOG (Gaia Object Generator).

GOG's main purpose will be the generation of Main DataBase data (MDB data). *MDB data* is the name given to those datasets that are stored in the main ESAC database and generated in any of the several steps that compose the reduction process, including the Initial Data Treatment. Additionally, GOG will also catalogue data directly generated by the Universe model.

One should note that MDB data could also be obtained from the telemetry stream generated by GASS – and in some cases, processed through the main ESAC database – or after processing GIBIS images. However, this method could be very time consuming since it would generate a large amount of unneeded information. GOG will be a shortcut designed to avoid this overhead in an efficient way.

GOG will thus be a tool to directly get catalogue and Main database (MDB) data without the use of GASS telemetry, GIBIS or the main ESAC database. This concept is represented in Fig. 47.

The first step for GOG will then be the production of catalogue data (that is, simulated objects from the Universe Model). The object information will include:

- Astrometry: position, proper motion, radial velocity.
- Photometry: G magnitude and other magnitudes and/or colors needed.
- Astrophysical parameters: g, T_{eff}, etc., needed to select the spectra of the source.
- Others such light curves (variables), orbital elements (binaries, extrasolar planets), etc.

From the catalogue data GOG will use the nominal instrument model to generate *nominal MDB data*, including:

- Photometry: flux in G band.
- Spectrophotometric data: spectra convolved with the nominal instrument response
- RVS data

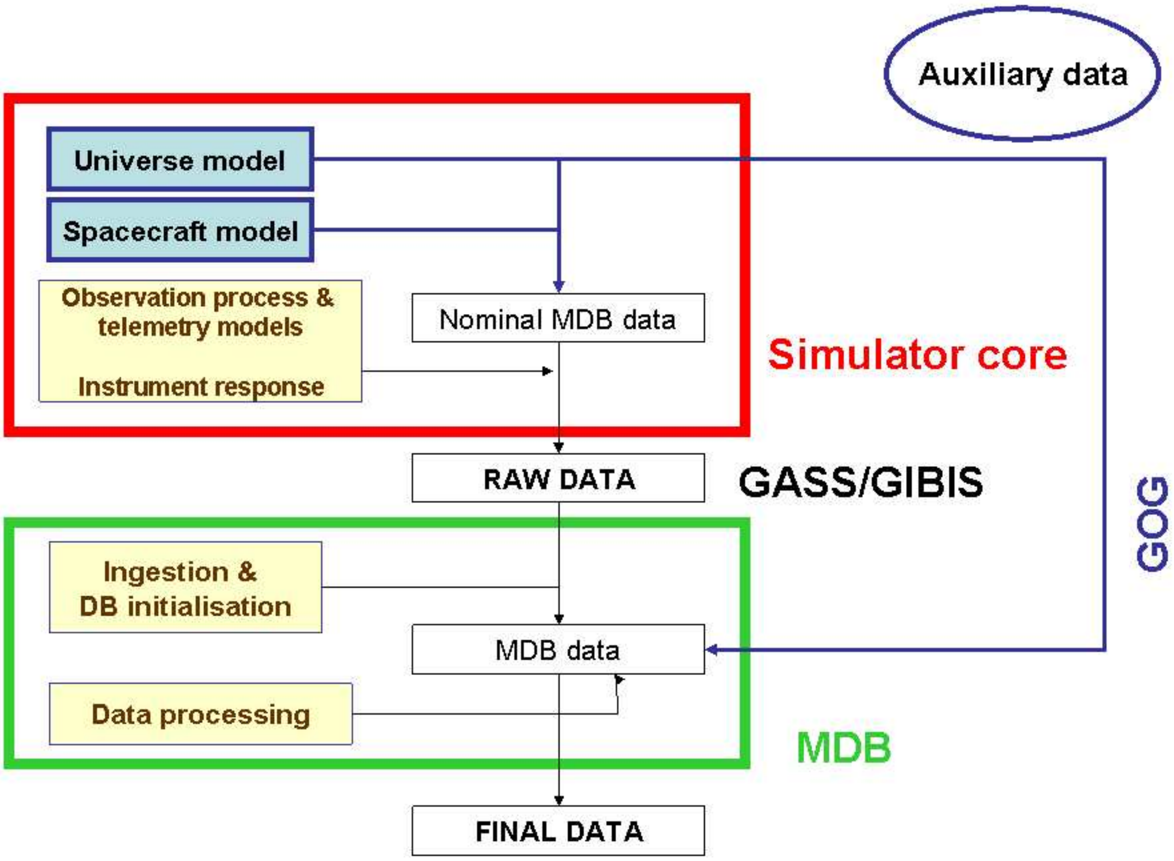


Figure 47: GOG concept

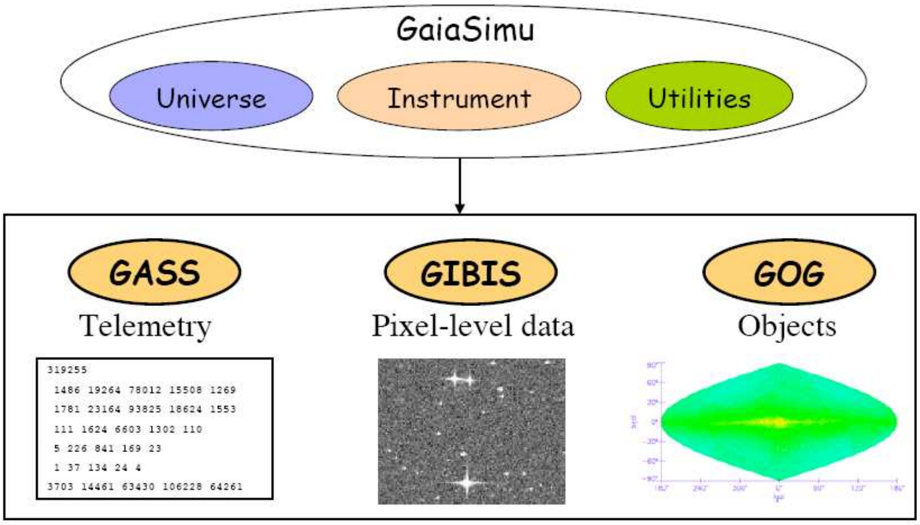
- Transit times: exact time when the star crosses the center of the readout column of each CCD (assuming the nominal geometry of the FoV)

- Field angles: angular coordinates of the source image on the focal plane

- Others

Of course, this nominal MDB data will not correspond to any kind of real data ever generated during the mission, but they will rather be an ideal reference for the validation of the reduction procedure.

Later, by using error models mimicking the behaviour of the different reduction algorithms, this data will be finally converted by GOG into *simulated MDB data*, which will constitute its main product.



```
319255
 1486 19264 78012 15508 1269
 1781 23164 93825 18624 1553
 111 1624 6603 1302 110
 5 226 841 169 23
 1 37 134 24 4
3703 14461 63430 106228 64261
```

Figure 48: Organisation of the simulation development

6.6.2 Boundaries

The development of the Gaia data processing system is a very complex project carried on by many teams around Europe. During this development many subsystems and modules will be developed, in different temporal sequences. The available resources will not allow to tailor the GOG simulations of data to each and every one of the needs of these separate modules, and therefore the boundaries of GOG have to be defined.

These boundaries can be summarized, in short, by this statement: *GOG will simulate MDB data plus the addition of any error model provided to mimic the result of any reduction algorithm or reduction step*. Therefore, an essential element of GOG will be inclusion of these error models, whose provision will be a responsibility of the algorithm developers.

6.6.3 Deployment

GOG will be deployed as a Java `jar` file. In this way it will be possible to run it on different platforms, either desktop computers for short specific simulations or large facilities in the data centers for massive simulations.

6.7 Development plan

In Sect. 6.1.3 the simulator building blocks (described in the previous sections) have been listed. These building blocks are depicted in Fig. 48.

The development of the Gaia simulator and the production of simulated data started already some years ago, in order to guide the instrument design and to feed the prototype of the reduction system known as GDAAS. These tasks are thus in a quite unique situation with respect to other DPAC tasks.

Due to this fact, the development of the data simulation system for the DPAC is in a very advanced state, even before the formal constitution of the consortia. The existing simulation system (over 100,000 lines of code, in Java and C) does already cover a part of the needs of the DPAC teams and will be further expanded to cover the remaining needs: both GASS and GIBIS are already on production, generating data on demand, and a first GOG implementation will be available in early 2007.

Starting from this existing system, the plans for the development are aimed to increase the level of detail and realism of the simulations, including but not limited to:

- New types of objects and more detailed physical models in the Universe Model
- More detailed models of the instrument and the spacecraft: optics, CCD behaviour, satellite dynamics, etc.
- A version of the simulator allowing the detailed simulation of radiation damage effects
- A version of the simulator including fully realistic telemetry formatting
- A library of error models for GOG allowing to simulate the results of the reduction algorithms at any intermediate step of the reduction process

These improvements will be implemented following the development cycles of the overall project, thus producing a new version of the simulator in each cycle incrementally implementing more realistic and detailed modules.

The expected development effort in each module of the simulation varies with time (Fig. 49), following the development of the mission and of the implementation of the reduction system. Thus, a large effort will be required during phase B2 of the mission to adapt the simulator to the changes in the spacecraft and instrument design, but once phase C is started such changes will be severely restricted. On the contrary, the needs for a more realistic and detailed universe model will increase as the development of the Gaia Data Processing system advances, culminating some time before mission launch when the final tests of the full system should be carried out. Finally, the effort on software engineering and scientific validation are expected to remain constant until mission launch.

From mission launch onwards the need for simulated data is expected to drastically drop, as the first batches of real data will arrive. However, maintenance of the simulator for some years is required in case some additional simulated data is requested during the mission.

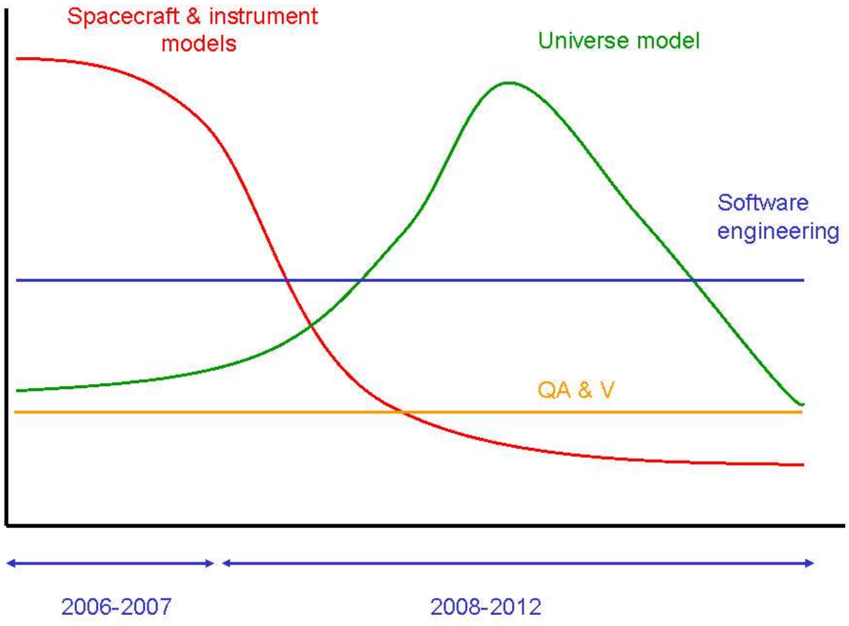


Figure 49: Expected effort for the development of the simulator

The main milestones for the simulator development are:

Early 2007: release of the first full version of GOG. All three data simulators are operative.

Mid 2007: end of Gaia phase B2. Instrument and spacecraft design are mostly completed. Implementation of final spacecraft and instrument models can start.

Mid 2010: GASS generation of a full-mission simulation for end-to-end testing of the reduction system

Early 2011: simulator development starts to wind down. Preparation for minimum maintenance phase.

Mid 2012: start of minimum maintenance phase.

2017: end of simulator operation

7 Data Processing Architecture and Operations

This section describes the the organization of the data processing from the HW/computational point of view with the hub-like organisation around the unique and centralised MDB. It introduces the DPCs (detailed later) and their relationship to the MDB.

7.1 System architecture

7.1.1 Approach

Any large system is normally broken down into logical components to allow distributed development. Gaia data processing is on a very large and highly distributed scale. The approach taken to the decomposition has been to identify major parts of the system which may operate relatively independently, although practically all parts of the Gaia processing are in fact interdependent from the point of view of the data. From a development point of view however, a well defined ICD (Interface Control Document) would allow completely decoupled components to be developed and even operated in disparate locations. The approach is driven by the fact that this is a large system which will be developed in many countries and by teams of various competencies.

Hence at this level of decomposition libraries or infrastructure are not considered to be components. At some lower level these components may indeed share libraries and infrastructure but this is not a cornerstone for the architecture. Only the top level components and their interaction are considered in this decomposition.

7.1.2 Logical Components

The list of components for Gaia Data processing has indeed been emerging for some time, the first indication of them in their current form was in [MBJ05]. Fig. 50 show the logical components of the system and the data flow between them.

- Mission Control System (MCS)[18]
- Data Distribution System (DDS)
- Initial Data Treatment and First Look (IDT/FL)
- Simulation (SIM)
- Intermediate Data Update (IDU)

---

[18] The MCS and DDS are MOC responsibilities, not part of DPAC and are included here for completeness.

- Astrometric Global Iterative Solution (AGIS)
- Astrometric Verification Unit (AVU)
- Object Processing (OBJ)
- Photometric Processing (PHOT)
- Spectroscopic Processing (SPEC)
- Variability Processing (VARI)
- Astrophysical Parameters (ASTP)
- Main Database (MDB)
- Archive

In the diagram the (Coordination Unit) CU notion is retained as it provides an agreed top level division of the processing effort and system. The CUs are small in number, with clearly-defined responsibilities and interfaces, and their boundaries fit naturally with the main relationships between tasks and the associated data flow. The CUs are described in detail in Sect. 8.

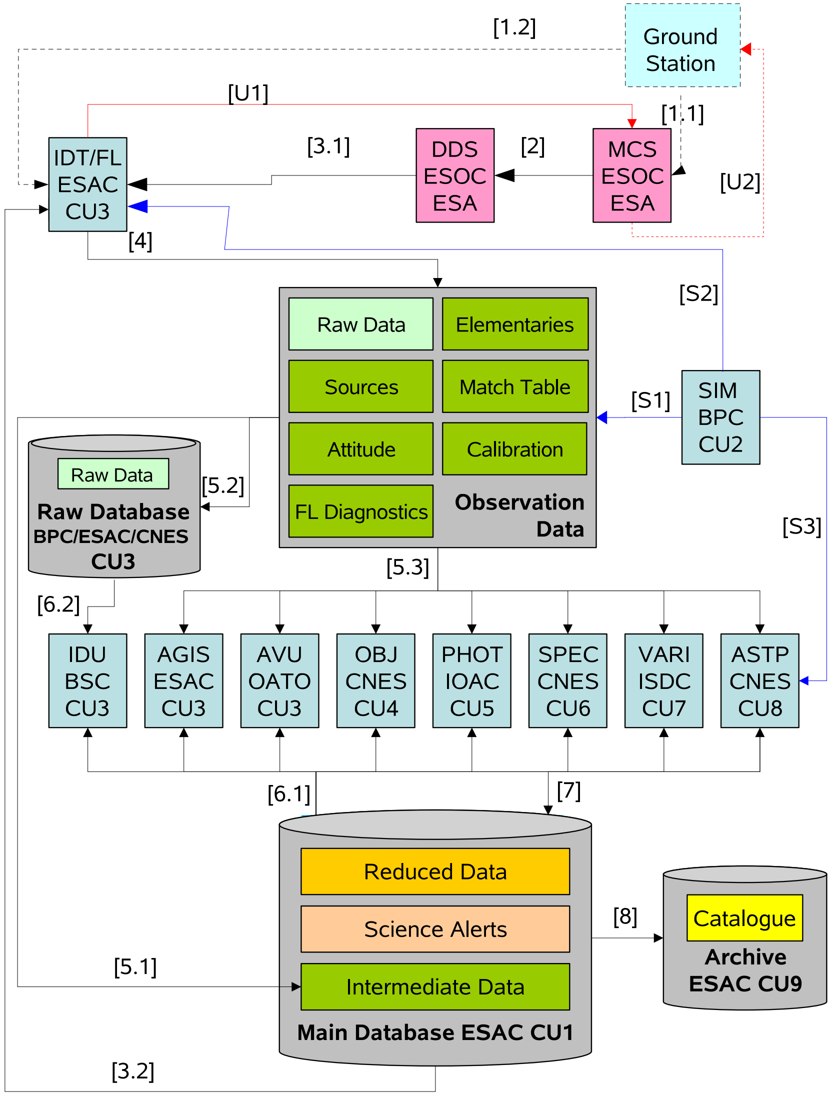
7.2 Data Flow

Gaia processing is all about data. The data flow is the most important description of the system and has been under discussion within the community for some time. From these discussions we see the data flow depicted in Fig. 50. The flow lines in Fig. 50 are labelled and these labels are referred to in the text below. The data flow is divided into two categories, Near-Realtime and Scientific Processing.

7.2.1 Near-Realtime dataflow

Near Realtime data flow represents the data flow on a time scale of approximately 1 or 2 days, corresponding to the activities of the Mission Operations Ground Segment.

The Mission Operation Centre (MOC) at ESOC receives all telemetry from the Space Segment[1.1] via the ground stations. The Science Operations Centre (SOC) at ESAC will receive all telemetry directly from the ground station also [1.2]. This data flow is to be finalised in the MOC-SOC IRD [Hoa07] here we give a simple view of the situation only. This raw data flow from the satellite is not shown explicitly in the diagram. Over the nominal mission duration of five years the payload will yield a total uncompressed data volume of roughly 100 TB. The satellite will have contact with the ground station once a day for a mean duration of 11 hours. During this



[U1] IDT/FL inputs to MCS – Planning/Calibration
[U2] MOC satellite uplink , TC, Calibrations
[1.1] HK from Gaia to Mission Control System (MCS)
[1.2] TM delivered to Initial Data Treatment and First Look
at ESAC directly from ground station
[2] HK data delivered to DDS.
[3.1] IDT/FL Orbit and Aux Data from DDS
[3.2] IDT/FL inputs from Main DB (Obs, Calibration)
[4] Daily observational files created
[5.1] Daily data ingested to Main Database
[5.2] Raw Data Archived in Raw Database
[5.3] Daily data delivered to Subsystems as required.
[6.1] Main Database version delivered to Subsystems
[6.2] Raw Database input to Intermediate Update system
[7] Subsystem updates delivered to main database.
[8] Main Database delivered for Catalogue production
[S1] Possibility to produce observational simulation
[S2] Possibility of TM simulation
[S3] Special Training Set Simulations for CU8

Figure 50: Top Level Components Data Flow for the DPAC

period, or “pass”, an uncompressed data volume of roughly 50 GB is downlinked from the satellite via its medium-gain antenna, at a rate of about 5 Mbit/s (0.625 MB/s).

7.2.2 Mission Control System

The raw telemetry data received by the ground station will be transmitted to the Mission Control System (MCS) at the MOC in two datastreams. ‘High Priority Telemetry’ will be transmitted to the MCS approximately one hour after reception at the Ground Station and includes Housekeeping Data, Science Housekeeping. ‘Regular Telemetry’ will be sent to the SOC as it becomes available (this is to be finalised in [Hoa07]). The MCS will provide an immediate assessment on the spacecraft and instrument status through analysis of the Housekeeping data.

7.2.3 Data Distribution System

All telemetry received by the MOC systems will be ingested into the Data Distribution System (DDS) [2]. The DDS will also contain data that were generated on-ground (e.g. orbit data, time correlation data), operational reports (telecommand history and timeline status), Satellite Databases used by the MCS and a copy of all telecommands send to the spacecraft.

7.2.4 Initial Data Treatment and First Look

Science Telemetry is received by SOC [1.2] for processing by IDT. Data is also retrieved from the DDS by the MOC Interface Task at the SOC and passed to IDT [3.1]. The IDT processing will decode and decompress the Telemetry. It will also extract higher-level image parameters and provide an initial cross matching of observations to known sources. Finally it will provide an initial satellite attitude.

The primary objective of First Look (FL) is to ensure the scientific health of Gaia. This information is returned to the MCS [U1]. First Look processing will carry out a restricted astrometric solution on a data set from a small number of great-circle scans.

To perform some of its tasks IDT/FL requires reference data, such as up-to-date calibration data and source positions of bright objects that are expected to be observed by Gaia during the time period to be processed. This data will be made available to IDT/FL [3.2] from the MDB. FL will also calibrate the current data set itself and this calibration will be used by IDT. The precise interactions of IDT and FL in this area are still TBD.

7.2.5 Uplink

Telemetry is received by the MCS which does basic system monitoring.

The First Look Diagnostics produced by FL [U1] will indicate if there are anomalies in the scientific output of the satellite which can be corrected on-board. After interpretation of the diagnostics, the FCT (Flight Control Team) is informed of the anomaly, which can be resolved either through immediate commanding or during the next mission planning cycle.

On a regular basis the MCS will send the prepared command schedule to Gaia [U3], taking into account normal planning and inputs from IDT/FL. During a Ground Station Pass, immediate commanding is also possible.

7.2.6 Daily transfers and Raw Database

The output of IDT/FL are made available to all tasks on a daily basis[4,5]. This comprises:

- Raw Data: invariant.
- Attitude: the best attitude spline coefficients derived by IDT/FL
- Calibration: calibrations derived by IDT/FL
- Elementaries: the higher-level image parameters for each observation.
- Sources: new sources derived from the cross match process
- Match Table: match of elementaries to sources
- FL Diagnostics: information about the satellite and science data from First Look processing.

The scientific output of IDT/FL will be ingested into the Main Database [5.1]. The Raw Database will be a repository for all raw data [5.2]. Copies of the Raw Database are expected at ESAC, BPC and CNES. Other tasks may retrieve the data according to their requirements [5.3]. Raw data will only be transmitted on a daily basis i.e. it does not form part of the Main Database and is not foreseen to be sent again later.

Data Processing Centres may produce Science Alerts from Observation Data. Science Alerts are sent to the SOC for immediate distribution to the scientific community and archiving in the Main Database [7].

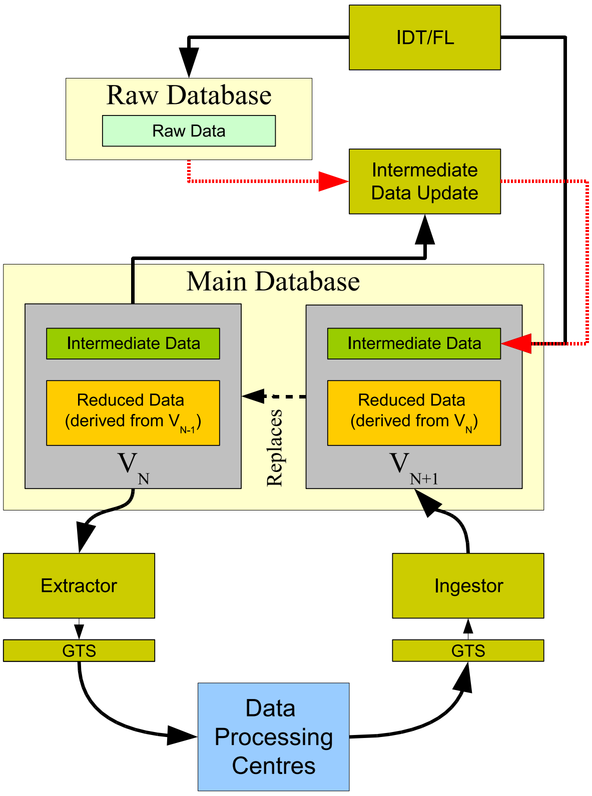


Figure 51: Versioning of the Main Database

7.3 Main Database

The Main database forms a Versioned repository of all Reduced Data as well as some Intermediate Data. A subset of the latest version of the Main Database will be sent to the DPCs for processing by the processing tasks ([6.1] in Fig. 50). After processing the DPCs will return updates (Reduced Data) for inclusion in the Main Database [7] . The version scheme is explained in more detail in Sect. 7.3.1. Interaction with the Main Database[5,6,7] will be governed by an Interface Control Document (ICD) [Her05]

7.3.1 Contents and versioning of the Main Database

The MDB will contain a number of version-controlled results databases that will form a sequence V_0, V_1, V_2, . . . Each V_n comprises two main parts, viz.

1. Intermediate Data: Elementary and Source/Spectro Window observations data expressed in local plane coordinates [Lin05b] plus a cross-match table linking observations with the source data etc.

2. Reduced Data: The astrophysical parameters of the identified celestial objects as obtained from the Astrometric, Photometric, Spectroscopic, and Object Processing plus attitude, and global parameters from the Astrometric processing as well as calibration and PSF/LSF parameters from a number of processes

This is depicted in figure 51.

The Intermediate Data will be generated continuously as IDT processes the telemetry received at ESAC. The Reduced Data will be the result of performing the different distributed processing tasks. All the Intermediate Data will be regenerated by IDT when new Reduced Data is available, this together with the Reduced Data produced by the DPCs will be ingested in the new version of MDB.

With the exception of the initial catalogue, V_0, each database version V_n represents the combined results of large-scale, coordinated runs in the processing systems (see 50). These Data Reduction Cycles will take place at regular intervals, e.g. every 6 months. The process of creating V_{n+1} from the contents of V_n is as follows (see also figure 51):

1. At $T_n = (n+1) * 6$ month, $n >= 0$, the result database V_n is available in the MDB. This event will trigger the transfer of the MDB extract to the Data Processing Centres. The exact contents of the MDB extract will be different for each DPC and will be determined by the ICD [Her05].

2. Each processing centre will start a number of designated tasks once the respective local extract of V_n has become available. All input data needed by all tasks of a processing centre must be available locally, i.e., no cross-communication between tasks in different centres is foreseen. In practice this may mean that if a task A in centre X requires (as input) the output from a task B running in centre Y, the A-processing has to be postponed until the next iteration when the results of B will be available as part of Y's local extract of V_{n+1}. This restriction is a limitation of the proposed design, but the alternative, allowing inter-centre task communications, would entail an unmanageable level of complexity. At this stage it is still too early to know in detail the time needed to perform the different processing tasks, bearing in mind that the first full data reduction cycle will be performed on a small set of Intermediate Data we could foresee a smaller interval of time between the V_0 and V_1 databases.

3. Once the Data Processing Center finishes the processing of the data, it will transfer the output to ESAC using the GTS (Gaia transfer System). Once the data are received they can be ingested in the new V_{n+1} MDB. There will be an ingestor task that will perform certain validation on the data and then ingest it on to the MDB. In general if any of the DPCs misses the delivery of its newly reduced Data, the new MDB will use the previous version of it.

4. A final step in creating V$_{n+1}$ is to reprocess with IDT all the Raw Data up to that moment of time with the new PSF/LSF and calibration values derived from the different Reduced Data sets. The new Intermediate Data will be ingested into V$_{n+1}$.
5. Once V$_{n+1}$ is complete, a Global QC (Quality Control) process is performed to ensure the overall quality of V$_{n+1}$.
6. At this point V$_{n+1}$ replaces logically V$_n$, and a new iteration begins.

The first version V$_0$ of the results database will have to be constructed in a special way before the iterative cycle can commence. Attitude and calibration parameters for the first six months will come from First Look processing, while the source data will be generated from a starting star catalogue (e.g. USNO-B, GSC 2.3 or equivalent).

Note that each version V$_n$ of the Database (with the exception of V$_0$) shall contain only a single list of sources but each source can have two different origins: It was either already included in the previous version V$_{n-1}$ of the Database or newly created by the second stage Cross Matching as part of the IDT.

7.3.2 Size of the Main Database

The main database will grow in volume with each version. A rough approximation of the size may be made by simply considering the downlink volume and allowing for an expansion in size for the added products. If we consider 40GB per day over six months then we would have around 7 TB. A nominal expansion of around 3 would then suggest a V$_1$ of about 20TB. The growth each version should remain fairly steady hence a final Main database would be at least 200TB. We shall estimate this number more rigorously using the data model which is currently under construction. Effectively the number of expected objects is well known, hence if we can tie down the description of each object in terms of bytes we will have an extremely accurate estimate of the database size. The current estimate using this approach is more like 300TB but it is too early to settle on a final number.

7.3.3 Database Backup

Backing up a 300TB database locally is not trivial. Space permitting the ideal backup would be on spare possibly slower disks. The database will partially be replicated at many locations thus providing a distributed backup. Our aim is to have a complete copy of the MDB at CNES. We envisage complete copies of RAW data at ESAC,CNES and BPC. Such distributed copies offer more security than a local back up. In addition one must consider that for a given MDB version the previous version will be available as well as all inputs which make up the next version. Hence it should be possible to reconstruct the database. We are currently not considering a large tape storage archive or other dedicated backup system.

7.3.4 Structure of the database

Astrometric data is normally perceived in the community in a tabular manner. As such there is a good match to relational database systems. Selecting for a table driven system early on also allows for other simplifications in the system design. If we consider all data to be tabular then in the software we may have a table abstraction (or interface) which can be implemented against any number of relation database management systems. Even more interestingly the abstraction may be easily implemented for other table like data structures such as FITS [NAS95] or plain ASCII files Fig. 52. The software sitting on top of such an abstraction layer can then easily switch from files to databases without any rewriting of code. This allows us to have a testbed with some FITS files while maintaining the database for the production system. This approach has already been successfully employed in the work at ESAC.

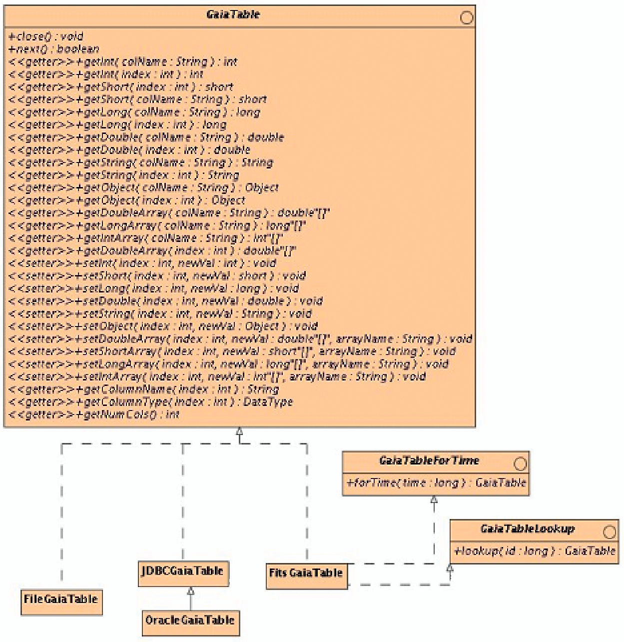


Figure 52: GaiaTable interface with multiple implementations.

The natural ordering for data within the MDB is spatial. Further work is needed in this area but HTM and/or HEALPix would seem to offer ideal candidates for a sky indexing scheme. Such a scheme would allow for partitioning of the database to multiple disks or even multiple machines. Certain database systems facilitate such partitioning still allowing the complete dataset to be viewed as a consistent whole. If such a commercial solution proves inefficient or ineffective we still have the opportunity to split the MDB into multiple separate databases on physically separate machines with independent disk subsystems.

It is clear that for the astrophysical parameters of an object and even the elementary observations of an object they should be stored within the database system and managed by database software. For the spectra the case is not clear and requires further study. Most database systems support the Binary Large Object (BLOB) feature which effectively allows one column in a table to act rather like a file - hence there is a possibility to keep this in the database also - the trade offs need to be experimented with.

7.4 Further Scientific Processing

Scientific Processing represents the production of the Gaia data products by the Data Processing Ground Segment from Intermediate Data. The timescale for each iteration of this process is much longer than Near Realtime processing, of the order to six months or more. It will continue after routine satellite operations have finished and culminates with the production of the final science products. The outputs of processing from each CU will be sent for incorporation in the Main Database [7]. The Main Database is described in more detail in Sect. 7.3 while the the science processing is described in detail in the in Sect. 5.

7.4.1 Access to the main database

Access to the main database is governed by the the ICD[Her05]. This document forms the baseline agreement for all CUs and will go under formal configuration control around late 2007. CU1 will be the curators of this document.

Each of the DPCs will require access to the database, it forms the hub of the DPAC system as depicted in Fig. 63. Intermediate and raw data will be transmitted to each DPC where processing will be performed and results returned to the main database. This may be viewed as a large distributed database but we feel the technology is not there to handle the volume of updates required. Instead we have a notion to make versions on more fixed points in time as describe in Sect. 7.3.1. The current baseline is to dump the main database to files and push these files to the other DPCs. At a fixed point in time the DPCs would return the results of their processing for inclusion in the next version of the database - also in the form of files. In this manner no single DBMS (Data Base Management System) architecture is enforced on the consortium and each processing centre can organise the data in a manner most efficient for their processing. Such data organisation is a key point for the efficiency of any of the processing systems in DPAC.

Hence only SOC directly accesses the main database - we feel this leads to a more secure, maintainable and consistent database. The fixed time versioning also allows changes to the schema and ICD to be introduced at these fixed periods - hence there will be no schema evolution of any given version of the database.

The transfer of files between DPCs will be governed by the Gaia Transfer System (GTS). This should provide monitoring, resending of files, error alerts etc. This would be built on existing technology or use an existing system if a suitable one meets the requirements. Studies will be carried out in this area as the transfer of potentially a few hundred terabytes over the academic network will require very efficient tools. DPAC acknowledges that a physical media transfer may be required for this system.

The current baseline for the file format of the transfers is FITS[NAS95]. Each table would be split in multiple FITS files. Tools for reading and writing FITS are readily available and FITS access has been built in to the Gaia Data Access Layer (DAL) in the Gaia Toolbox already - hence the FitsGaiaTable implementation in Fig. 52.

7.4.2 Intermediate Data Update (IDU)

IDU is a very demanding process, both in terms of storage and computing capabilities, that must run every six months. The raw data collected every six months of mission amounts to around 3 TB. Hence at the end of the nominal mission some 30 TB of raw astrometric data must be treated by IDU. As depicted in Fig. 50 a copy of the RAW data is located in Barcelona where IDU is intended to run. The IDU, like the other processing tasks, processes this growing amount of data in relative isolation from the other parts of the system. It interacts only through the MDB which is governed by the ICD.

In Fig. 53 we schematically show how the IDU will work. IDT running at ESAC produces raw data which are stored locally and sent to the raw data base hosted at BPC (see Sect. 9.7) while the intermediates are stored in the MDB version n. After some time (six months) the intermediate data are treated by reduction processes to give the $n+1$ version of the reduced data which are stored in the MDB version $n+1$. At the time Tn, when the production of new reduced data starts, the MBDn is frozen and the flux of intermediates coming from IDT is stored in MDB v $n+1$. This $n+1$ version will contain source, calibration and attitude data from the start of the mission to Tn. To run an IDU process, all the raw data observed between the start of the mission T0 and the time Tn enter IDU and are extracted from the local Raw Database, along with calibration, attitude and source data extracted from the version $n+1$ of the MDB.

It should be noted that while IDU is improving the existing intermediates, new raw data and new intermediates coming from IDT are being stored in the raw databases and in the MDB version $n+1$. That is, improved and not improved intermediate data

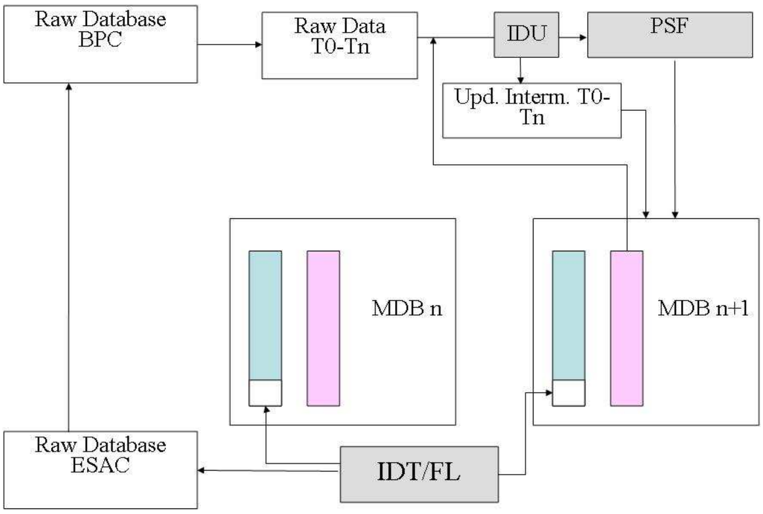


Figure 53: The interaction of IDU with the Main Database

coexist in all the versions of the MDB. To show that, the intermediate area in MDB in Fig. 53 has been separated into two sections.

It has been considered that a computer like Mare Nostrum (Sect. 9.7, or better, its successors) with large scalable computing facilities will be the right place to run such a task in a distributed manner. Each node can cope with the objects in a given area of the sky or in a given interval of time. The right environment has to be designed there to implement IDU and the raw database.

7.5 Software development approach

7.5.1 Project Phases and Milestones

For the purpose of resource planning and WP definitions we adopt the following coarse project phasing based on the ESA mission phases [Div03]:

Phase	**time frame**	**title** (reviews ECSS-E-40 Part 1B)
B	2006-07	definition (System Requirements Review)
C	2008-09	development and production (Critical Design Review)
D1	2010-11	verification (Qualification and Acceptance Reviews)
D2	2012	commissioning
E1a	2012-14	early mission
E1b	2014-17	late mission
E2	2017-18	extended mission
F	2018-20	post mission

Although the extended mission is not guaranteed it is assumed in planning as it is listed in the SMP (Science Management Plan). Note also that the cyclical development approach and how it ties into these phases is described in Sect. 7.5.2. A details of the reviews are provided in Sect. 7.5.4.

7.5.2 Development planning approach

The DPAC will follow a cyclical development approach up to launch. By this we mean many incremental system release will be made, each building on the previous experience. This going from 'working' system to 'working' system is possibly the only way we may hope to build the complex processing system required by DPAC and is not a particularly new idea [LB03]. This is essentially a formalisation on a large scale of the eXtreme programming [Bec99] approach. The important factor will be to complete the cycle and deliver the software on time. The exact deliverables will be captured in the requirements task at the beginning of the cycle while a broad outline of all cycle deliveries should be made in cycle 1. There are many reasons to have many short cycles:

- Problems should be encountered sooner rather than later.
- It is easier to adapt to changing or ill defined requirements e.g. due to problems encountered.
- Short projects have a better chance of success - we will see a working system sooner.
- In our very distributed development we need to synchronise some work e.g. large scale simulations.

One of the important tenets of eXtreme programming and the one most important for a development like DPAC's is to cope easily with changes, as not every detail of the DP can be fully planned in advance. The whole structure must have the responsiveness to adapt smoothly to significant modifications to the initial plan.

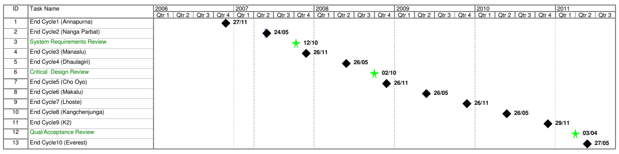


Figure 54: Development cycles and reviews to launch. Each cycle is around six months in duration. The review dates must be agreed and fixed with ESA.

In Fig. 54 the cycles are shown. Each cycle has a number and a name[19]. The important point here is the end date of the cycle. Cycles may overlap and indeed requirements will be captured in early cycles but may not be implemented until far later. In the same manner some CUs may wish to deliver slightly ahead of the milestone for planning purposes. The point is that around the milestone a complete set of DPAC software should be stable and available. This will facilitate end to end testing and generally get the DPAC in phase.

With this in mind simulations have already been identified as critical with most CUs depending on CU2 for simulations. Hence the CU2 development cycle starts one month in advance of other CUs cycles aiming to be able to provide simulations to them in one of the following scenarios:

1. Making the simulations available about one month before the end of CUx cycle N, during the CUx testing and integration phases

2. Making the simulations available in advance of cycle N+1 for CUx

3. In some cases, also partial releases during the cycle could be agreed for early delivering of limited-feature simulations to CUx

Therefore, In order to be able to produce the simulations in due time, the CU2 needs to receive the requirements for simulations from other CUs at the beginning of its cycle, that is, one month in advance of the other CUs cycles. The overall cycle structure is depicted in Fig. 55.

7.5.3 Configuration Management

All source code from DPAC will be in a source code repository (currently subversion). Access to the repository is restricted to DPAC members. Upon release of a subsystem it will be tagged in the repository and a distribution made available along with the standard documentation as laid out in [LJMD$^+$07]. Changes are only made to code

---

[19] The ten highest peaks in the world

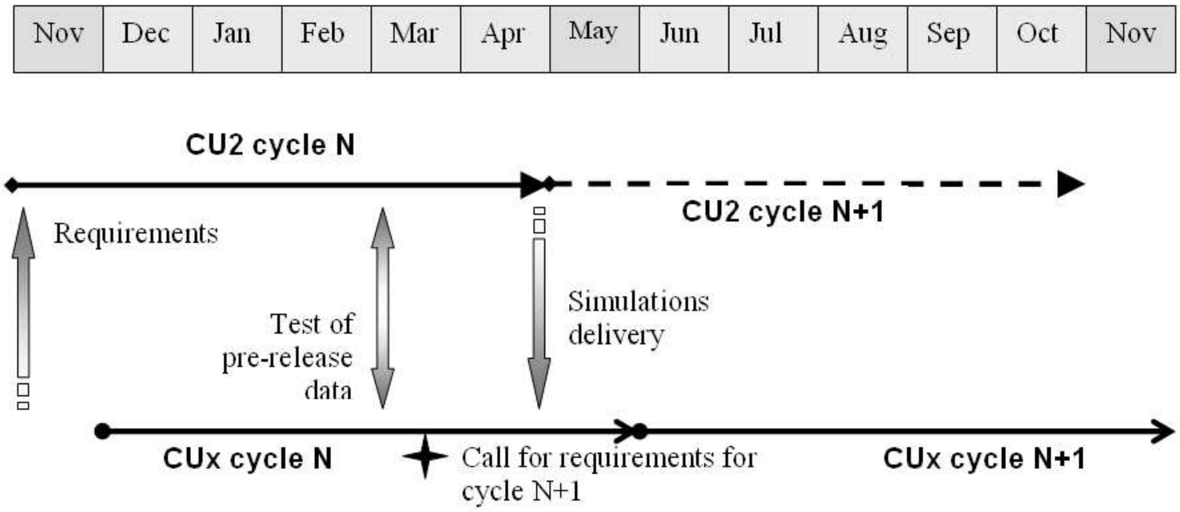


Figure 55: Development cycles for CU2 are in advance of other CUs to make sure simulations are available on time.

if an issues has been raised and accepted. Similarly all hardware systems shall be under configuration control with no changes to operation systems without appropriate requests and tracking.

7.5.4 DPAC Reviews and Testing

The DPAC Consortium intends to put into place a reporting system to insure documentation and quality product assurance, compliant with ECSS standards. Current efforts are to tailor these standards to the DPAC, taking into account its multinational structure and large scientific component; most of the algorithm development will be done by scientists while final implementation will be done by people with proper IT experience, and the documentation standards must allow these two worlds to meet efficiently and effectively. The tailored standards will then be adopted across the DPAC.

Each CU will carry out reviews of documentation and code as part of the QA role within the CU during each cycle. However, a number of DPAC wide reviews, as prescribed by ECSS, are also considered important. In Fig. 54 the approximate dates of these reviews are proposed. Here is a list of the ECSS Review dates with the review purpose and initial deliverables

Phase B 2007 Q4: System Requirements Review
Check requirements have been captured before investing a lot of effort in development. Without good requirements DPAC will not know if it has achieved

its goal. To be reviewed:

- Science Implementation Plan
- Software development Plans (per CU)
- Software Quality Assurance Plan
- Software Configuration Management Plan
- Science Ground Segment Definition Document
- Requirements documents for each software subsystem (min. one per CU).
- Prototype of difficult subsystems.
- Draft End to End System Test plan
- Risk Management Plan and Risk Register
- Initial Interfaces Control Documents
- MOC SOC Interface Requirements Document
- Software Product Assurance Reports

Phase C 2008 Q4: Critical Design Review
Check the design and the approach taken to the entire system development. There is still time to fix major flaws or to assess if a particular CU has a particularly effective approach which could benefit the others. To be reviewed:

- Design Documents, overall and for each software subsystem (at least one per CU).
- End to End System Test plan and Test plans for each software subsystem
- Existing software
- Risk Register

Phase D1 2011 Q2 : Qualification and Acceptance Review
Check requirements have been met and that all tests have been carried out.

- Software User and Installation Manuals
- Software Maintenance Plan
- Detailed Design Documents for each Software Subsystem
- End to End Systems Test report and Test Reports for each Software Sub System
- Risk Register
- Complete Interfaces Control Document(s) (incl. MOC SOC)

These reviews will help DPAC track progress and should highlight possible problems earlier in the development. DPAC welcomes ESA, as well as external participation, in these reviews and in fact sees this as essential to the proper review process.

The exact operational boundaries of the systems remains to be defined. Hence the list of which systems may be included in end to end tests with the ground segment is incomplete. Here the End to End Testing means DPAC testing not including the Mission Operations Centre. Participation in SVT is however foreseen in CU1.

7.6 Work breakdown approach

A preliminary Work Breakdown Structure (WBS) has been defined for the DPAC activities. The following sections describe the schema addopted for the Work Package (WP) descriptions, the WP numbering scheme and a description of a series of management WPs common across all the CUs. The actual WBS is listed in Appendix. B.

7.6.1 Work Package Descriptions

In Appendix. C the top level work packages are listed providing information under a list of headings (derived from ECSS), as follows

Heading:	WP identifier and title .
Project Title:	Gaia DPAC.
Providers:	names people doing the work, entity only if individuals are unknown.
Manager:	WP manager's name and name of deputy where appropriate.
Objectives:	description of the objectives of the WP, what it is to achieve.
Tasks:	description of the tasks, what needs to be done.
Inputs:	list of the inputs necessary to achieve the tasks.
Interfaces:	links with other tasks, WPs, or CUs.
Dependencies:	WPs which must complete for this to begin.
Outputs:	list of the expected outputs e.g. data resulting from this package.
Deliverables:	list of deliverables e.g. software or reports under configuration control.
Start:	expected start date.
End:	expected end date.
Total Effort(MM):	the total approximate effort (in Man Months) required to complete the package over the expected duration. Available effort, or work, is not considered here and should be done as a scheduling activity.
Remarks:	some free text to make comments on the WP.

7.6.2 Work Package Numbering

To facilitate organisational breakdown into CUs and DUs a two-tier numbering scheme has been adopted as follows:

GWP-T-CNN-PPPPP[-CY]

- **GWP** is just a string to denote a work package, **GWP** (for Gaia Work Package) has been adopted.

- **T** denotes a type code which identifies the following generic WP types:

M	Management or top level packages gathering science and technical parts
T	Technical design and analysis of system/framework definition technical modules
S	investigation and elaboration of Science algorithms, definition of modules to plug in framework, analysis of science output

A clear distinction between technical and science packages should be maintained below the top level package. Top-level packages should normally be M type and may contain one or more S or T type packages. M packages may appear again at any level to indicate a management activity. S or T type package may also appear at the top level if an entire package is completely scientific or technical.

When breaking down the S or T types further we may introduce the following packages:

I	Interface, either an ICD development or meetings with members of another CU
C	Component - an algorithm to be plugged in to or a software module forming part of an infrastructure
O	Operations package (operation of system)

In this proposal only the top level packages are presented.

- **CNN** is the top-level WP number, where **C** represents the CU. **NN** are just running numbers. Proximate numbers should mean nothing. Package numbers at this level should be agreed at the DPACE level.

Special numbers exist for common work packages as discussed in Sect. 7.6.3. These **CNN** numbers will be the level that the DPACE will work with — a summary of all top level packages is presented in Appendix. B.

- **PPPPP** provide further structure within the higher-level breakdown. All work packages with the same **CNN** number are considered to be subordinate to the **CNN** package. The **PPPPP** part will also have a few special numbers defined for common packages (Sect. 7.6.3). Several different types of special **PPPPP** packages may exist within a given **CNN** package. The **00000** is considered the 'top level' package grouping the subordinates.

- **[-CY]** An optional cycle number for a work package which repeats in many cycles

(See Sect. 7.5.2).

7.6.3 Common work packages for all CUs (C0N)

Each CU has some common work packages, and these have been given the same number, differing only in the first digit (**x**) CU assignment. The **x0** number of each CU is reserved for common packages. The current list is:

Table 8: Workpackages which are common across CUs.

WP Number	Description
GWP-M-x01-00000	Management and scientific coordination of CUx
GWP-T-x02-00000	Architecture and technical coordination CUx
GWP-T-x03-00000	Quality assurance and config management for CUx
GWP-M-x04-00000	Integration,Validation and Operation of CUx systems

The above exist for each CU, e.g. for CU1 GWP-M-101-PPPPP, GWP-T-102-PPPPP, etc. GWP-C-110-PPPPP and greater would be used for other tasks in CU1, which are not similar across the CUs.

These packages are described in general here while the detailed descriptions for each CU in Appendix. C contains only CU-specific information.

7.6.3.1 GWP-M-x01-00000 Management and scientific coordination of CUx

Manage the coordination unit. Define the product tree/work breakdown structure and the development schedule and management plan for the coordination unit. It is also important to adhere to the schedule once put in place and to monitor development unit progress. In other words all of the classical management tasks. It is now agreed that the schedule for the CU will be done in MS Project such that the different CU schedules may be included in one common project if we wish to have an overview at DPACE level.

The DPACE and Project Scientist should be kept informed through regular reporting and attendance at DPACE meetings.

Interaction with the DPC should be undertaken to ensure sufficient resources to perform the tasks of the coordination unit.

Interaction with the members of the CU should be undertaken to ensure sufficient resources to perform the tasks of the development units.

In addition to the classical management role there is a scientific angle to the coordination unit which must also be considered. Scientific requirements must prioritised and monitored to make sure the software products will deliver the required science.

Tasks and sub tasks of GWP-M-x01-00000

1. Make the CU plan: WBS, planning, milestones, and scheduling
2. Maintain the Risk register.
3. Monitor progress - intervene where necessary.
4. Resource management at DPC level
5. Personnel management at DU level
6. Interfaces with other CUs
7. Reporting to DPACE and Project Scientist
8. Monitoring and action item tracking, follow-up

Deliverables: CU Plan, Regular inputs to DPACE reporting.

7.6.3.2 GWP-T-x02-00000 Architecture and technical coordination of CUx

Generally define the coordination unit architecture and its requirement for hardware, networking, infrastructure etc. This work package will be undertaken in conjunction with, or entirely by, the data processing centre for the CU. More specific requirements will come from individual work packages in the system but these must be consolidated. This WP also considers writing of certain documents which ESA will require such as the Software Requirement Specification (SRS) or Software Design Document (SDD) . Each CU will need to provide input to the operational scenario document and the Ground Segment Design Description document.

Tasks and sub tasks of GWP-T-x02-00000:

1. Define functional requirements
2. Define database requirements
3. Define data model
4. Define data flow
5. Estimate data volumes
6. Documentation

Deliverables: SRS, SDD, other required documents.

7.6.3.3 GWP-T-x03-00000 Quality assurance and config management for CUx

The objective of this work package is to provide the required configuration management services to the coordination unit and ensure that it meets its quality assurance responsibilities. The QA responsibilities are defined in [LJMD$^+$07] 'Product Assurance and Engineering Dispositions for Scientific Development'.

Tasks and sub tasks of GWP-T-x03-00000:

1. Software configuration management. This includes monitoring the state and contents of the CU source code stored in the repository at ESAC and providing input into software releases (ChangeLogs etc).
2. Software release management. Managing the production and delivery of a software release ready for deployment at the DPC, including production or collation of the Configuration Description File, software release documentation and software test results.
3. Operational configuration management. Management of the software and hardware configuration at the data processing centre, including maintenance of the Operational Configuration Reference
4. Issue tracking. Ensuring that non-conformances and change requests are tracked and actioned according to the standards defined in [LJMD$^+$07]
5. CCB activities. Convening the Configuration Control Board.
6. Document and code reviewing for compliance with CU/DPAC standards.

Inputs: The CU management plan and [LJMD$^+$07].

Deliverables: Tailoring of [LJMD$^+$07] with compliance matrix for the coordination unit. In ECSS terms this is the Software Quality Management Plan (SQMP)

7.6.3.4 GWP-M-x04-00000 Integration, Validation and Operation of CUx systems

The objective of this work package is to perform system (DPC) level integration and validation of CUx systems. Given the foreseen cyclical nature of the DPAC development, this work package will be activated multiple times during development and operations phases in line with the overall planning contained in the Software Development Plan of the CU.

Given the formal nature of the acceptance process, completion of acceptance by this work package (for a given cycle) represents a key milestone in the activities of CUx.

This workpackage also contains a sub work package for the involvement in the end to end testing of the entire system (see [OL05]) details will be defined in the CU1 end to end system test plan.

Tasks and sub tasks of GWP-M-x04-00000:

1. Integration of delivered systems
2. Validation and Acceptance testing of integrated systems according to test plans
3. Participation in End to End System Testing
4. Operation of integrated systems during all mission phases

Integration of delivered systems The CUx systems will be delivered according the CU Software Development Plan, written in line with the overall DPAC schedule. The delivery procedure will be described in the Product Assurance Plan. The systems will be integrated into the operational environment according to the System Integration Plan, which defines the procedure to be followed to safely integrate a new release of the systems into an operational system.

Validation and Acceptance testing For each development cycle, the Software Validation Plan will be updated in line with the new or revised requirements for that cycle. Once all CUx systems are integrated, the SVP is performed. The acceptance of the systems delivered in that cycle is based on successful completion of all, or an agreed subset, of the tests in the SVP.

Operations The operation of the CUx systems at the DPC will be conducted according to the Operations Plan, which defines the planned operations of systems through the mission phases and which is agreed at DPAC level.

The operations of the systems themselves are defined in the Operations Procedure Handbook, which describes the routine activities to be performed in operating the CUx systems installed at the DPC.

DPC Workpackages will exist for the operations of individual systems but the CU still needs this work package to ensure handbooks etc. are written.

Inputs:

1. Product Assurance Plan
2. Software Development Plan
3. System Integration Plan

Deliverables:

1. Completed Software Validation Plan
2. Accepted CUx systems for each cycle
3. Operational CUx systems after DPAC Acceptance Review

4. Completed Software Test Plan (including test results)
5. Operations Plan
6. Operations Procedure Handbook

7.7 Software development support

The software development support activities will fall under the responsibility of CU1 but certain efforts are important enough to merit a mention at this level to inform the reader of the ongoing activities.

7.7.1 A common approach to development and product assurance

The consortium feels it is important to have guidelines and procedures to maintain forward momentum. A common work package numbering scheme and software life cycle were seen as essential already in 2005 with the first issue of [OL05]. This approach has been used to define the work packages as described in this document. Indeed the inclusion of CU1 itself as a coordination unit to look after global aspects of the processing system is an important aspect of the DPAC organisation.

As discussed in [O'M05] it is important to lay down rules in the beginning of a project - relaxing them is much easier than instituting them later. Hence we have also considered it quite important to have a common approach to software configuration control and quality management. To this end the common product assurance plan [LJMD$^+$07] has already been drafted and circulated to the DPAC by CU1. The DPAC intends to follow this plan right through to the end of the mission. The QA Document [LJMD$^+$07] lays down the types and schedule for document and code delivery of software components e.g. requirements, designs etc. as per ECSS. All documents will adhere to agreed standards (currently all released documents must be in PDF form and put in Livelink).

To assist developers in the various CUs several workshops have taken place and are in plan for the future (see Sect. 7.7.4). Furthermore it is a key role of the DPC to assist and guide the CUs in technical matters.

7.7.2 A common language - Java

It has long been considered advantageous in the Gaia community to have a common language for the development of the the processing system. This of course leaves the uneasy question of which language to choose. DPAC has selected Java as the main programming language [HOHL06]. It is understood that certain elements of the system may require other languages but the bulk of the code will be in Java.

With this in mind, and again to institute standards sooner rather than later, coding guidelines for Java within DPAC have already been issued by CU1 [OHH+06] early in 2006.

The common language also allows the development of a common toolbox of routines. The GaiaTools library [Ang05] had its origins with the Gaia Simulation Working Group efforts as early as 1999 and continues to grow in usefulness and completeness. The toolbox is already in use for the Java workshops and within the AGIS and simulation developments. The toolbox may ultimately contain the core Gaia Data Model and the ingestion and export software required to adhere to the Main Database ICD [Her05].

7.7.3 Common tools

In addition to the Gaia toolbox several other tools will be used by all CUs, a few examples are given here.

The Gaia MyPortal hosted by RSSD provides a terrific suite of communication and management facilities in use by the community. All released documents are archived in the Livelink system. Planning activities, news and information distribution take place through the Gaia Wiki pages. Mantis, an issue tracking tool, is already in use by several CUs for tracking issues, problems and actions.

For collaborative projects a versioning control system is an essential component. This can be used not only for software code but also for documents being prepared by teams whose members are spread around Europe. The Subversion (SVN) repository hosted at ESAC is already being used by the DPAC community.

The Gaia Parameter Database [Lam03] is perhaps one of the most important common tools in use in DPAC. The intention is to have all constants used in the Gaia software system defined in this central database. From the database a set of classes in Java, and other languages, may be extracted. Indeed the Gaia toolbox contains and relies on these derived classes. Hence all processing software should be consistently using the same set of parameters.

7.7.4 Workshops

As pointed out in [O'M05] the management of a large scientific project such as the Gaia processing is often underestimated and poorly undertaken. Key CU managers have already participated in a specially tailored management course held at ESTEC. This allowed the participants to consider how to break down, track performance, and deliver the processing system as well as contributing to the formation of a team spirit.

Although Java has been in use for the Gaia simulations for several years, Java remains a new language for many in the Gaia community. Moreover a large part of the community has never been involved in a large distributed, collaborative development before. The tools for supporting these activities are new to many. To augment the DPAC productivity, a series of workshops have been planned to introduce the community to collaborative development tools and to the Java language. These are delivered by developers working on Gaia and using the tools regularly. The workshops are tailored to the Gaia tools and use many features of the Gaia toolbox. Participants work on their own laptops and leave with all tools and working examples installed, i.e. a complete development environment. The past workshops have been very positively received.

7.7.5 Technical studies in the areas of computation and infrastructure

The purpose of the technical studies undertaken by the DPCs is to provide answers to any problem that may arise concerning technical feasibility, to sketch out effective solutions for Gaia and to prototype these solutions. To make the best use of the resources available within the consortium, it is essential to avoid any duplication of activity. CU1 is in charge of the overall coordination of the studies. In addition, apart from their special fields of interest, the DPCs need to find solutions to certain problems they have in common. They must therefore not only maintain considerable transparency as to the studies they are working on, but also ensure that their results, the prototypes created and the operational products developed from these prototypes can be used by other DPCs, thereby facilitating the reuse of software components by several DPCs. Each DPC can then adapt these products to its own context if necessary.

We note that as well as satellite related studies ESA started data processing related studies in 2003, e.g. Gaia Grid [tLdWG05] [ALtL05] and Gaia Data Access and Analysis Study [GST$^+$02]. These studies proved useful to confirm the complexity of the core processing system and show how the grid may be harnessed for Gaia processing.

A new set of studies and industrial experiments underway should enable us to clarify processing issues further. These studies, started in mid-2006 by CNES, will start to show results in 2007. They cover the following aspects:

- experimenting with sequencing tools and their evaluation, also the suitability of these tools for distributed processing;
- analysing the toolkit developed by the Barcelona Supercomputing Centre as it has optimisation features which might be reusable for Gaia;
- studying solutions for automatic data transfer between the DPCs and the central database;
- studying data organisation solutions at the CNES DPC level taking into account the specific data models of CU4, 6 and 8;

- consolidating ESAC's work on the DAL (Data Access Layer) to adapt it to the DPC context;
- evaluating various Java compilers/interpreters.

Further studies can then be carried out on the basis of the results obtained. Several DPCs plan to participate.

7.8 Required Processing Power

7.8.1 Global view

In many areas the actual processing required for Gaia remains uncertain. We may consider the estimated one million images per second for photometry, or the 10 TFLOP/s estimate in [Lam06b], or the rough IDT numbers suggesting a sub TFLOP/s machine would work in [PFCT]. Further tests are needed of course, but as of yet nothing suggests that the estimate of between 10^{20} and 10^{21} FLOPs [Per04] is too far away from the truth. Factor in to this the major uncertainty now facing DPAC concerning CTI effects in the CCDs and the high estimate of 10^{21} may not be enough. This also does not include simulations which are needed far sooner and are very demanding in themselves.

Hence taking this number of 10^{21}FLOPs we come to the trade off between time and money. What is an acceptable amount of time to run the Gaia processing over? This number is an integral of all iterative processing - unfortunately this bunches up toward the end of the mission and hence we can not simply spread it over seven years. It is also a reason to buy hardware for Gaia processing as late as possible. Let's pessimistically say we need this amount of processing over one year, it means a computer which can produce 3.17×10^{13} FLOP/s[20] or around 30 TFLOP/s (something like Marenostrum). Of course putting a price on this is very difficult, when one looks for a machine of this nature companies like IBM are happy to reduce cost for marketing, at least if one is doing something interesting. Looking at some prices today it could range anywhere between 15 Million (in house cluster) and 150 Million Euro (certainly an overly pessimistic value). The 1 to 6 Million or so for disk is palatable beside this.

But these are today numbers, Gordon Moore [Moo65] told the world (back in 1965) that component density (read processor power) would double every year. His formula has only been slight revised to eighteen months rather than one year for fitting the last forty years of data. This trend now seems set to last until at least 2017, the production of multi core processors (Intel have hinted at a 64 core processor) leads one to have faith in this. We may be sure that the huge variation in price will not

---

[20] Note that our convention is FLOPs for some number of operations and FLOP/s for some number of operations per second.

diminish but the range outlined above for our Gaia machine in the year 2017 would be between 300,000 and 50 Million[21]. Using a PERT costing on a numbers like this we get about 9 Million for the machine.

Of course this would not be one machine but several spread over the many DPC's. Hence the prices are not too frightening. One of the main complexities faced by DPAC is data access - getting the bits to the processors is the bottleneck in most of the current experiments. Also this is not homogeneous, AGIS needs to run quickly and for it to do its job in one month ESAC will need at least a 10 TFLOP/s machine itself.

It should be clear from the above that buying the Gaia hardware later is the best approach - there is more processing to be done later when more Gaia data is available. Meanwhile the price of hardware will drop dramatically over time. Hence it is foreseen to buy hardware for operations in at least two steps, modestly before launch and again closer to the end of the mission. The second advantage apart from price is an absolute knowledge of the required processing power after one or two years processing real data.

7.8.2 Estimates from current DPAC softwares

In the following we try to scale the current estimate of the computing power requirements based on actual experiments done with the current version of the Data Processing software when this makes sense. Otherwise the amount of computation is assessed in term of the volume of data to be processed over a particular interval of time (day, month or cycle). The hardware power to be available for the DPAC is also considered.

7.8.2.1 Processing power required at CNES DPC

Many of the DPAC processings will be carried out at the CNES DPC (for CU2, CU4, CU6 and CU8) and the computing power availability has been seriously investigated. The risk has been evaluated as part of the CNES Phase A which ended up in December 2006. The evaluation and the needs and how they will be satisfied is detailed in Sect. 9.3.6). This risk has been evaluated as ”medium risk” by a review committee. Although the required computing power looks very impressive for the current standard, the extrapolation to 2012 seems tractable within a large computing centre. In addition for CU4-CU6-CU8 the processing is made of numerous relatively small processings that can be distributed or run in parallel in case of real difficulty. Optimisation is also an area with major impact for nearly identical processings that are repeated billion times.

---

[21] This pessimistic number also assumes Moore's law does not hold true

7.8.2.2 Processing power required for IDT

A first operative version of IDT has been implemented and scientifically tested on the ESAC and CESCA(BPC) computers. From the results of these runs we found that about 7 hours of CPU (in a 70 GFLOP/s machine) are requested to process the data equivalent to one day of real mission. This leads to a total of about 2×10^{15}FLOPs for one day of data. The final IDT will be more complex than the one we have now, which, on the other side, is being optimized. We guess that these effects will somehow compensate, so 6–7 hours daily as a maximum devoted to IDT can be a good estimate. If we assume that ESAC will have a TFLOP machine, we can reduce this figure by a factor of 10, but that will depend on how much of the ESAC's computer is devoted to IDT. Being it a parallel process we can reduce the time using more processors. IDT time will depend on the area Gaia is scanning, so probably, we can use more processors (that is more power) for dense areas and less for the lighter ones, but never using the full ESAC's capability. In conclusion we think that the above estimate is a safe one even for the future ESAC's computer. An additional and very demanding aspect of IDT is the transfer of raw data from ESAC to the DPCs where they are needed to run other processes. This transfer amounts to some 30 GB per day.

7.8.2.3 Processing power required for AGIS

A first functional version of the AGIS system is available at ESAC since end 2005 and has been used since then in a number of testing campaigns to validate the software and also to forecast the processing effort needed to construct the final astrometric solution.

In [Lam06b, Sect. 4] a simple method was proposed to estimate the total computational effort F^{tot} of the operational AGIS as a function of various assumed and measured parameters, viz:

$$F^{\mathrm{tot}} = \frac{L/M+1}{2} \cdot C n_r N_T N_s \frac{T^{\mathrm{ref}} P^{\mathrm{ref}}}{N_T^{\mathrm{ref}}} (I \cdot f_p + f_{\mathrm{src}} - f_p f_{\mathrm{src}}) \qquad (27)$$

A central quantity is the mean number of GIS iterations I (see Sect. 5.1.4) for which recent test results [LR06] suggest a value of not less than 25. It is clear that I sensitively depends on circumstances that are not accurately known now and perhaps will remain unclear until some point well into the mission, e.g. noise characteristics and initial noise levels of the unknowns and how efficiently these will be damped by the iterative scheme. It is believed however that $I = 25$ represents a sound working hypothesis for now.

The tests results presented in [LR06] also largely confirmed the parameter values used in [Lam06b, Sect. 4] which yields:

$$\begin{aligned} F^{\text{tot}} &= 3 \cdot 2 \cdot \frac{72/6+1}{2} \frac{100 \cdot 1.2 \cdot 10^9 \cdot 14440 \cdot 140 \cdot 10^9\,\text{FLOPs}}{100 \cdot 10^6} \cdot (25 \cdot 0.1 + 0.5 - 0.1 \cdot 0.5) \\ &= 2.79 \times 10^{20}\,\text{FLOPs} \end{aligned}$$

The execution of 2.8×10^{20} FLOPs on an assumed operational AGIS hardware platform with a floating point performance of 10 TFLOP/s would hence take about $2.8 \cdot 10^{20}\,\text{FLOPs}/10 \cdot 10^{12}\,\text{FLOP/s} \approx 324\,\text{d}$ ignoring any I/O overheads. Taking into account a conservative factor 2 for this would then suggest a total run time of about 650 d. Note that this is a hypothetical number accounting for the total computational effort over the full six years. Evaluating Eq. 27 for only $i = L/M$ gives the total effort for the final AGIS cycle:

$$\begin{aligned} F^{\text{tot}}_{L/M} &= 3 \cdot 2 \cdot 100 \cdot 1.2 \cdot 10^9 \cdot \frac{14440 \cdot 140 \cdot 10^9\,\text{FLOPs}}{100 \cdot 10^6} \cdot (25 \cdot 0.1 + 0.5 - 0.1 \cdot 0.5) \\ &= 4.29 \times 10^{19}\,\text{FLOPs} \end{aligned}$$

Hence, the last AGIS cycle on the targeted 10 TFLOP/s machine, considering an assumed factor 2 for I/O overhead, will take approximately $4.29 \cdot 10^{19}\,\text{FLOP}/10 \cdot 10^{12}\,\text{FLOP/s} \times 2 = 5.58 \times 10^6\,\text{s} \approx 99\,\text{d}$.

7.8.2.4 Processing power required for IDU and associated processes

The IDU process is not fully defined today. A first list of requirements has been brought together and a prototype is being implemented. From this work an approach that considers the successive running of PSF calibration, 2-D imaging and the IDU proper has been adopted in order to minimize the I/O of the system by putting together processes that make use of the raw data. A crude estimation for this three processes, nowadays recognized as the more demanding of the whole Gaia data processing, gives that some 10^{20} FLOPs are needed at each six-month cycle, thus leaving this tasks for a Marenostrum-like computer, where, according to its present capabilities, they will request at least the 10% of the full machine during a period of 10-20 days. We must recall that it is advisable that IDU and PSF run only once every six months. We should bear in mind that once IDU is finished the data produced (5 TB at the beginning of the mission, 50 TB at the last cycle) must be uploaded to the MDB requiring a very efficient link between BPC and ESAC. Alternative proposals to this transfer, like the physical transport of disks has been considered and could be more easily achievable.

7.8.2.5 Processing power required for SGIS

The Spectroscopic Global Iterative Solution (SGIS), whose task is the calibration of the RVS instrument, is one of the big blocks of the spectroscopic processing system. To calibrate the RVS over the 5 years of the mission, the SGIS will have to process about 300 millions spectra. A first prototype of SGIS has been developed over the past 3 years. It allowed us to perform an order of magnitude estimation of the total FLOPs for 5 years of RVS calibration: i.e. $\sim 4 \times 10^{18}$ FLOPs. For a 40 GFLOP/s computer (reasonably powerful computer by todays standard), the full calibration of the RVS for 5 years should require ~1100 days (including a safety margin of 2). The increase of the computing power by 2012 (and even more by 2017) should reduce the effective required time to significantly less than a year.

The full spectroscopic processing system is made of 3 other large blocks (i.e. extraction of the spectra, single transit analysis, multiple transit analysis) of similar sizes as SGIS.

7.8.2.6 Processing power required for simulations

The estimation of the processing power required for simulations can be reliably extrapolated from the simulations run in the last years, given that the simulator team has been producing simulated data for quite a while already.

The estimations are discussed in terms of the processing power needed every cycle (six months) up to mission launch, that is, five years from now. We can split this estimation in three parts, corresponding to the three data generators built in the Gaia simulator.

GIBIS: this data generator has been running in a linux single-node station for some years and has been recently moved to a CNES cluster to provide more scalability.

The CNES cluster consists of a set of 24 biproc Opteron, 12 biproc being affected to operational activity, GIBIS simulation jobs are running on this operation part. The frontier between development part and operation is tunable if necessary. Up to now the CPU consumption of GIBIS has been moderate, about 270 CPU hours were used from mid January to end of March 2007, and has been easily accommodated in the available hardware. We expect the demand to increase in future cycles, peaking some time before launch. In this peak GIBIS will be requested to generate maximum-realism simulations for longer time intervals in order to test several of the data processing components. The present environment at CNES can cope with an increase of a factor of 100 with respect the present demands for normal operations and could be complemented with extra processing power from the Mare Nostrum supercomputer to cater for one-shot, high demand simulations.

GASS: this data generator is probably the most demanding in terms of CPU consumption. It has been designed from the start to generate large amounts of realistic telemetry and has been run in supercomputing centers (CESCA, BSC) for several years. GASS has already generated large simulations for IDT/AGIS 15 million objects, five years of observations, about 1.5 TB of data) and the demands on it will grow with the cycles: 30 million objects in 2008, 150 million in 2009 and up to 400 million in 2010-2011 (or up to full mission, one billion, if possible), with increasing complexity. The largest runs have been executed in the Mare Nostrum supercomputer, and extrapolating from the available data and taking into account the increase in complexity, a full-mission (one billion objects) run would be feasible, but would take a large fraction (about 50%) of this supercomputer during a six-month period. Although very difficult, even this extreme case is conceivable, even more taking into account the planned upgrades of Mare Nostrum before 2011. The storage, however, might become a problem given the large amount of generated data, but can be alleviated by immediate transfer to the users as soon as the data is generated.

GOG: this data generator is an intermediate case between the two previous ones. In some cases it will be used to run massive simulations at style of GASS, while on the other it will be used to run specific tailored simulations at the style of GIBIS. It is also the one for which less previous information is available, given its recent release.

In the case of GOG the problem of computing resources is eased by the fact that it will be run in several sites, including CNES, the BSC but also local computers or clusters in various institutes participating in the DPAC. This distributed nature will help accommodating the needs during the preparation of the data reduction, and in extreme cases the Mare Nostrum will be available for large runs.

7.8.2.7 Processing power required for CU7

The most CPU demanding task in variability processing (CU7) is the period search. Because of the irregular nature of the time sampling of the measurements, optimised methods, such as the Fast Fourier Transform (FFT) cannot be applied. All relevant algorithms involve the calculation of some kinds of amplitude for each value of a large set of possible periods. The number of periods that should be evaluated and compared is very large, of the order of a few tens of thousands, and consequently the processing is relatively slow. The Lomb-Scargle method is a relatively fast such method that has been used to analyse OGLE data. Applying this method to the about 10^8 Gaia sources found to be variable from simpler statistical tests requires a computing power of the order of 2×10^{18} FLOPs at the end of the mission when an average of 80 measurements per source will be available.

As of today, a common AMD A64 or Intel Pentium 4 general-purpose PC has a power of around 10^{10} FLOP/s. With a PC farm including about 200 CPUs, it would be possible to process variable Gaia sources with the Lomb-Scargle algorithm in about 10^6 seconds, i.e., in about 10 days.

These numbers are currently used to dimension the hardware required at the variability data processing center. It is assumed that a processing of less than 10^{19} FLOPs will be achieved every 6 months on a PC farm containing a few hundred CPUs. It is also anticipated that the FLOP/s per CPU will increase in the coming years even though it may not follow Moore's law anymore.

7.8.2.8 Processing power required for CU8

The final algorithms to be used for the classification and astrophysical parameter estimation have not been chosen: part of our work is to identify and tailor the optimal algorithms. However, we can calculate an approximate figure using those algorithms for the Discrete Source Classifier (DSC) and Generalized Stellar Parametrizer (Photometry) (GSP-phot) (see section 5.5.3) created during DPAC development cycle 2. In both cases the algorithm is a support vector machine (SVM), used in classification mode for DSC and regression mode for GSP-phot. Extrapolating our tests on 8000 sources to an assumed 10^9 sources, the FLOPs required are 6×10^{14} for DSC and 6×10^{15} for GSP-phot. (The training times are, in comparison, much shorter.) The FLOPs for the binaries parameterizer (GWP-S-834) is assumed equal to GSP-phot.

The other algorithms in CU8 will use similar methods but will operate on a much smaller number of objects, so we simply assume that all together they also require the same FLOPs as GSP-phot. All of the CU8 algorithms will run several times during the mission, e.g. at each operation cycle. We assume they all run ten times on the whole data set. Putting this all together, and increasing the result by a factor of four as a margin (e.g. for nonlinear extrapolation) gives a total requirement of 8×10^{17} FLOPs. Armed with a 100 GFLOPs/s computer, this would require 86 days spread over the whole mission and post-processing period, or about 10 days per year. Note that most CU8 algorithms operate on sources individually, so sources can be processed in parallel. It must be emphasised that this is a requirement which could get the job done using today's algorithms. We will develop more sophisticated algorithms (i.e. which make more accurate predictions, which also provide individual uncertainty estimates) which may well require more FLOPs.

Part III

The DPAC

8 Structure of the Consortium

This section describes the overall organisation that has been adopted for the Consortium and presents its top level divisions into *Coordination Units*. A subsections describes the objectives of each of the units, their boundaries and interfaces, with appropriate references to Sec. 9 dealing with the Data Processing Centres.

8.1 Overall philosophy

8.1.1 Coordination Units

The central concept introduced in this document is to build the Gaia Data Analysis Consortium around a series of 'coordination units' (CU), a breakdown level introduced in 2005 by F. van Leeuwen. The CUs are small in number, with clearly-defined responsibilities and interfaces, and their boundaries fit naturally with the main relationships between tasks and the associated data flow. There will be several areas of involvement across these boundaries, but in the first instance it is up to the coordination units to ensure that a group of tasks is prepared and optimised, as well as fully tested and documented, as required by the project.

The coordination units will have a reasonable amount of autonomy in their internal organisation and in developing what they consider as the best solution for their task. However they are constrained by the fact that any such solution has to meet the requirements and time schedules determined by the Consortium Executive for the overall data processing. In this respect the data exchange protocol and the adherence to the DP cycles are mandatory to ensure that every group can access the data it needs in the right format, with the required content and at the right time. The coordination units reflect the top level structure of the data processing, with well-defined responsibilities and commitments to the DPAC. For practical reasons they are, in some cases, organized into more manageable components, called development units (DUs). This a more operational level with a lighter management which will take the responsibility for the development of a specific part of the software with well defined boundaries. Not every CU will organise its DUs (if any) in the same way and how they interact is left to the CU management.

The tasks of the CU are

1. Coordinate and supervise the activities of their constituent DUs, by providing a clear, one-dimensional task allocation and reporting structure (i.e. WP; DU; CU; DPACE). Each coordination unit is headed by a scientific manager supported by his/her deputy who are responsible for these tasks. Full-time Gaia posts are considered mandatory for these positions. Specific tasks include defining in detail of the DP tasks and establish the internal CU structure, development

priorities, schedules, development plans, simulation requirements etc.

2. Write (or procure), optimize and test algorithms which fulfill the DP requirements. This is the bulk of the work to be done by the CU members.
3. Convert the algorithms and/or software into Gaia-compatible software. In some cases, DUs may have their own local expertise to produce this. This is is is certainly desirable because it simplifies communication, problem tracking etc. In other cases, however, a DU may lack the manpower and expertise for this, in which case this would be a service provided by software engineers associated with the CU itself, perhaps through the DPC. Flexibility rather than a rigid structure is required here to reflect individual needs.

Each CU manager will report periodically to the DPACE on the work under development, the updated schedule, the technical design and will communicate any problem area (technical, manpower or funding) that impacts on the DPAC activities. These reports will form the basis of the DPAC advancement reports to ESA. The CU managers will be invited to attend the Gaia Science Team meeting with the status of observer.

The list of Coordination Units making up the DPAC with their manager and deputies is given in Table 9 and has been agreed by the executive of the Consortium. The CU list should not change over the lifetime of the DPAC (although the managers may) and constitutes the backbone of its internal organisation upon which the development, testing, implementation and operations phases will gradually unfold.

Attached to each CU there is at least one Data Processing Centre (DPC) Sect. 9. The DPC is responsible for the computer hardware and software infrastructure which will be used to carry out the actual processing of the data. During the software development phase (prior to launch) the DPC plays an important role concerning integrating software, defining software interfaces to the data, testing algorithms, running system-wide tests, measuring performances and providing general software support to the rest of the CU. The DPC comprises primarily software and hardware engineers. A technical manager from this center belongs to the management structure of every CU. The scientific software development and its implementation and operation in a DPC are parallel activities. Their mutual adequacy must be closely monitored by the CU manager and the DPC manager. Note that some CUs have more than one DPC, and some DPCs support the activities of more than one CU (Fig. 56).

Sections 8.2 to 8.10 describe each of the CUs in more detail. Due to the significant autonomy of the CUs and the freedom they have in their internal organization, the reader will note the somewhat heterogeneous treatment and presentation between these sections. The DPAC believes that external management and standards

Table 9: *List of Coordination Units, managers and deputies in the DPAC.*

Label	Name	Manager, deputies
CU1	System architecture	W. O'Mullane, U. Lammers, T. Levoir
CU2	Data simulations	X. Luri, C. Babusiaux, F. Mignard
CU3	Core processing	U. Bastian, J. Torra, M. Lattanzi
CU4	Object processing	D. Pourbaix, P. Tanga
CU5	Photometric processing	F. van Leeuwen, A. Brown, C. Cacciari, C. Jordi
CU6	Spectroscopic reduction	D. Katz
CU7	Variability processing	L. Eyer, D.W. Evans, P. Dubath
CU8	Astrophysical parameters	C.A.L. Bailer-Jones, F. Thévenin
CU9	Catalogue access and scientific exploration	(to be activated later)

should only be imposed where they are necessary to insure timely delivery of quality software which meets DPAC requirements. To minimize unnecessary management overheads, other internal CU issues are left to the discretion of the CU. The DPAC executive is the glue which binds the CUs. It comprises all of the CU managers and so is the forum for good communication between the CUs; it sets DPAC-wide standards, requirements, priorities and schedules and defines the inter-CU boundaries (see Sect. 10.4).

8.2 CU1: System architecture

ESAC, as part of ESA's contribution to the DPAC, as specified by the SMP makes a substantial contribution to CU1 in the fomr of leadership and manpower. The services and resources provided to CU1 would be availble to any consortium, and are presented here for completeness. A detailed accounting of these resources are given in Appendix. D.1.

CU1 takes the lead in helping the DPAC define the overall system processing philosophy, architecture and strategy.

CU1 will provide advice and support to the DPACE and DPAC in the areas of software design and technology. CU1 does not intend to enforce rules, rather the intention

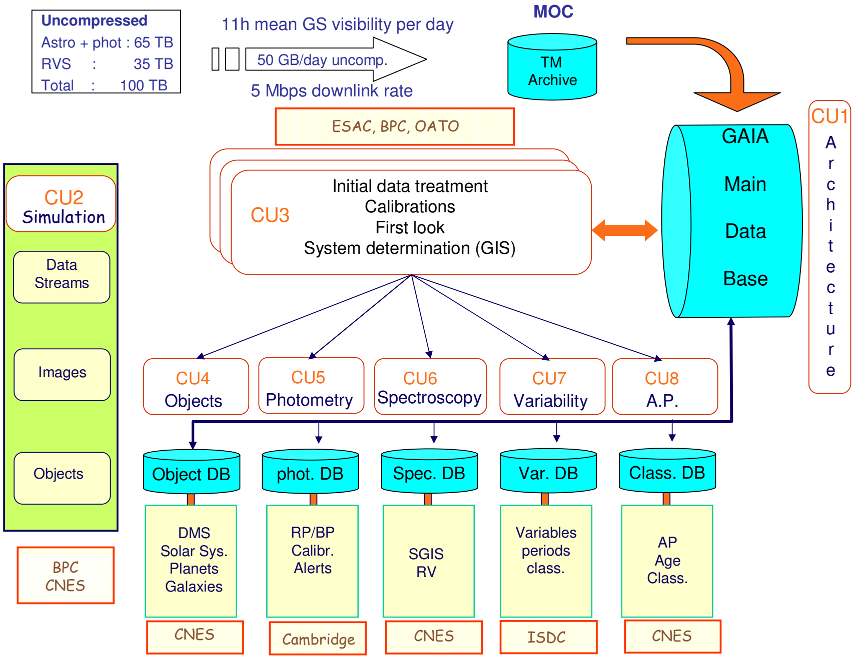


Figure 56: Relationship within the DPAC between the coordination units, the data flow and the data processing centers.

is to make recommendations which should be considered by each CU. Rules must be agreed and enforced at DPAC level. CU1 shall draft coding guidelines, product assurance plans, configuration management guides etc. covering the entire DPAC scope - the DPACE must enforce the agreed rules and standards.

Additionally some central/common software and ICDs fall in the remit of CU1. For example the main database, the Gaia transfer system an the Toolbox are considered central and used by all CUs hence it is felt CU1 should take responsibility for these.

8.2.1 Structure of CU1

The CNES and ESA teams are the main players in CU1. The coordination unit is lead by William O'Mullane (ESAC) and the Deputy chairs are Thierry Levoir (CNES) and Uwe Lammers (ESAC). The current top level Organigramme is as shown in Fig. 57.

Each of the blocks in Fig. 57 may be seen as a development unit with the identified person in charge of that unit. The top level workpackages are summarised in Appendix. B while detailed descriptions may be found in Appendix. C.2.

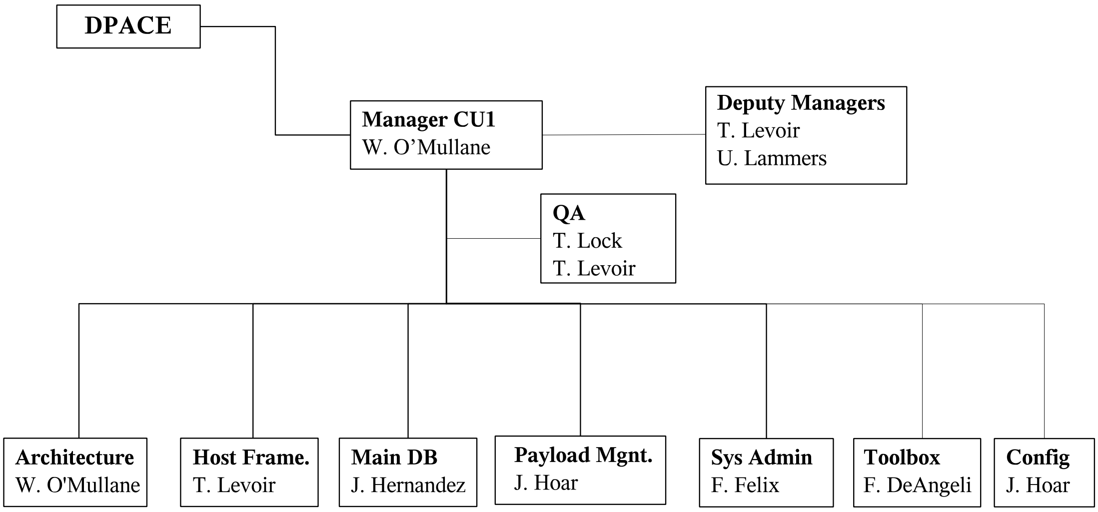


Figure 57: CU1 Organisation Diagram showing major areas of work and the person responsible for that area.

8.2.2 Milestones and schedule for CU1

CU1 will follow a cyclical development up to launch as outlined in Sect. 7.5.2. The complete details of each cycle are refined in the requirements gathering phase at the beginning of the cycle but some milestone highlights are mentioned here.

8.2.2.1 Up to Launch

- **Cycle 1:** Preliminary system up with difficult technology prototyped. MDB initial schema from ICD , export in ICD format, transfer to other DPCs. MDB deals with 3 Million Objects.
- **Cycle 2:** MDB preliminary Data Dictionary system. Initial Integrator. MDB deals with 10Million Objects. Preliminary requirements and design for Payload Operations Software.
- **Cycle 3:** MDB Data Dictionary on line for CUs. CU4,5,6 in ICD and integrator. Preliminary implementation of Payload Operations software.
- **Cycle 4:** ICD and Integrator includes CU7,8. MDB Deals with 50 Million Objects.
- **Cycle 5:** Fully specified and well on the way to implemented system. MDB full schema and ICD incorporating all CUs. GTS up and running. Transfer in and out from all DPCs.

- **Cycle 6:** ICD and Integrator includes CU7,8.
- **Cycle 7:** More complex data - issues from other CUs.
- **Cycle 8:** More complex data - issues from other CUs.
- **Cycle 9:** More complex data. MDB deals with 100 Million objects. Payload operations software complete for MRR.
- **Cycle 10:** Complete system working with as near to real simulation data as is available. More complex Data in MDB. Fully tested and qualified MDB and Payload Operations system.

8.2.2.2 Post Launch Operations

More detailed planning is required for operations in the next two years. However it is clear that CU1 will need to maintain and operate the MDB. The schedule of MDB versions will need to be agreed with DPACE. Data will have to be received and transmitted to other DPCs.

CU1 also has responsibility to provide ESOC with calibration and time line changes for the satellite information. This requires providing ESOC with information resulting from DFL in an appropriate format.

8.2.2.3 Catalogue Production

Payload operations will cease. Although no more satellite data will be downloaded in this period the transfer of data between DPC will continue over the two years. At least three iterations will be required of the MDB after the last data is received from Gaia.

8.2.3 CU1 interfaces with other CUs

CU1 has programmatic interfaces to all other CUs via the main database which shall be governed by the ICD [Her05]. The main database is described in more detail in Sect. 7.3. CU1 is the hub of DPAC as depicted in Fig. 58. The hub and spokes architecture reduces interdependencies between the individual CUs but puts considerable pressure on CU1 to perform its tasks efficiently and accurately. CU1 also has interface with all other CU1 from the perspective of standards and tools. For this reason CU1 is seen as an advisory body to the DPACE and somewhat orthogonal to other CUs.

8.3 CU2: Data simulations

The main task of the CU2 is to develop a software system capable of covering the simulation needs of the Gaia Data Processing Consortia. For this the CU2 will need

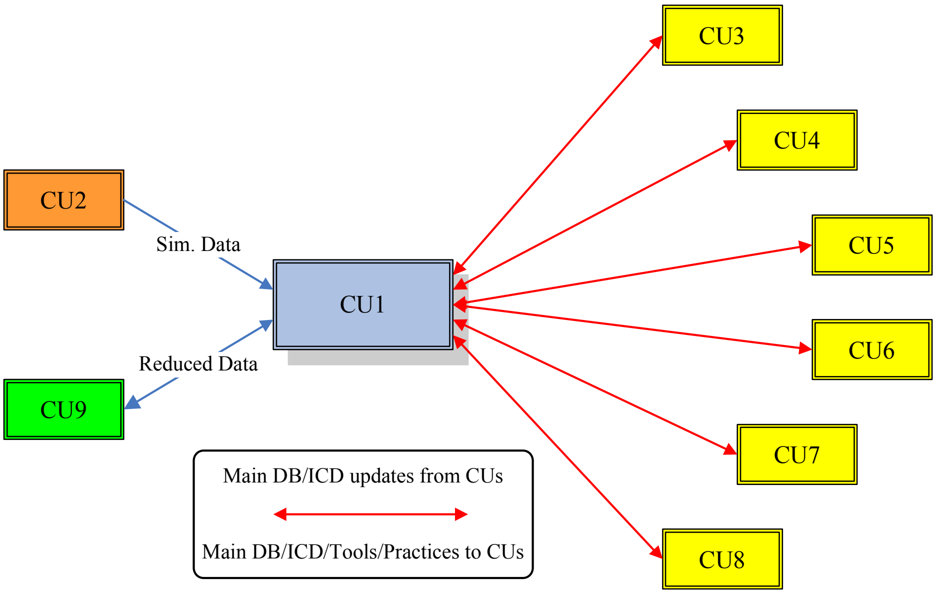


Figure 58: CU1 Interaction Diagram

a strong software engineering base, capable of handling the development of such a complex system in a professional way. However, software engineering competence alone is not sufficient for the task ahead; a strong scientific component is also needed to ensure that the system fulfills the scientific needs of the DPAC. DPAC, after all, has the essentially scientific task to reduce the Gaia data and produce the Gaia catalogue.

8.3.1 Structure

The CU2 structure will reflect this dual nature and, as described in Sect. 6.1, will be organised around four teams

1. A core software engineering team
2. A scientific for the development of a Universe Model
3. A scientific team providing the expertise to develop models of the Gaia spacecraft and its instruments.
4. A Quality Assurance and Validation team (QA&V)

This structure (Fig. 43) is mapped into Development Units, as discussed in the next paragraphs and depicted in Fig. 59, with clear definition of responsibilities and interfaces.

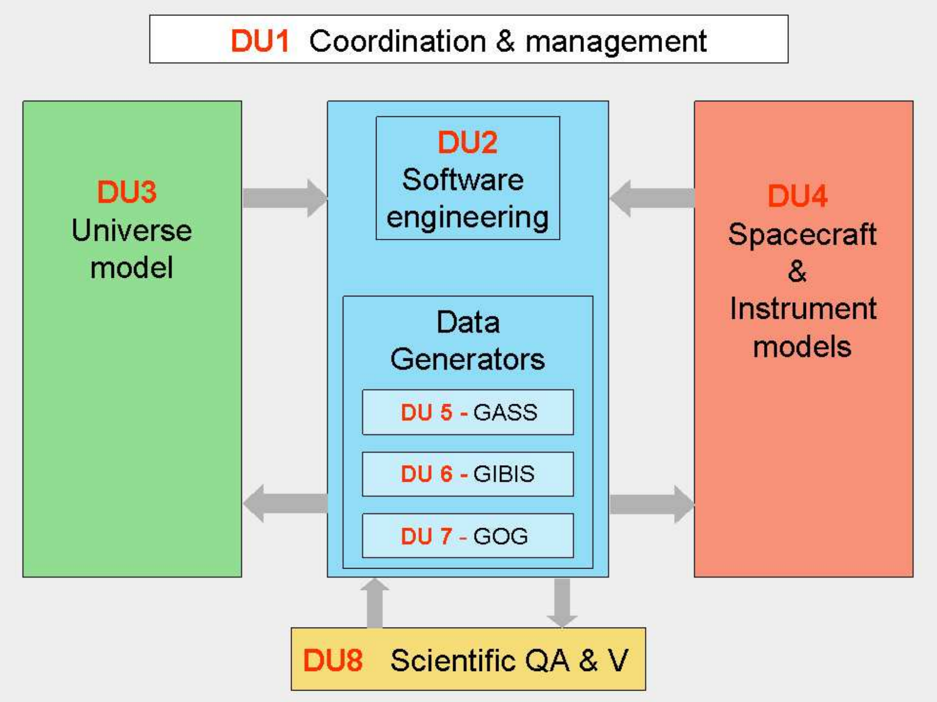


Figure 59: CU2 breakdown into development units, mapped into the basic structure of the CU

Taking this structure as reference, the CU2 is subdivided into Development Units (DUs), which will take the responsibility of specific parts of the CU2 overall task.

DU1 Coordination & management

Manager: X. Luri (University of Barcelona)

This DU is a placeholder for the management structure of CU2. It is in charge of the overall coordination and management of CU2, from requirements gathering to project control, schedule definition and priority selection. It is constituted by the CU2 and DU managers.

DU2 Software engineering

Manager: J.M. Wallut (CNES)

This DU is in charge of the software engineering aspects of the project, a core software engineering team able to professionally manage the development of a complex software system, the Gaia simulator. This team should be mainly composed of software engineers but should also include adequate scientific expertise to ensure proper scientific management and coordination with other CUs.

DU3 Universe model
Manager: A. Robin (Observatoire de Besançon) **Deputy manager:** C. Reylé (Observatoire de Besançon)

This DU is in charge of the definition, development and integration of Universe Model into the Gaia simulator. The team composing this group should be essentially made of scientists, in charge of interacting with the wide European Scientific Community to gather the necessary expertise to build the models of the wide variety of objects to be observed by Gaia, but should also be competent in software engineering to enable its integration in the Gaia simulator with the support of the other DUs.

DU4 Spacecraft & instrument models
Manager: M. Gai (Osservatorio Astronomico di Torino) **Deputy manager:** J. Rebordão (INETI - DSIC)

This DU is in charge of the definition, development and integration of the spacecraft and instrument models into the Gaia simulator. The team composing this group should be a mix of engineers, software engineers and scientists whose task will involve close coordination with the industrial teams building Gaia through the project team and with the CUs involved in instrument design.

DU5 GASS
Manager: E. Masana (University of Barcelona)

This DU is in charge of the development of the GASS data generator, that will provide simulations of the telemetry stream of the mission. The simulations use some simplifications of the instrument and Universe models allowing a large amount of data to be simulated over a significant period of time. The building of the data generators presents very specific needs for code development and coordination with other DUs inside CU2 and also with other CUs, and therefore have been grouped together in its own DU. The team in this DU should be a mix of software engineers and scientists, working in close cooperation with the other DUs to integrate the developed code into functional data generators. A description of GASS can be found at [YIL06].

DU6 GIBIS
Manager: C. Babusiaux (Observatoire de Paris-Meudon)

This DU is in charge of the development of the GIBIS data generator, that will provide simulations of the data at the pixel level. The resulting simulations should be as realistic as possible for a limited region of a sky and over a short period of time. The building of the data generators presents very specific needs for code development and coordination with other DUs inside CU2 and also with other CUs, and therefore have been grouped together in its own DU. The team in this DU should be a mix of software engineers and scientists, working in close cooperation with the other DUs to integrate the developed code into functional data generators. A description of GIBIS can be found at [Bab06].

DU7 GOG
Manager: X. Luri (University of Barcelona)

This DU is in charge of the development of the GOG data generator, that will provide simulations of number counts and lists of observable objects from the Universe model and, for a given source or a collection of sources, simulations of intermediate and end-of-mission Gaia data. The building of the data generators presents very specific needs for code development and coordination with other DUs inside CU2 and also with other CUs, and therefore have been grouped together in its own DU. The team in this DU should be a mix of software engineers and scientists, working in close cooperation with the other DUs to integrate the developed code into functional data generators. GOG is now in design phase, and a proposal for its implementation can be found at [EM06].

DU8 Scientific Quality Assurance & Validation
Manager: D. Egret (Observatoire de Paris)

This DU will be in charge of the quality assurance and validation of the Gaia simulator. This will include checking compliance with requirements and design of tests for the validation of simulations, both internal (self-consistency) and external (comparison with real data). The team composing this group should be essentially made of scientists to ensure the scientific validation of the simulated data, with also some software engineering expertise. This DU will be activated in 2007 once CU2 is fully operational.

8.3.2 Milestones and schedule

As discussed in Sect. 6.7, the simulator development and the production of Gaia simulated data started already some years ago. Due to this fact, the development of the Gaia simulator is in a very advanced state, even before the formal constitution of the consortia.
Starting from this existing system, the plans for the development are aimed to increase the level of detail and realism of the simulations. The goals, milestones and schedule for simulations are also presented in Sect. 6.7. The top level workpackages are summarised in Appendix. B while detailed descriptions may be found in Appendix. C.3.

8.3.3 Interface with other CUs

As discussed in Sect. 6.1.3, the simulation needs in a given development cycle will substantially depend on the schedule and development status of the rest of the consortia. Therefore, an essential element for the development of the simulator is a good interaction with the "consumers" of simulated data, that is, with the members of other CUs.

To cover this need, clear interfaces with the rest of the DPAC CUs have been defined as part of the CU2 structure and are established in two ways: either through the manager of one of the CU2 development units (listed in Sect. 8.3.1) or through one of the so-called *CU2 thematic coordinators*.

Thematic coordinators have been introduced to ensure that some critical areas are properly coordinated in all the simulation activities through the boundaries of the development units. There are three thematic coordinators in CU2:

Astrometry coordinator: C. Fabricius (University of Barcelona)

Spectroscopy coordinator: P. Sartoretti (Paris-Meudon Observatory)

Photometry coordinator: C. Jordi (University of Barcelona)

The interface with other CUs is therefore defined as follows:

CU1: established through DU2 manager

CU3: established through the astrometry coordinator

CU4: established through the DU3 and DU7 managers

CU5: established through the photometry coordinator

CU6: established through the spectroscopy coordinator

CU7: established through the DU3 and DU7 managers

CU8: established through the DU3 and DU7 managers

8.4 CU3: Core processing

CU3 ("Core Processing") covers the entire processing chain going from the raw telemetry to the astrometric core solution. In particular, the unit

- receives the raw telemetry from the Gaia ground segment, unpacks and decompresses it, and ingests it into the 'raw database' at the Science Operations Centre.

- processes the raw windows to provide astrometric and photometric parameters for images
- provides the assignment of images to sources through the cross-matching process
- thereby constructs the main source catalogue
- stores the results of all these operations in the Main Database at the Science Operations Centre.
- does the daily First Look processing for data quality and instrument health monitoring, including daily CCD, PSF/LSF, astrometric, photometric and RVS calibrations
- produces the astrometric core solution, i.e. the AGIS
- undertakes an independent verification of the raw-data treatment and of the AGIS
- produces and assesses trial 5-parameter astrometric solutions for all sources
- coordinates the relativistic modelling and tests relevant for Gaia

Evidently, several of the tasks of unit CU3 will require very close interaction with the photometric, spectroscopic and object processing CUs, both for the algorithm development and for the actual processing. The boundaries of CU3 are defined mainly by the interfaces to those other CUs, and by the activities of CU1. In particular, software modules for most of the above-mentioned daily calibrations will be provided by other CUs. They will be coordinated and integrated into the Initial Data Treatment (IDT) and First Look (FL) processing chain and operated by CU3.

8.4.1 Structure

8.4.1.1 Motivation At its first meeting (Heidelberg Feb 23/24, 2006) CU3 has agreed on the organizational structure listed in the next subsection. It was constructed in accordance with boundary conditions set by the DACC (through the minutes of the various DACC meetings and through the guidelines set out in GAIA-C1-TN-ESAC-WOM-001), and in accordance with previous developments within the former Gaia working groups (those relevant for CU3). The structure aims at preserving the pre-existing motivation, expertise and creativity of the (now dissolved) Gaia working groups, and transfer these resources to the forming CU3. Traditionally, what has become CU3 now, has consisted of five major scientific ‘building blocks’ (with their informal scientific leaders indicated):

Astrometric core solution (AGIS)	L. Lindegren
Relativistic models and tests (REMAT)	S. Klioner
Initial data treatment (IDT)	J. Torra
First Look (FL)	S. Jordan
Astrometric verification (AVU)	M. Lattanzi

Two more such 'building blocks' were added in the course of the formation of CU3:

Intermediate-data updating (IDU)	J. Torra
The ESAC data processing centre	W. O'Mullane

The agreed organizational structure of CU3 reflects and represents all these building blocks.

8.4.1.2 The agreed CU3 structure The formal structure consists of the following items. The names of the leading persons (since February 2006) are given.

CU3 Scientific Manager:	U. Bastian
Deputies:	J. Torra M. Lattanzi
CU3 Technical Manager:	W. O'Mullane
Deputy:	U. Lammers
CU3 Data Processing Centre:	Villafranca, ESAC, W. O'Mullane
Subcentres:	Barcelona (for IDU) Torino (for AVU)

For the practical coordination of the work of CU3, a 7-person CU3 Steering Committee was set up, representing both the formal structural items listed above, and the topical 'building blocks' listed above. The members are listed in alphabetical order below. The parentheses behind the names indicate the aspects represented by each of the persons.

CU3 Steering Committee:

U. Bastian (chair, scientific manager)
S. Jordan (FL)
S. Klioner (REMAT)
M. Lattanzi (deputy scientific manager, AVU)
L. Lindegren (AGIS)
W. O'Mullane (technical manager, data processing centre)
J. Torra (deputy scientific manager, IDT, IDU)

The actual data processing development work, as well as the operations and interpretation of the data after launch, are done in 12 top-level workpackages, which are subdivided into lower-level workpackages as appropriate. These top-level workpackages are summarized in Appendix A and described in some detail in Appendix B. To each top-level workpackage there is a manager who is responsible to organize and survey the work, in accordance with the CU3 Steering Committee and the guidelines set by the DPACE. In some cases there are two managers, a scientific and a technical one who are shall closely cooperate. In the case of the overall management and coordination workpackage (GWP-M-301-00000) there are two deputy managers, which are the deputy scientific coordinators of CU3.

8.4.1.3 Contributors As of November 21, 2006, the membership list of CU3 contains 63 entries. They correspond to about 30 full-time equivalents. This is not yet considered fully satisfactory, but the resources are presently being built up at all the major contributors. The full level needed will be reached in the course of the next two to three years. Taking this into consideration, the overall funding situation of CU3 is tight but not critical.

The most important contributors are the ESA (the Gaia Science Operations Centre at Villafranca), Germany (Astronomisches Rechen-Institut at Heidelberg University and Lohrmann Observatory at Technical University Dresden), Italy (INAF, Osservatorio Astronomico of Torino), Spain (Departament d'Astronomia i Meteorologia at the University of Barcelona) and Sweden (Lund Observatory at Lund University). Tab. 10 lists all the contributors.

The top level workpackages are summarised in Appendix. B while detailed descriptions may be found in Appendix. C.4.

8.4.2 Milestones and schedule

CU3 will follow development cycles as outlined in Sect. 7.5.2. Bellow a few high level milestones are listed. The cycle 0 and cycle 1 milestones have already been achieved.

- End 2005: Prototype astrometric GIS running at ESAC
- Mid 2006: Common Gaia software framework (prototype) ready at ESAC. Gaia Parameters Database represents “Gaia-3”; framework uses it. Prototype IDT modules (adapted from GDAAS) running at ESAC
- End 2006: End of development cycle no 1; some new modules/algorithms from CU3 delivered and integrated into the common software framework at ESAC

Table 10: Home institutions of the CU3 members. Short names for the institutions are given, along with the town and country in which they are located. The numbers in parentheses are the total number of CU3 members in each location (as of Sep. 18, 2006). The number of scientific personnel actually involved with CU3 work may be higher. On the other hand, not all members contribute their full time to CU3.

INAF-OATo, Torino	Italy	14
Univ. Barcelona	Spain	9
ARI, Univ. Heidelberg	Germany	8
ESAC, Villafranca	Spain	8
Lohrmann Obs., TU Dresden	Germany	4
IMCCE Paris	France	3
Univ. Leicester	UK	2
Obs. Paris-Meudon	France	2
ZARM, Univ. Bremen	Germany	2
ESTEC, Noordwijk	Holland	1
OCA, Nice	France	1
Lund Obs., Univ. Lund	Sweden	1
Univ. Besancon	France	1
Univ. Leiden	Holland	1
Univ. Padova	Italy	1

- October 2007: System Requirements Review for CU3

- End 2007: End of development cycle no 3; full-featured software framework ready at ESAC ; at least prototypes of all main algorithms delivered and integrated at ESAC

- 2008–2010: Development cycles nos. 4–9; quasi-continuous integration and updating of algorithms into the framework, with gradually more detailed functionality and more realistic adaptations to the real Gaia

- Jan–Apr 2011: Software Readiness Reviews (subsystem level)

- Jun 2011: End of development cycle no. 10 (6 months before launch); Ready for Software Readiness Review (system level)

- Dec 2011: Launch; start of operations in space and in the data processing centres; IDT and FL running daily.

- early 2012–2017: IDT and FL running daily; IDU, AGIS and AVU running about semi-annually in the global iterations according to the overall system architecture; all processing softwares being maintained and updated as needed; all output data continually being validated by the responsible scientists.

- early 2017: End of operations in space (for an assumed 5-year scientific mission); end of IDT and FL processing.
- 2017–2019/20: Post-mission processing; IDU, AGIS and AVU running about semi-annually; final updates of processing softwares; final cleaning and validations of output data in the last few global iterations according to the overall system architecture; preparation for catalogue production with CU9.
- about 2019/2020: End of data reduction operations; final delivery of all data products to MDB; catalogue production with CU9.

8.4.3 Interfaces with other CUs

As for other CUs interaction and data transfer will be through the MDB and governed by the MDB ICD. With that in mind we list here specific points of interest.

- **CU1:** The interface to CU1 consists of
  - the basic processing infrastructure (software processing framework and hardware) of the Gaia SOC, into which the CU3 processes (IDT, IDU, FL, AGIS) have to fit
  - the interface control documents for the main database (MDB), GAIA-C3-SP-ESAC-JH-001, and for the raw database (RDB)
  - common software toolboxes and other common software resources by CU1
  - software coding rules by CU1
  - software configuration control and quality assurance by CU1
- **CU2:** This interface consists of the data streams simulated by CU2, and their definitions/descriptions. Details of the individual streams will be defined as the project goes along. A first specific data set, delivered in mid 2006, is defined in GAIA-C3-SP-ESAC-UL-016-1. — All requests from CU3 to CU2 are to be channeled through and coordinated by the CU3 manager, according to a well-justified wish of CU2.
- **CU4:** The data flow from CU3 to CU4 and vice versa is organized through the MDB ICD and the controlled versions of the MDB.
  For the treatment of non-single stars CU4 receives the astrometrically pre-reduced individual Gaia measurements for all stars in the form defined in GAIA-LL-061-2, along with the corresponding trial single-star astrometric solutions produced in the astrometric core processing. For the detection and treatment of resolved and partially resolved non-single stars, the ‘shape parameters’ and flags from the IDT centroiding process will be used by CU4 (but the main work there will be based on the 2-d imaging results from CU5).

For the astrometric reduction of Gaia observation on solar system bodies, original single-CCD transit times and field coordinates will be transferred from CU3 to CU4, as discussed in GAIA-LL-061-2.
CU4 returns duplicity flags to CU3, which are used for the selection of AGIS primary stars.

- **CU5:** The data flow from CU3 to CU5 and vice versa is organized through the MDB ICD and the controlled versions of the MDB. For the main photometric processing, CU5 receives cross-matched intermediate data from the SM, AF and RP/BP instruments. In return CU5 delivers calibrated magnitudes and colours for all objects. They are used for the determination and correction of colour and magnitude terms in the astrometric reduction, and also to improve the centroiding and flux estimation (in the IDU process).
  For the 2-d imaging process, CU5 receives special astrometric parameters for the superposition of pixel data. They are defined in GAIA-LL-061-2.
  The third part of the interface between CU3 and CU5 consists of the CCD and PSF/LSF calibrations that CU5 produces, and which are used by CU3 in the IDU centroiding processes.

8.5 CU4: Object processing

CU4 (Object Processing) will further process any ill-behaved objects which pop up in CU3, CU5, or CU6 default reduction as well as those identified as eclipsing binaries by CU7. Such objects include Non-Single Stars (NSS), Solar System Objects (SSO), and Extended Objects (EO).

Even though the identification of all these objects will result from the original processing by CU3, 5, 6, and 7, those CUs will not provide the list of objects to be processed by CU4. Instead, the former will provide a quality indicator of the default model fit for each object and CU4 will set a threshold above which an object deserves a quest for a revised model. That threshold will be set to maximise the scientific return of the mission achievable within the resource constraints (computing resources, schedules for data and output releases, . . .).

The Data Processing Centre associated with CU4 is the CNES. The reduction frequency will be quite different depending on whether one deals with NSS or SSO. In the one hand, SSO require daily data due to their rapid sky motion and also because observations of some particular objects might be immediately very useful for special cases, like Earth-crossing asteroids on peculiar trajectories. In the other hand, NSS could essentially afford to wait till the end of the mission to be processed. So, if one ever needs some unexpected computing resources, the reduction of NSS could be postponed without affecting the rest of the mission. This is also true for EO.

If one omits the Earth-crossing asteroids, the three groups of objects could wait until the mission is over to be processed. However, owing to the iterative nature of that processing, updating the solution as new data come in might be a way to speed up the fitting stage with respect to an approach where all the data would be ingested at once. Besides, it may well happen in the first phases of the data reduction that objects are incorrectly classified as ”ill-behaved” by CU3, 5, 6, or 7, simply due to undetected calibration problems. It may thus be of interest that some early CU4 processing is performed to help the concerned CU to recognize and correct such cases ASAP.

8.5.1 Structure

The structure of CU4 closely follows the main division of its components, i.e. non-single stars (essentially built up on the former Double and Multiple Star working group) and the Solar System Objects (as a continuation of the Solar System working group). CU4 is therefore subdivided into several DUs, two of which are responsible for a group of SSO DUs and NSS DUs respectively (boldface in Tab. 11). The foreseen activity around EO does not so far justify any further division of that unique DU.

The *Simulation/Test Data Management* DU provides the interface with CU2 whereas the two *Simulations* DU under NSS and SSO are in charge of simulating object features which need to be available for internal debugging purpose. It is unclear, for the time being, whether the level of details and the number of test cases make it possible to presently rely upon CU2 tools only for the first development cycle. Once the simulator (Gibis/GOB) is fully operational and contains detailed object simulations directly accessible, those two DU will become obsolete. Then the role of these DU may merely be the production of test data for the algorithm validation beside the role of natural interfaces between CU4 and CU2.

CNES will play a key role being responsible for all the common work packages related to the infrastructure Tab. 12.

The Unit is chaired by a manager (presently Dimitri Pourbaix) and a deputy manager (currently Paolo Tanga) assisted by a Steering Committee composed of the manager, his/her deputy, a representative from CNES (for the time being, Thierry Levoir) and three scientists (Frederic Arenou, Albert Cellino, and Christine Ducourant at the time of writing).

In terms of workforce, CU4 is built upon 56 persons, essentially scientists spread in 16 cities all around Europe (Athens, Bergamo, Besançon, Bordeaux, Brussels, Dresden, Geneva, Heidelberg, Helsinki, Liège, Lubljana, Nice, Paris, Torino, Toulouse, Uppsala) and two outside Europe (Gainesville and Moscow). Some of these cities hosts two or more distinct institutions. Although there are a few exceptions, namely Brussels, Paris and Torino, each location focuses on either EO, NSS or SSO exclusively.

Thirty percents of these sites contribute to the manpower with only one individual, almost all with tenure positions. Those singletons are therefore likely to grow thanks to PhD students or postdocs joining the teams. Still, strictly speaking, these singletons constitute a risk since some of these individuals could be assigned to other duties by

Table 11: DU structure of CU4

Responsible	Title
D. Pourbaix	Management and scientific coordination of CU4
D. Pourbaix	**Management and implementation of NSS processing**
J.-L. Halbwachs	Astrometric Binaries
D. Pourbaix	Resolved Multiples
E. Gosset	Spectroscopic Binaries
O. Malkov	Photometric Analysis
C. Siopis	Eclipsing systems
A. Sozzetti	Extrasolar Planets
F. Arenou	Simulated Test Data
P. Tanga	**Management and implementation of SSO processing**
J. Berthier	Auxiliary data
F. Mignard	Solar System objects cross matching
A. Dell'Oro	CCD processing
J.-E. Arlot	Astrometric reduction
J.-M. Petit	Object threading
K. Muinonen	Orbital inversion
D. Hestroffer	Global Effect on Dynamics
A. Cellino	Physical parameters
W. Thuillot	Ground based observations
F. Mignard	Simulated Test Data
C. Ducourant	**Management and implementation of EO processing**

their home institutions. However, there is enough know-how redundancy in the whole CU to prevent any shortage of resources in case of problems with any of these singletons.

8.5.2 Milestones and schedule

CU4 will follow a cyclical development up to launch as outlined in Sect. 7.5.2. The complete details of each cycle will be refined in the requirements gathering phase at the beginning of the cycle but some milestone highlights are mentioned here.

8.5.2.1 Cycle 2 At least one prototype D.U. of each object type will be put in place, including the different work packages and the task scheduler.

Table 12: Essential work packages of CU4 under CNES responsibility

Title
Architecture and technical coordination of CU4
Define System architecture
System Administration
Quality Assurance and configuration management for CU4
Configuration Management
Integration
Subsystem Integration/Validation at DPC
Optimise System
Subsystem operations and monitoring
Maintain Subsystem
CU4 host software framework

8.5.2.2 Cycle 4 At the end of cycle 3 (November 2007), all codes relying on observations from one instrument exclusively should be running smoothly with respect to the existing simulations. With cycle 4 should begin the combination of data of different types.

8.5.2.3 Cycle 10 Complete system working with as near to real simulation data as is available.

The top level workpackages are summarised in Appendix. B while detailed descriptions may be found in Appendix. C.5.

8.5.3 Interface with other CUs

The interfaces are with:

- CU1: This interface is mainly technical. The contact will therefore be between CU1 and the DPC (CNES) essentially.
- CU2: Members of CU4 are responsible for supplying the Besançon Galaxy generator with models for non-single stars (including extrasolar planets), solar system and extended objects.
- CU3: CU4 will receive data from CU3 either on a daily basis (for SSO) or after convergence of the astrometric GIS and processing of every star in that reference frame (NSS). For NSS, fully calibrated residuals (together with their

weights and the single star model and its derivatives) will be made available [Lin05b]. Once CU4 successfully model an object as binary, it can be definitively removed from AGIS. Raw sample data are also needed.

- CU5: Besides epoch fluxes/magnitudes, some of the CU4 DUs (Resolved Doubles, Extended Objects) will also use the products of CU5 DU18 '2D image restoration' led by A. Brown.
- CU6: CU4 expects nothing but barycentric epoch radial velocities and uncertainties from CU6. However, if the derivation of those RV involves cross-correlation, epoch spectra might be preferred instead, thus making it possible to use some disentangling methods, such as Todcor or Korel. In such cases, the radial velocities and the orbital parameters are derived all together.
- CU7: The contact will be mostly on Eclipsing binaries and Planetary transits. CU4 will also take advantage of the period search implementation provided by that CU.
- CU8: Some of the processing in CU4 will take advantage of some astrophysical parameters (surface gravity, effective temperature, . . .) derived by CU8, even if they are initial guess obtained assuming a single star model. In the case of eclipsing binaries, those values will be updated by the corresponding DU from CU4. The multiplicity flag will always be confirmed or updated (single source changed into double or binary star changed into a triple), thus making CU8 aware that its own results might not be for a single star.

As mentioned in the introduction, CU3, 5, 6, and 7 will have to provide a quality indicator (goodness of fit, . . .) of their initial solution from which CU4 will decide whether an alternative model is worth fitting or not.

8.6 CU5: Photometric processing

CU5 is responsible for the photometric processing.

8.6.1 Structure

The structure of CU5 is mapped into Development Units, as discussed in the next paragraphs, with clear definition of responsibilities and interfaces.

DU01 Planning, management, and coordination of CU5 activities
Manager: F. van Leeuwen (Institute of Astronomy, Cambridge) This DU takes care of planning, management, and coordination of CU5 activities: the resource management, communications (reporting) and representation of CU5 at various levels.

DU02 Architecture and technical coordination of CU5

Manager: P. Bunclark (Institute of Astronomy, Cambridge) This DU takes care of the architecture and technical coordination of CU5, in particular System Architecture and System Administration.

DU03 Quality assurance and configuration management for CU5

Manager: P. Richards (Rutherford Appleton Laboratory) This DU takes care of the quality assurance and configuration management for CU5, and specifically Software quality assurance, configuration management and scientific quality assurance.

DU04 Integration, validation and operation of CU5

Manager: F. De Angeli (Institute of Astronomy, Cambridge) This DU takes care of the integration, validation and operation of CU5, and in particular the simulation and test data management, the system integration and validation at the DPC, the system optimization, the system operation and monitoring and the system maintenance.

DU05 Technical support

Manager: F. De Angeli (Institute of Astronomy, Cambridge) This DU takes care of the technical support within CU5, providing and maintaining development tools, developing software tools for common use, and providing general software support to CU5 developers.

DU10 PSF and LSF calibration

Manager: M. Barstow (University of Leicester)

This DU is in charge of the design, development and testing of the software for the PSF and LSF calibrations. The calibration software will include both the reconstruction and the application of the PSF and LSF for the SM, AF, BP/RP CCDs, and will process the raw transit data. It will be able to operate with or without the knowledge of transit positions and intensities as obtained from the astrometric and photometric GIS. An interface with laboratory data will provide supplementary information.

DU11 BP/RP flux extraction and initial data treatment

Manager: A. Brown (University of Leiden)

This DU is in charge of the design, development and testing of the BP/RP spectral extraction package for use in the initial data treatment. This implies defining broadband flux and colour parameters from the dispersed photometry for the astrometric GIS and exploring the possibility of using different parameter sets for saturated images. Provisional background extraction over the full range of brightness will also be part of this package.

DU12 Research on SM, AF, BP/RP photometric calibration model

Manager: C. Jordi (University of Barcelona)

This DU is in charge of the research into the internal photometric calibration model. It includes exploring provisions for the wide range of large- and small-scale influences on the observed fluxes, defining methods for comparing and combining different dispersion spectra, and accommodating in the calibration models the effects of CTI. The implementation of the models in the CU5 pipeline is the responsibility of DU15.

DU13 Instrument absolute response characterisation: ground-based preparation

Manager: E. Pancino (Osservatorio Astronomico di Bologna)

This DU is in charge of providing the ground-based observations required for the absolute calibration of the Gaia photometric data. This includes the assessment of the needed ground-based observations, the acquisition and reduction of the observations and the required preparation of the data for the application to the Gaia photometry.

DU14 Instrument absolute response characterisation: definition and application

Manager: C. Cacciari (Osservatorio Astronomico di Bologna)

This DU is in charge of the design, development and testing of the software for the absolute calibration of the photometric data. It includes an assessment of the expected accuracies. A function for the application of the absolute calibration to all observations is also part of these developments.

DU15 Internal photometric calibration and its application

Manager: D. W. Evans (Institute of Astronomy, Cambridge)

This unit is responsible for implementing the internal calibration model developed in DU12 yielding calibration parameters, and for the the application of the internal and external calibration models to all sources. It is furthermore in charge of the design, development and testing of the methods and software for the accumulation of the mean flux information and the variability detection. It is also responsible for the software and methods required for the release of the CU5 data to the central database at ESAC.

DU16 Selection and preparation of internal calibration sources

Manager: C. Jordi (Universitat de Barcelona)

This unit is responsible for the selection of reference sources for internal calibration of fluxes (G and BP/RP), wavelength scale and absolute zero wavelength (BP/RP). This includes establishing suitable criteria using all available information of the sources and their observations and the design, development and test of the methods and software for the selection.

DU17 Flux and classification-based science alerts
Manager: V. Belokurov (Institute of Astronomy, Cambridge)

This DU is in charge of the design, development and testing of the methods for flux and classification-based anomaly detection. This includes the rapid detection of flux anomalies in the initial data treatment as well as classification-based anomalies to be done during normal operation activities. Methods to filter the anomaly candidates will also need to be explored and developed together with the assessment of the statistical success rate of the detection process and thus the reliability of the alerts.

DU18 2D image restoration
Manager: A. Brown (University of Leiden)

This DU is in charge of the design, development and testing of the software required for reconstructing images from SM (1 and 2) and AF (5, possibly also 2 and 8) transits. 2D maps will have to be produced from stacked images and analysed automatically to characterise disturbing sources.

DU19 Data archive and database
Manager: N. Hambly (Institute for Astronomy, Edinburgh)

This DU is in charge of the design, development and testing of the photometric database, as based on a specification of user requirements. The main requirements will be from the process in charge of the preparation of the data release and detection of variability, DU16. This unit is also responsible for the digestion of general data-base upgrades as received from the central data base at ESAC.

8.6.2 Milestones and schedule

CU5 has scheduled 6-monthly internal reviews, to be held at alternating institutes participating in CU5. At these reviews the managers and developers will come together to present, discuss and plan their progress. Each internal review will be concluded with a written report, presenting an overview of the current status and the planning to completion.

CU5 also participates in the development cycles for DPAC. At the conclusion of each development cycle there will be an assessment, by the management team, of the progress, man power distribution over the tasks, planning, risks etc. These meetings may result in adjustments of requirements in exceptional cases.

In the current planning, the internal reviews will take place halfway each cycle, providing effectively 3-monthly reference points for the management team to monitor and direct the progress.

The top level workpackages are summarised in Appendix. B while detailed descriptions may be found in Appendix. C.6.

8.6.3 Interface with other CUs

CU5 will have direct interfaces with CU2 (simulations) and CU1 (central data base). Indirect (requirements) interfaces will exist with all other CUs, as all will need processed photometric data or photometric calibration models.

8.7 CU6: Spectroscopic processing

The CU6 is responsible for the following aspects of the spectroscopic processing:

- Calibrate the characteristics of the Radial Velocity Spectrometer: e.g. spectral dispersion law, overall throughput, etc.
- Conduct a ground-based observation campaign to define a library of radial velocity standards. These standards will be used to calibrate the RVS wavelength scale and to define the zero point of the radial velocities.
- Design, develop and operate the programs that will monitor the good health of the RVS during the Gaia operational phase. The analysis of the output of the good health monitoring program will be under the responsibility of CU3, in the context of its "First Look" activity (see Sect. 8.7.3).
- Extract, clean, calibrate the raw spectra collected by the RVS. These spectra will be used by CU6 to derive some of the characteristics of the sources (see below). They will also be delivered (through the Gaia main database) to other CUs which are in charge of specific facets of the spectroscopic processing (see Sect. 8.7.3).
- Derive epochs radial velocities and mean radial velocities for single and multiple sources.
- Derive epochs rotational velocities and mean rotational velocities for single and multiple sources.
- Diagnose potential binary/multiple sources.
- Diagnose potential variable sources.

- Identify objects that may require a “rapid” ground-based follow-up and issue alerts.

Other facets of the spectroscopic processing are under the responsibility of other CUs: e.g. the derivation of the stellar atmospheric parameters (T_{eff}, $\log g$, [Fe/H]) is under the responsibility of CU8. The spectroscopic tasks under the responsibility of other CUs are summarised in Sect. 8.7.3.

8.7.1 Structure

The CU6 is managed by a CU coordinator which also act as science coordinator. The CU coordinator is supported by a technical coordinator and a steering committee. The work breakdown structure (WBS) of CU6 is divided in Top-level work packages (TWP) and work packages (WP). Each TWP and WP is managed by a coordinator supported by development teams. The CU6 web site and documents are managed by a documentalist. Table 13 presents the CU6 organisation.

Table 13: CU6 organisation.

CU6 coordinator	Katz
Science coordinator	Katz
Technical coordinator	Jean-Antoine
Steering committee	Cropper, Jasniewicz, Jean-Antoine, Katz, Levoir, Viala
Documentalist	Turon
TWP coordinators	See text.
WP coordinators	See text.

The CU6 counts 45 members working in 15 different institutes (see Table 14). The CU6 data processing centre is the Toulouse site of the “Centre National d’Etudes Spatiales - CNES” (National Centre for Space Studies).

The CU6 work breakdown structure is made of 10 Top-level work packages which can be divided in 4 common Top-level work packages ((i) Management & scientific coordination of CU6, (ii) Architecture & technical coordination of CU6, (iii) Quality assurance & configuration management of CU6 and (iv) Integration, validation & operation of CU6 system) and 6 top-level work packages that are specific to CU6:

- GWP-S-610 CU6 host framework infrastructure.
- GWP-S-620 Spectra extraction.
- GWP-S-630 Calibration of spectroscopic instrument.
- GWP-S-640 Radial velocity zero-point.

Table 14: Institutes involved in CU6.

Institute	Country
INAF,Astronomical Observatory of Padova	Italy
Astrophysical institute of Potsdam	Germany
Group of research in A & A of Languedoc (GRAAL)	France
Institute of Astrophysics of Paris (IAP)	France
Mullard Space Science Laboratory (MSSL)	United Kingdom
National Centre for Space Studies (CNES)	France
Observatory of Bordeaux	France
Observatory of "Côte d'Azur" (OCA)	France
Observatory of Geneva	Switzerland
Observatory of Paris	France
Royal Observatory of Belgium (Brussels)	Belgium
University of Antwerp	Belgium
University of Bonn	Germany
University of Liege	Belgium
University of Ljubljana	Slovenia

- GWP-S-650 Single transit analysis.
- GWP-S-660 Multiple transits analysis.

The 6 CU6-specific Top-level work packages are briefly presented below.

GWP-S-610 CU6 host framework infrastructure.
TWP coordinator: A. Jean-Antoine (CNES); WP coordinators: A. Jean-Antoine (CNES). This TWP is devoted to the definition, implementation, test and operation of the infrastructure which will host the spectroscopic processing and analysis pipeline: e.g. database, data access layer, job scheduler, etc.

GWP-S-620 Spectra extraction.
TWP coordinator: M. Cropper (MSSL); WP coordinators: M. Cropper (MSSL), M. Steinmetz (Potsdam).
This TWP is devoted to the development of the software that will reconstruct the windows (when truncated), model the background and subtract it, deblend the overlapped spectra of close neighbours, clean the spectra (from e.g. cosmic rays), apply the calibrations and normalise the spectra to a pseudo-continuum.

GWP-S-630 Calibration of spectroscopic instrument.
TWP coordinator: M. Cropper (MSSL); WP coordinators: M. Cropper (MSSL), S. Mignot (Paris), M. Steinmetz (Potsdam).
This TWP is devoted to the development of the software that will calibrate the characteristics of the Radial Velocity Spectrometer: e.g. CCD characteristics, PSF profile, wavelength scale, ... This TWP also includes WP devoted to the sanity checks of raw and faint data.

GWP-S-640 Radial velocity zero-point.
TWP coordinator: G. Jasniewicz (GRAAL); WP coordinators: G. Jasniewicz (GRAAL), C. Soubiran (Bordeaux).
This TWP has two objectives: (i) build a library of radial velocity reference sources (stars and asteroids) that will contribute to the calibration of the RVS wavelength scale and to fix the zero-point of the radial velocities and (ii) develop the softwares to compute the corrections needed to transform the measured radial velocities into kinematical radial velocities (e.g. compute the gravitational redshift correction).

GWP-S-650 Single transit analysis.
TWP coordinator: Y. Viala (Paris); WP coordinators: R. Blomme (Brussels), C. Delle Luche (Paris), J.-M. Désert (IAP), Y. Frémat (Brussels), E Gosset (Liege), C. Martayan (Paris), F. Royer (Paris), Y. Viala (Paris)
This TWP is devoted to the development of the software that will: perform a coarse characterisation of the spectra, derive the single transit radial and rotational velocities and issue science alerts for object requiring a rapid ground-based follow-up.

GWP-S-660 Multiple transits analysis.
TWP coordinator: M. Cropper (MSSL); WP coordinators: M. Cropper (MSSL), P. Dubath (Geneva).
This TWP is devoted to the development of the software that will: combine the information obtained over several transits to derive the multiple transits radial and rotational velocities and to identify the variable sources.

8.7.2 Milestones and schedule

As for the other coordination units, the CU6 schedule will be structured around development cycles of 6 months. The cycle 2 is an exception in this scheme. It will last 7 months.

8.7.2.1 Overall objectives for the period 2006 - 2008

The overall objectives for the period 2006 - 2008, i.e. cycle 2 to cycle 5, are:

Cycle 2 [mid-Oct 06 - mid-May 07]:

- Validate the procedures and protocols defined in the "Gaia Assurance and Engineering Dispositions for Software Development" [LJMD$^+$07]: the definition of the requirements, the definition of the algorithms design, the delivery of the algorithms to CNES, the integration of the algorithms at CNES, etc.
- Specify the requirements and design the software products of the TWP host framework, extraction, calibration and single transit analysis and the WP astrophysical zero point and source variability.
- Implement, deliver to CNES and integrate at CNES the software products for which the specification/design will be "quick": i.e. 8 software products, see detailed objectives of cycle 2 in Sect. 8.7.2.2.

Cycle 3 [mid-May 07 - end-Nov 07]:

- Implement and deliver/integrate at CNES the software products belonging to the TWP host framework, extraction, calibration and single transit analysis and the WP astrophysical zero point and source variability.
- Perform an end-to-end test of the processing chain from extraction to single transit analysis.

Cycle 4 [Dec 07 - end-May 08]:

- Optimize and deliver/integrate at CNES the software products belonging to the TWP host framework, extraction, calibration and single transit analysis and the WP astrophysical zero point and source variability.
- Specify and design the software products of the TWP multiple transit analysis.

Cycle 5 [June 08 - end-Nov 08]:

- Optimize and deliver/integrate at CNES the software products belonging to the TWP host framework, extraction, calibration and single transit analysis and the WP astrophysical zero point and source variability.
- Implement and deliver/integrate at CNES the software products belonging to the TWP multiple transit analysis.
- Perform an end-to-end test of the full processing chain.

8.7.2.2 Detailed objectives and agenda of cycle 2

Some CU6 WPs aim to implement and optimise (to the Gaia-RVS specific case) "classical" astronomical methods. For those WPs, the first 1/3 of cycle 2 will be devoted to the specification of the requirements and to the design of the software products. The remaining 2/3 will be devoted to the implementation, delivery, integration and test of the software products. The same agenda will apply to the algorithms which are already in development for several months (before the start of cycle 2). The concerned WP/software products are listed in Table 15.

Table 15: List of WPs/software products to be specified, designed, implemented, delivered, integrated and tested during cycle 2

610-30000	Design, develop and test [host framework infrastructure]
620-08000	Normalisation to the continuum
630-10000	Wavelength scale and distortion map
650-07000	Radial & rot. vel. by CC with a template/mask in data space
650-08000	Radial velocity in Fourrier space
650-10000	Radial & rot. velocity by minimum distance method
650-11000	Rotational velocities by neural network
650-12000	Radial & rot. vel. for multi-lines by TODCOR-like method

A large fraction of the CU6 WPs are concerned with questions that are very specific to the RVS. The definition/choice of the "optimal" method to solve these problems require a significant amount of work/time. For those WPs, the full cycle 2 will be devoted to the specification of the requirements and to the definition of the design of the software products. The definition of the requirements and the design of the software products may require some prototyping, but these prototypes will not be delivered to CNES during cycle 2. The level of details of the definition of the design of the software product reached during cycle 2 should allow for a quick start of the implementation of the software products during cycle 3. The WP/software products to be "only" specified and designed during cycle 2 are listed in Table 16.

Table 16: List of TWPs and WPs/software products to be specified and designed during cycle 2.

620-all	Spectra extraction
630-all	Calibration of the spectroscopic instrument
640-05000	Astrophysical zero point
650-06000	Coarse characterisation of sources
650-14000	Detailed first look and validation: single transit
650-15000	Science alerts
660-05000	Assess source spectroscopic stability/variability

The CU6 main milestones for cycle 2 are:

- Conclusion of cycle 1 - Kick-off of cycle 2
  - 12-13 Oct. 06: CU6 Workshop 2 Brussels.
- Specification, design, implementation, delivery, integration and test of the 8 software products defined in Table 15.
  - 3 Nov. 06: Delivery of the Software Requirements Document (SRD).
  - 1 Dec. 06: Delivery of the Software Design Document (SDD).
  - 30 March 07: Delivery of the software products, of the Performance Report Document (PRD) and of the Software User Manual (SUM).
  - 11 May 07: completion of the software integration, validation and test.
- Specify and design the software products of the TWP and WP listed in Table 16.
  - 10 Nov. 06: Delivery of the draft Software Requirements Document (SRD).
  - 15 Dec. 06: Validation of the draft software requirements.
  - 15 Dec. 06: Delivery of the draft Software Design Document (SDD).
  - 26 Jan. 07: Validation of the draft software designs.
  - 16 March 07: Delivery of the revised Software Requirements Document (SRD).
  - 13 Apr. 07: Validation of the revised software requirements.
  - 13 Apr. 07: Delivery of the revised Software Design Document (SDD).
  - 11 May 07: Validation of the revised software designs.

8.7.3 Interfaces with other CUs

CU1

- CU1 defines the quality assurance (QA) rules for the whole DPAC. These rules apply to: e.g. the structure of the development cycles (phases and deliverables), the software to be used (e.g. subversion, eclipse, ...), the coding standard. As the other CUs, CU6 will implement the QA rules.
- CU1 defines the structure and operates the Gaia main database (MDB). The Gaia main database will store the raw data as well as the processed data from all CUs. During the operation phase, processed data (from all CUs) will be transmitted from the Gaia main database to the Spectroscopic Data Processing Centre (SDPC - CNES) on a half yearly basis. Processed spectroscopic data will be transferred from the SDPC to the Gaia main database on the same half yearly period. During the development phases, data will be exchanged between the Gaia main database and the SDPC for large scale tests.

CU2

- CU2 will provide simulated RVS-like data to CU6. These data will be used by CU6 to develop, test and assess the performance of the spectroscopic programs. Three categories of simulated data will be available: pixel level (GIBIS software), telemetry level (GASS software) and processed data level (GOG software). The implementation of the three software packages and the computation of the simulated data is under the responsibility of CU2.

- CU6 will provide CU2 with error models (i.e. mathematical formulation) to simulate the spectroscopic processes errors (e.g. wavelength calibration error) as well as catalogue data precisions (e.g. radial velocity errors).

CU3

- CU6 is responsible for the definition of the spectroscopic IDT and IDU. The responsibility of the implementation will be discussed and defined jointly with CU3, on a case by case basis, according to the skills required (e.g. the implementation of the derivation of the barycentric correction which requires astrometric knowledge is under the responsibility of CU3).

- CU6 is responsible for the definition and implementation of the spectroscopic FL programs. Some of the them (those checking the good-health of the raw data) will be installed and will run in the Science Operation Center. The rest of them (looking to the RVS calibration parameters and to the derived astrophysical quantities) will be installed and will run in the Spectroscopic Data Processing Center (CNES - Toulouse). All the first look diagnostics will be send to the CU3 first look center, who will centralize the astrometric, photometric and spectroscopic diagnostics, analyze them and react in case some anomaly is detected.

CU4

- CU6 is responsible for deriving radial velocities for single and multiple-lines systems. It will also identify and flags potential spectroscopic multiple systems.

- CU6 will provide CU4 (via the Gaia MDB) with radial velocities time series and flags of potential multiple systems.

- CU4 is responsible for the derivation of the orbital parameters of the multiple systems.

CU5

- CU5 will provide CU6, via the Gaia MDB, with the G_{RVS} magnitudes for the sources observed by the RVS. These magnitudes will, in particular, be used by CU6 to model and correct for the mutual contamination of neighbouring spectra.

CU7

- CU6 is responsible for deriving the radial and rotational velocities and for identifying and flagging potential variable sources.
- CU6 will provide CU7 (via the Gaia MDB) with flags, radial and rotational velocities time series and single epoch spectra time series for the potential variables.
- CU7 is responsible for the characterisation of the variable sources.

CU8

- CU6 will provide CU8, via the Gaia MDB, with cleaned, calibrated, normalised, rest-frame spectra both at single epoch and combined over several epochs of observation.
- CU8 will provide CU6, via the Gaia MDB, with the characteristics of the stars. They will be used in many spectroscopic processes: e.g. modelling of the mutual contamination of neighbouring spectra, selection of the appropriate template and mask for the derivation of radial and rotational velocities.
- CU8 is responsible for the on-ground observation of the spectra of reference stars (with the exception of the radial velocity reference stars - see below) and for the computation of synthetic spectra. These spectra will be used by CU6 as auxiliary data: e.g. as calibration stars or as template to derive the stars radial velocities.
- CU6 is responsible for the on-ground observation of radial velocity reference stars, which will be used to calibrate the spectrograph in wavelength (together with other, i.e. non-ground-based-standard, stars) and to define the zero point of the radial velocities. The reason why this task is under CU6 responsibility (and not CU8 as the other spectroscopic ground-based observations) is that it is very closely connected to the calibration and radial velocity derivation tasks which are also under CU6 responsibility.

8.8 CU7: Variability processing

The objective of CU7 is to characterise the photometric and spectral variability. The motivations for having the CU7 structure we present is based on the experience gained from the Hipparcos mission and other surveys and on the following facts:

- The importance to determine calibrators for astrometry, photometry and RVS.
- The importance to have stringent procedures for validation of the calibrations.
- The importance to have an added value to the Gaia mission in the intermediate releases and final catalogue that allows an efficient and timely analysis by the scientific community based on these catalogues, this especially in the perspective of other contemporary large scale surveys.
- The fact that the processing would be difficult to realise outside the DPAC due to the heaviness of the computational effort required to process one billion objects.

The different CU7 tasks are described in details in Sect. 5.2.5. The variability behaviour is first characterised, deriving statistical parameters, searching for periods and fitting simple models. The variable sources (stars/QSOs) can then be classified, and further analyses of particular source types can be carried out. The variability results will also serve as an investigation database to further validate the photometric and spectroscopic calibrations. A wrong calibration can indeed result in false variability that can be detected during the quality assessment of the variability catalogue.

8.8.1 Structure

The CU7 is managed by Laurent Eyer (Observatoire de Genève) which also acts as science coordinator. He is seconded by two deputies Dafydd Evans (Institute of Astronomy, Cambridge) and Pierre Dubath (Observatoire de Genève). The technical coordinator is Mathias Beck (Observatoire de Genève/ISDC). The data processing center associated with CU7 is at the Geneva Observatory.

The CU7 includes 45 active members[22] located in about 20 different institutes. These active members form the CU7 consortium, which meets 3 to 4 times a year.

The top-level work packages (WPs) for CU7 are listed in Tab. 17. They can be divided into 4 groups. The first (numbers 701 - 705) are the management and technical WPs common to all CUs. All other WPs concern the implementation of the actual algorithms developed to realize the functional tasks identified in Sect. 5.2.5. They can be divided in another three groups (710 - 712, 720 - 721, and 730 - 732). The first group includes the lower level tasks to identify, characterise, and classify variable sources. The systematic variable processing either of all sources, or of specific type of sources, is the concern of the second group, while the third one comprises tasks which do not fall into the other groups. More details of each WP are given in the appendix. These WPs are also decomposed further into several levels of sub-WPs

---

[22] There is also a list of about 60 affiliated members. These members are not expected to provide work package contributions, but rather act punctually as scientific consultants. They are not further mentioned in this document.

Table 17: Top-level work packages for CU7

WP number	Name	Manager
GWP-M-701	Management and Scientific Coordination	L.Eyer (Geneva)
GWP-T-702	Architecture and Technical Coordination	M.Beck (Geneva)
GWP-T-703	Quality Assurance	I.Lecoeur (Geneva)
GWP-M-704	Integration, Validation and Operation	M.Beck (Geneva)
GWP-M-705	Variability S/W Framework	M.Beck (Geneva)
GWP-M-710	Special Variability Detection & Analysis	A. Lanzafame (Catania)
GWP-M-711	Variability Characterisation	J.Cuypers (Brussels)
GWP-M-712	Classification	C.Aerts (Leuven)
GWP-M-720	Specific Object Studies	N.Mowlavi (Geneva)
GWP-M-721	Global Variability Studies	L.Sarro (Madrid)
GWP-M-730	Unexpected Feature Analysis	P.Dubath (Geneva)
GWP-M-731	Analysis of impacts on Astrometry	A.Jorissen (Brussels)
GWP-M-732	Supplementary Observations	G.Clementini (Bologna)

in CU7 organisation and documentations, but these ones are not described in this document.

8.8.2 Milestones and schedule

The pre-launch phase focuses on the development of algorithms and their implementation in software. The first point must be emphasized, because this is not simply a matter of re-writing existing algorithms in Java. To maximally exploit the Gaia data, schemes must be developed which are tuned to the Gaia requirements and available data. Numerous different existing algorithms must be tested (to avoid “reinventing the wheel”), and modified or extended where necessary. The very large number of expected Gaia sources imposes very stringent constraints on algorithm performances in terms of processing time.

CU7 will develop and test software in accordance with the DPAC development cycles (see Fig. 54). Each cycle involves the development or improvement of algorithms, their implementations into Java and the deliveries to the DPC. The early cycles will focus on setting up a baseline system with simplified or even dummy algorithms. The objective of later cycles is then to add functionalities and improve performances.

Upon delivery of the different software packages for each scientific WP, the DPC implements the algorithm and performs system integration tests.

Details of the first CU7 cycles are listed below (the dates correspond to the end of the cycle). For cycles 5–9 only an outline of the tasks are given. Many things are improved in all algorithms at all cycles, the exact details of which are not entered into here.

Cycle 2 (mid May 2007) The main objectives of cycle 2 is to develop a first Vari system, and to gather a version of all the Software Requirement Specifications (SRSs) and Software Design Documents (SDDs). The Vari system includes the complete infrastructure required to implement specific algorithms, as well as a first data model including a few basic, important elements. A first version of the system is released internally by the end of February. Ample time is planned for discussions and reviews, and a second version will be released at the end of the cycle to support software development in cycle 3.

We request a SRS and a SDD for each work package even if they are very incomplete. The idea is to have specifications of simplified (sometimes perhaps almost dummy) algorithms in order to allow to start coding in cycle 3. In this way, a complete work cycle spread over cycle 2 and 3 can already be exercised for all WPs.

Cycle 2 does not include the complete set of activities. The pace is slower, first, to take into account the learning process (most people involved have to get familiar with the tools used such as Java, Subversion, and Eclipse for example), and second, because of the large amount of work involved in developing the first version of the system.

Cycle 3 (end November 2007) Cycle 3 is the second “learning” cycle. For each work package, the simplified/incomplete schemes specified in cycle 2 will be coded, tested and delivered to the DPC. A first integration will be carried out at the DPC at the end of the cycle. At the same time the documentations (SRSs and SDDs) will be improved and completed to serve as a basis for cycle 4 development. Feed backs and new requirements on the CU7 framework will be collected and a new version developed and issued.

Cycle 4 (end May 2008) An important goal of cycle 4 is to exercise for the first time a cycle with a complete set of activities. Codes corresponding to the SRSs/SDDs of cycle 2 will be developed, tested and delivered. They will be integrated at the DPC and system test will be performed on a complete set of simulated data. In this way, test data of different processing levels will be produced and provided to developers to support algorithm implementation in cycle 5.

As in previous cycle, updated SRS and SDD versions, and a new Vari system will be produced to serve as a basis for cycle 5. The same set of WPs now delivers algorithms which show greatly improved functionality, i.e. providing scientifically good estimates of their outputs. They may still be short of the final algorithm in the sense that the scientific or computational performance may not be optimal, and some of their functionality may be lacking or they may not be entirely robust (e.g. they may assume complete data). Reassessment of overall design. Critical design review follows.

The sequence of activities of cycle 4 serves as a model for the following cycles. In the following, we do not repeat the sequence description, but we just list some preeminent features of the cycles.

Cycle 5 (end November 2008) Individual algorithms improved in terms of quality, performance and functionality.

Cycle 6 (end May 2009) Software optimizations (e.g. parallelization, improved multidimensional optimization for algorithm training, improved neighbour searching) implemented where necessary. Final SRSs/SDDs for WPs 710 and 711.

Cycle 7 (end November 2009) Final code for WPs 710 and 711, and final SRSs/SDDs for WPs 712. Test data produced with integration of WPs 710 and 711 code.

Cycle 8 (end May 2010) Extension of algorithms to relevant optimally exploit heterogeneous data. Improvement in error estimate methods (e.g. covariances) for all algorithms. Final code for WP 712 and final SRSs/SDDs for all WPs. Test data produced with integration of WPs 712 code. Qualification review follows.

Cycle 9 (end November 2010) Final code for all WPs. Internal large scale testing. Identification of major remaining issues. Priority plan and schedule to correct these plus make additional improvements to performance, code quality (for maintenance) etc.

Cycle 10 (end May 2011) Final large scale testing, followed by final performance estimates and documentation. Acceptance review follows.

Figure 60 shows a CU7 planning showing the main tasks from cycle 2 to cycle 4. The dependencies and the data test flows are not depicted to avoid overloading the picture. Dependencies are described in the above cycle presentation. Test data will be provided for each cycle by the CU2 group. These test data will be provided to all WP developers. They will also be systematically processed during the integration test at the DPC to produce test data of higher levels at the end of all cycles (except for cycle 2) . The sequence of events in cycle 4 will serve as a model for later cycle.

8.8.3 Interface with other CUs

The interfaces are with:

- CU1 (MainDatabase): There will be a close contact with the CU1, and a thorough study of the CU1 proposed solution for the whole DPAC consortium will be made.

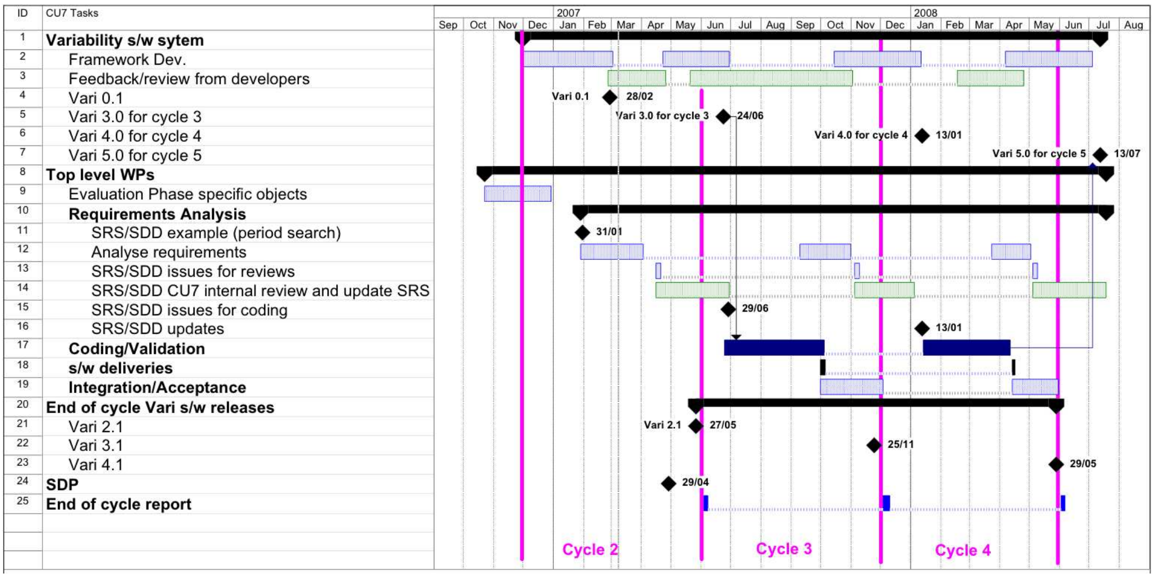


Figure 60: CU7 Software development schedule for cycles 2 to 4 showing the main tasks. See the text for a full description of the tasks and their dependencies. Cycle 4 will serve as a model for schedule of later cycles.

- CU2 (Simulations): CU7 will have the task to provide methods and data to put variable objects in the model of the Universe and to ask for data which will permit us to perform the tests of the software developed within CU7.

- CU4 (Objects): the contact will be mostly on eclipsing binaries and planetary transits.

- CU5/6 (Variability detection): Pierre Dubath and Dafydd Evans will be our connection to the core processing Coordination Units. These groups are in charge of the general statistics for assessing the level of variation in flux.

- CU8 (Astroph. parameters): Connections with CU8 are important since they deliver astrophysical parameters for the stars, which may help our own variability classification. On the other hand, the fact that an object is variable may lift some possible degeneracies. As an example the variability information for the identification of QSOs is important.

- Variability announcement: in the operation tasks, there will be a number of announcements to a wider community which will be made, this in agreement with the GST. These announcements are of interest for general public outreach, amateur astronomers, and scientists.

8.9 CU8: Astrophysical parameters

The objective of CU8 is to classify and determine astrophysical parameters for all of the sources which Gaia observes (see [BJ02] [BJ03] [BJ05] for more details). The context for this and the exact objectives were described in section 5.5, but are summarized here:

- Assign probabilities that a source belongs to each class in a list (e.g. single star, physical binary, optical double, galaxy, quasar, asteroid). From this, a sample of quasars with low non-quasar contamination can be constructed to define the extragalactic astrometric reference frame.

- Estimate astrophysical parameters (APs) for the sources. For stars these include effective temperature, T_{eff}, surface gravity, $\log g$, overall metallicity, [Fe/H], abundance of alpha-process elements, $[\alpha/\mathrm{Fe}]$, and line-of-sight interstellar extinction, A_G. These APs – in particular T_{eff} or perhaps an empirical classification – are also used by CU6 to aid in the RVS data processing.

- Accommodate the identification of new types of objects (such as rare stars, abnormal abundance patterns or rare multiple systems) which are inevitably discovered in any large scale survey.

The classification and AP estimation algorithms make use of all data which Gaia provides, namely RP/BP, RVS, proper motions and parallaxes. A description of the overall classification system and example algorithms were given in Sect. 5.5. CU8 is responsible for the end-to-end data processing to fulfill these tasks: requirements statement; task definition; algorithm development; software implementation; testing; execution; analysis; publication in the Gaia-DPAC catalogue.

The boundaries of CU8 with respect to the rest of the DPAC are best illustrated when the tasks which are *not* part of CU8 are considered:

- Use of morphological information. Nominally, all Gaia sources are point sources and are treated independently by the classification (CU8) processing. At the end of the mission, however, some 2D spatial information can be extracted from the multiple scans. These may be used for identifying planetary nebulae, galaxy halos, stellar disks etc. Exploitation of this is currently a task within CU4, although this could change if it is found that these data improve the primary classification work in CU8. This will be decided following more detailed simulations and, ultimately, mission data.

- *Detailed* classification of the many different types of (unresolved) multiple stellar systems. Preparatory work has shown that unresolved binary systems can be detected from their composite low resolution spectra or multiband photometry

[WKBJ04], and a CU8 algorithm (DSC) will attempt to detect these. A subsequent CU8 algorithm (MSC) will attempt to assign APs for the components, taking into account astrometric information on binaries (from CU4) where appropriate. Nonetheless, this task will be relatively limited in scope. Exactly how well we can classify binaries and parameterize their components will be determined during the pre-launch development.

- Orbital characterization of solar system objects. This concerns astrometric core processing and so is done by CU4. Taxonomic classification of asteroids (and identification of slow-moving ones from their spectra) is part of CU8.
- Detection and characterization of spectral variability. This is done by CU5 and CU7.
- Improvement of stellar APs using the light curves. This is of lower priority so is not considered now. It may be done by CU7, by the community following release of the Gaia catalogue or by CU8 in the post-mission processing phase.
- Science alerts. Development of software for these is coordinated by CU5, although quite likely using expertise (and even algorithms) from CU8.

CU8 produces most of the classification and astrophysical parameter information which appears in the final Gaia catalogue. Experience with Hipparcos and SDSS (for example) demonstrates that almost every user of the Gaia catalogue will require this information. Without it the astrometric and photometric catalogue is of limited use. Leaving this part of the processing for the community to do only after the rest of the processing is complete would (unacceptably) delay the fruitful exploitation of the Gaia data. Furthermore, because the astrophysical parameter work involves a processing of the entire data (and because it feeds back into some critical parts of the main processing – see section 8.9.3) it must be done within the DPAC, subject to its schedule, standards and coordination.

It must be emphasised that all the CU8 tasks are concerned with extracting the astrophysical information on *individual* objects. They are not about doing science with the catalogue. For example, CU8 does not include tasks which deal with bulk analyses of objects, such as abundance analyses of stellar clusters, determination of the QSO luminosity function, or searches for specific types of stars. While some analysis will be undertaken by CU8 or CU9 for demonstration and calibration purposes (e.g. it would be foolish not to check whether we correctly derive the properties of well-studied stars or clusters), this is clearly driven by the need to test the algorithms and perform quality control on the data.

During the Gaia phase A study, some of the classification issues have been studied under the auspices of the working group “Identification, Classification and Astrophysical Parameterization” (ICAP). Details of the work done and copies of the technical reports produced can be found on the ICAP web page at `http://www.mpia.de/GAIA`.

8.9.1 Structure

8.9.1.1 Organization

In common with all of the DPAC CUs, CU8 comprises a scientific and a technical wing. The scientific activities are geographically distributed at about 20 institutes in 9 countries, with between 1 and 5 people working (not necessarily full time) for CU8 in each. The total membership of CU8 is 56 people; the average number of Full Time Equivalents is currently around 15 (as of July 2006).

The CU8 manager and scientific coordinator is Coryn Bailer-Jones (Max-Planck-Institut für Astronomie (MPIA), Heidelberg) and the deputy is Frédéric Thévenin (Observatoire de la Côte d'Azur, Nice). The data processing centre is the Centre National d'Etudes Spatiales (CNES) in Toulouse where the technical coordinator, Anne-Marie Janotto, is located.

The top-level WPs for CU8 are listed in Tab. 18. They can be divided into three groups. The first (numbers 801–805) are the management and technical WPs common to all CUs. The second group (811 and 812) are the "support" WPs which provide, respectively, the simulated data required to build and test the classification algorithms, and advice on the modelling and parameterization of interstellar extinction. The third group (821–836) comprise the actual DP algorithms which will be developed: there is one top-level WP for each of the major algorithms identified in Fig. 39 in Sect. 5.5. More details of each WP are given in the appendix. This third group represents the major part of CU8 in terms of manpower. The CU and top-level WP managers, along with this overall WBS, were agreed at the CU8 kickoff meeting held on 16–17 March 2006.

8.9.1.2 Algorithms

The major algorithms which will be developed for CU8 are described in Sect. 5.5. An estimation of the processing effort required for these is given in section 7.8.2.

8.9.1.3 Simulated data

As described in section 5.5, the classification algorithms require extensive and accurate sets of simulated data on which they are trained. The activities in CU8 (specifically, GWP-S-811) include the development of new stellar model atmospheres for this purpose, involving several leading groups in Europe. These form the basis for the construction of the training (and testing) data.

A number of stages are involved in this as illustrated in Fig. 61. First, synthetic spectra of the different classes of objects, viz. stars, galaxies, QSOs and solar system objects, are simulated. The stellar data are provided by GWP-S-811. Solar system objects, QSO and galaxy spectra are provided by their respective algorithm WPs

Table 18: Top-level work packages for CU8.

WP number	Name	Manager
GWP-M-801	Management and scientific coordination	CU manager
GWP-T-802	Architecture and technical coordination	Technical coordinator
GWP-T-803	Quality assurance	Technical coordinator
GWP-M-804	Integration, validation and operation	Technical coordinator
GWP-T-805	Host software framework	Technical coordinator
GWP-T-806	Data model and utility library	Tiede (Heidelberg)
GWP-S-811	Training data	Thévenin (Nice)
GWP-S-812	Interstellar extinction	Drimmel (Torino)
GWP-S-821	Discrete Source Classifier	Bailer-Jones (Heidelberg)
GWP-S-822	Generalized Stellar Parameterizer (phot.)	Bailer-Jones (Heidelberg)
GWP-S-823	Generalized Stellar Parameterizer (spectro.)	Recio-Blanco (Nice)
GWP-S-824	Object Clustering Analysis	Sarro (Spanish VO)
GWP-S-825	Luminosity, Age and Mass Estimation	Lebreton (Paris)
GWP-S-831	Quasar Classifier	Claeskens (Liège)
GWP-S-832	Unresolved galaxy classifier	Kontizas (Athens)
GWP-S-833	Solar system object classifier	Lagerkvist (Uppsala)
GWP-S-834	Multiple Star Classifier	Bailer-Jones (Heidelberg)
GWP-S-835	Extended Stellar Parameterizer	Fremat (Brussels)
GWP-S-836	Outlier analysis	Manteiga (Coruña)

(GWP-S-831, -832 and -833). Likewise, specific details of non-stellar spectra, such as emission lines, carbon enhancements etc. will be provided within GWP-S-835. Within GWP-S-811, the WP “Provide calibrations and auxiliary data” provides any real data required to calibrate the synthetic data (see section 5.5). All of these data – real and synthetic – are assembled by the WP “Prepare training data”. This WP performs any corrections of synthetic data using the real data, and applies artificial interstellar extinction according to the guidelines laid out by GWP-S-812. The result is that a grid of spectra on all types of object is obtained showing the required variance in the APs at the necessary spectral resolution and wavelength range. These are passed to CU2 (Simulations) which processes them with the Gaia instrument models to simulate mission data. CU2 then passes these back to the “Prepare training data” WP which assembles these into the training data sets (or grids) required by the various classification algorithms. This includes making any necessary transformations of the inputs (photometry, spectroscopy, astrometry) and outputs (APs, classes). This group maintains these libraries of training data throughout the project which are accessed by the classification algorithms as required.

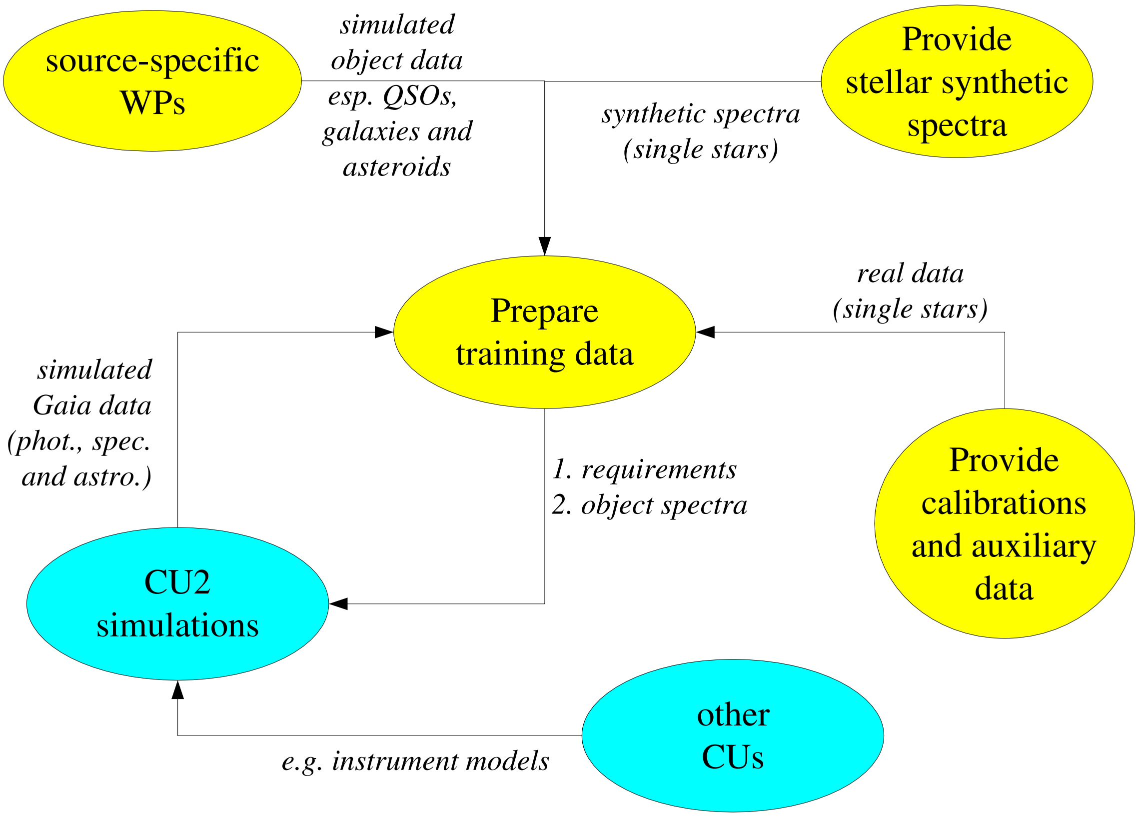


Figure 61: The different tasks and their connectivity in the generation of the simulated data sets used by CU8 for developing (training and testing) the classification algorithms. The yellow ellipses (the upper four) indicate WPs within CU8.

It should be stressed that CU2 is "only" responsible for performing the instrument simulations. CU8 is responsible for providing simulations of the source spectra for the entire DPAC and coordinates the requirements across the DPAC.

As was described in section 5.5, the synthetic data and the AP estimation algorithms will require calibration using real data. To some extent this can use existing data, but a dedicated ground-based programme will be necessary. This will comprise (1) acquiring high-resolution (echelle) optical spectra of targets to accurately determine their APs, and (2) multi-band photometric observations of the same stars with which the continuum of their corresponding synthetic model spectra can be corrected. Although only in the early stages of defining the requirements, it is expected that of order 1000 targets will have to be observed using 8m class telescopes (for the spectroscopy). CU8 is coordinating this with other CUs so that a single set of proposals can be made to observatories. The reduction and analysis will be coordinated by CU8 and the resulting spectroscopic and photometric database made publicly available.

8.9.2 Milestones and schedule

8.9.2.1 Pre-launch development

The pre-launch phase is concerned with the development of algorithms and their implementation in software. The first point must be emphasized, because this is not simply a matter of re-writing existing algorithms in Java. To maximally exploit the Gaia data, a classification system must be developed which is tuned to the Gaia requirements and available data. Numerous different existing machine learning algorithms must be tested (to avoid "reinventing the wheel"), and modified or extended where necessary.

CU8 will develop and test software in accordance with the DPAC development cycles (see Fig. 54). Each cycle involves the development of an algorithm, its implementation into Java and its delivery to the CU8 DPC. The early cycles will focus on the assessment and modification of existing algorithms (not necessarily within Java) and also the production of a baseline Java code. If necessary, the basic algorithm maybe entirely changed during the very early cycles if its overall design is found not to meet the specifications. Within a few cycles an algorithm will be converged upon. The objective of later cycles is then to add functionality (e.g. include parallaxes, give error estimates, extend the AP output space) or improve performance (e.g. more accurate regression, better optimization, faster neighbour searches).

Upon delivery of the different software packages from each scientific WP, the DPC implements the algorithm, performs system integration tests, runs the algorithms and delivers the results back to the provider. The providers then analyse the scientific results (and CNES the computational ones), the results of which feed into the specifications for the next development cycle.

Details of the first four CU8 cycles are listed below (the dates correspond to the end of the cycle). For cycles 5–9 only an outline of the tasks are given. Many things are improved in all algorithms at all cycles, the exact details of which are not entered into here. Fig. 62 shows the time plan for cycle 1 by way of example.

Cycle 1 (end January 2007) The objective of the first cycle is primarily to test and evaluate the code integration, testing, accessing and reporting procedures. It involves the delivery of just the algorithms Discrete Source Classifier (DSC; GWP-S-821) and Generalized Stellar Parameterizer (photometry) (GSP-phot; GWP-S-822). Both algorithms will take as inputs just end-of-mission RP/BP data plus any error (covariance) information associated with it. DSC will operate on six classes of source namely single stars, physical binaries, non-physical binaries, quasars, galaxies and solar system objects. It will produce relative probabilities of each class (plus the additional "unknown" class). GSP-phot will estimate $T_{\rm eff}$, $\log g$, [Fe/H] and A_V for those sources which DSC identifies as single stars.

Cycle 2 (mid May 2007) All of the algorithmic top-level WPs in CU8 (viz. GWP-S-

821 to GWP-S-826 inclusive) will deliver working code. This involves the delivery of 11 distinct algorithms and will be the first full test of the quality assurance procedures across most of CU8. All algorithms must be capable of working with end-of-mission RP/BP and RVS data (as appropriate). All algorithms must show some minimum functionality: they must provide the specified outputs given the input data, although not necessarily to a high degree of accuracy. DSC and GSP-phot will be improved over what was delivered at cycle 1.

Cycle 3 (end November 2007) Improvement of all algorithms from the top-level WPs, in particular to incorporate parallaxes and proper motions into the classifications (as appropriate). The following simulated data will be available for different classes of objects: single stars (RP/BP, RVS, astrometry); non-single stars (RP/BP, RVS, astrometry); quasars (RP/BP, astrometry); galaxies (RP/BP, astrometry); solar system objects (RP/BP); unknown objects (RP/BP, RVS, astrometry). Software will also show improved functionality over cycle 2, and could involve different algorithms or multiple algorithms (for comparison purposes) for a given WP. Tests of algorithm interoperability (internal to CU8).

Cycle 4 (end May 2008) The same set of WPs as in cycles 2 and 3 now delivers algorithms which show greatly improved functionality, i.e. providing scientifically good estimates of their outputs. They may still be short of the final algorithm in the sense that the scientific or computational performance may not be optimal, and some of their functionality may be lacking or they may not be entirely robust (e.g. they may assume complete data). Reassesment of overall design. Critical design review follows.

Cycle 5 (end November 2008) Connectivity of algorithms expanded to get them to work in a processing chain. Individual algorithms improved in terms of quality, performance and functionality.

Cycle 6 (end May 2009) Software optimizations (e.g. parallelization, improved multidimensional optimization for algorithm training, improved neighbour searching) implemented where necessary.

Cycle 7 (end November 2009) Incorporation of variability estimates. Extension of algorithms to operate on multi-epoch data (as opposed to single epoch or averaged end-of-mission data). Improve estimates of classificatin and AP accuracy performance for all types of objects.

Cycle 8 (end May 2010) Extension of algorithms to relevant optimally exploit heterogeneous data. Improvement in error estimate methods (e.g. covariances) on all APs for all algorithms. Qualification review follows.

Cycle 9 (end November 2010) Internal large scale testing. Indentification of major remaining issues. Priority plan and schedule to correct these plus make additional improvements to performance, code quality (for maintenance) etc.

Cycle 10 (end May 2011) Final large scale testing, followed by final performance estimates and documentation. Acceptance review follows.

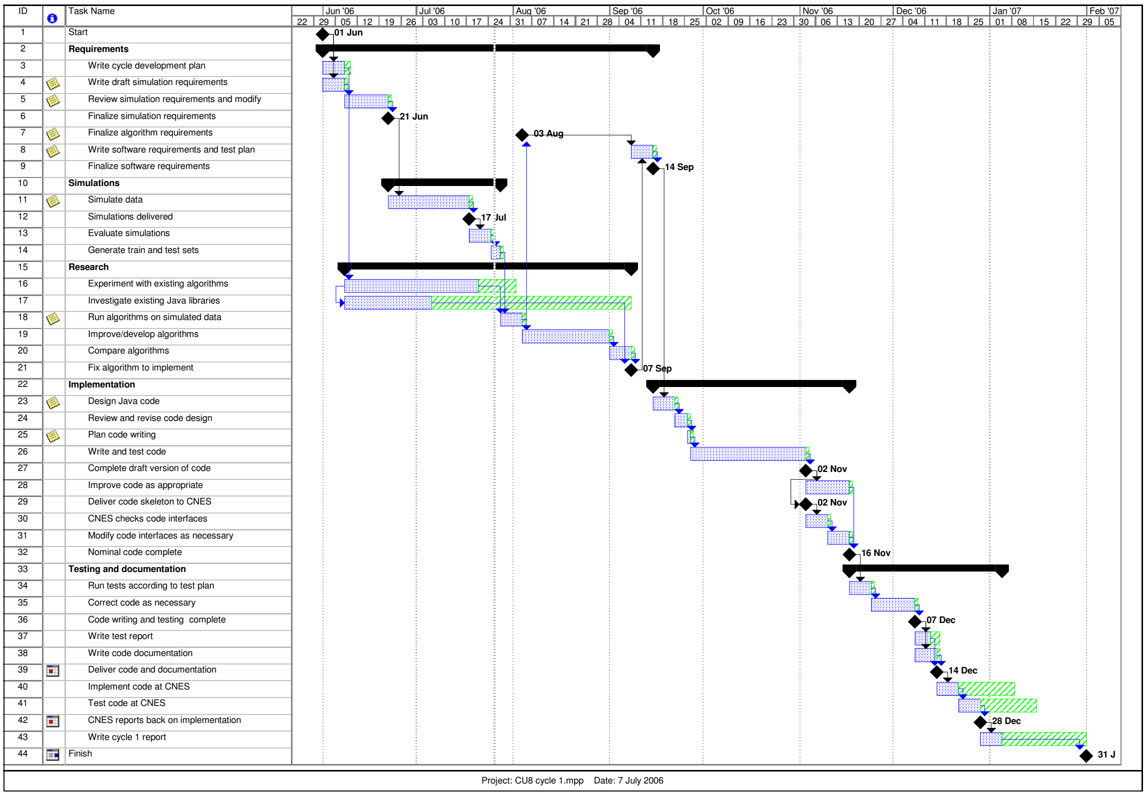


Figure 62: Software development cycle 1 in CU8 showing the main tasks and dependencies. In future cycles the durations of the tasks may change. The simulations requirements phase has been artificially compressed here: In practice the simulation requests and delivery take much longer (several months) and are therefore started much earlier (overlapping with the previous development cycle).

8.9.2.2 Mission operations and post-mission processing

Prior to the launch of Gaia, CU8 will have developed, tested and deployed a fully functioning data processing pipeline to fulfill the Gaia classification objectives. The mission phase is concerned with (1) the operation of the pipeline, (2) the analysis of the results, and (3) the improvement of the algorithms. Part (3) is particularly important. Gaia – and the classification work that will be done with it – is novel, and the simulations upon which we develop our system will inevitably not accurately nor completely reflect the panoply of celestial objects. During the mission, therefore, the algorithms will continue to be developed and improved, based on what is learned from the real mission data. (The exact schedule of this will be dictated to some degree by the early data releases.) This will include (but is not limited to):

- the need to recalibrate the classification algorithms using the Gaia observations of calibration stars (see Sect. 8.9.1.3)
- re-training algorithms using improved source models (e.g. stellar atmosphere codes, opacity tables etc.). This is necessary if the 2019 final catalogue is not to be based on 2010 input physics
- identification of new types of objects (from the Gaia data) which are then better characterized by external (non-Gaia) observations. Such knowledge can be fed back into the Gaia classification system to improve the identification of these objects and, where possible, provide more information on them (e.g. assign a specific class and/or APs)
- modification of the instrument simulations because the instrument characteristics or performance deviate from their ground-based assessment. (It is known that instruments will degrade as the mission proceeds due to radiation damage, but exactly how is impossible to predict.)

The DP system will be operated at CNES on each version of the data received from the main database. The results of the analysis will, as appropriate, be sent to a limited set of CU8 members (typically the algorithm developers) for assessment and analysis. This work includes

- inspection of the algorithm run-time logs to check for errors, convergence problems, unresolved degeneracies etc.
- sanity checks of the results via “manual” reprocessing of a subset of the data. This includes applying other methods/algorithms to this subset as well as visual and interactive analysis
- scientific validation of the results. This includes analysis of the results on “known” (i.e. well-studied) objects, to see whether the classes/APs are correctly reproduced. The work is carried out by all of the algorithmic top-level WPs in CU8, viz. GWP-S-821 to GWP-S-836 inclusive, and the effort for this is included in the manpower assessment for the operations phase.
- limited analyses of a known population of objects, such as globular or open clusters, to ensure consistency in AP estimates for different tyoes of objects. An example is known members of an appropriate globular or open cluster whereby our AP estimates reproduce the HR diagram for the cluster and indicate a common abundance, extinction etc. The results of this feed back into the calibration of the algorithms (e.g. [FBJL^{+}07]). Again, this is purely for scientific validation and not for scientific exploitation purposes.
- detailed analysis of all the “unknown” objects (GWP-S-836) and unclassifiable clusters emerging from the unsupervised analyses (GWP-S-824).

In some cases the results of these analyses will permit us to improve the classification algorithms and implement these in the next DP cycle.

The post-mission phase will see the final processing of all sources using all available data and the optimized algorithms. The main product is a contribution of classifications and APs to the final catalogue. But in addition the appropriate parts of the classification pipeline will be delivered to allow users to reclassify the data, using, for example, different training data sets or different parameter settings (e.g. thresholds on class probabilities, extinction curves). CU8 will also assist in the provision of visualization and data mining tools (boundaries to be agreed with CU9 once that CU is operational).

8.9.3 Interface with other CUs

The input data used by CU8 in its data processing are as follows, for all sources. [23]

- G-band measurement in magnitudes on a calibrated, physical scale (from CU5)
- Fully calibrated RP/BP data. Both single-epoch spectra and robust multi-epoch (e.g. end-of-mission) averages (from CU5)
- Fully calibrated RVS RP/BP data. Both single-epoch spectra and robust multi-epoch (e.g. end-of-mission) averages (from CU6)
- Parallaxes and proper motions (from CU3 or CU4, depending on whether the source is single or not)
- Summaries of photometric variability, perhaps as variability flags or indices (from CU7)

Initial uncertainty (error) estimates on the above are required by CU8. These can be improved as the development proceeds. By the end of the mission some estimate of the data covariances would be useful, e.g. from a data model.

The output data produced by CU8 and written to the main database are, for each source,

- probability that each source is a member of an astrophysical class (e.g. single star, physical binary, optical double, galaxy, quasar, asteroid, unknown). This is used by CU3 for selection of astrometric reference frame objects and may also be used by CU4 to aid binary star identification.

---

[23] As with all CUs, CU8 strictly takes all of its input data from the main database only. The CUs stated in the list indicate the origin of the data and thus with which CUs CU8 will need to coordinate specifications and requirements.

- for each class with a probability above some threshold, APs appropriate to that class derived from RP/BP (specifically, the GSP-phot algorithm)
- primary stellar APs ($T_{\rm eff}$, $\log g$, $[\alpha/{\rm Fe}]$ $A_{\rm G}$). These are used in particular by CU6 to select the template for the radial velocity determination (cross correlation).
- additional sets of APs for a given class, coming from GSP-spec, ESP (for stars) or class-specific parameterization algorithms. They may be used by CU7 to improve stellar AP estimates of variable stars.
- information on the outliers and natural clustering in the data (from an unsupervised analysis)
- computational and scientific analyses of the results to fulfill the Quality Assurance requirements (see section 8.9.2)

Uncertainty estimates (plus covariances where appropriate) will be provided for the APs.

As discussed earlier (Sect. 8.9.1.3), CU8 is responsible for producing simulations of the intrinsic spectra of all types of astrophysical sources as required by the whole DPAC. “Intrinsic spectra” means the spectra from the source prior to observation by Gaia. (CU2 then uses these to simulate mission data; see Fig. 61.) As several other CUs will also make use of these data (in particular CU6, but also CU4, 5 and 7), the requirements for the intrinsic spectra must be coordinated across all CUs. CU8 is responsible for this coordination. CU8 will likewise coordinate with other CUs concerning ground-based observations for the calibration of the AP estimation algorithms and synthetic data.

8.10 CU9: Catalogue access

The catalogue production for Gaia is understood to be covered by an AO to be issued at some later point in time. The DPAC is therefore concentrated on the data processing to produce the science products, not wishing to divert effort to considerations of how the catalogue will be presented. The DPAC will, in any case, have considerations concerning the new AO for the catalogue production and access. Therefore it has created as a placeholder a coordination unit (CU9) called ”Catalogue Access”, which will be activated at some future time.

8.10.1 Catalogue description

A great deal of effort in the catalogue production will be in the area of documentation. The catalogue itself is seen as an extract of the main database as depicted in Fig. 50. Given the volume and diversity of the Gaia science products, the real work will be making the accompanying documentation practically accessible to the astronomical community. Gaia has wide ranging applicability in many fields of astronomy, and precise and careful descriptions of all parameters and their derivation will be essential for the proper interpretation of the catalogue values. The virtual observatory community are working in the area of ontological descriptions of astronomy data, or as a minimum the creation of an agreed data model for astronomical data [McD04]. It would be premature to use this as a basis for our own data model from the beginning, but the final catalogue construction may certainly adopt any future standard. In any case, much of the effort behind the documentation will have to come from within the existing DPAC.

8.10.2 Catalogue Access

The access to the vast Gaia catalogue was already raised as an issue in the study report in 2000 [gai00]. The catalogue will be of the order of tens of terabytes: simply 'giving' the catalogue to an individual or institute will not be useful. Rendering the Gaia catalogue practically accessible is fundamental for making it scientifically useful; obviously the traditional printed catalogue is not to be considered. The organisation of the data will be very important and the way the data will be accessed will therefore be encoded in some software scheme. A present day example is the Sloan Digital Sky Survey(SDSS) catalogue, which is distributed as a database with special access software - the same system is available through the web[24].

Providing data mining facilities to the community is the most efficient way to provide access to a large catalogue. The SDSS catalogue is currently only a few terabytes - allowing for time similar problems will certainly have to be faced [GST$^+$03]. Some form of batch access, such as the CasJobs system [OLNS$^+$05], will be required. The logical integration of photometric, spectroscopic and astrometric data in a coherent system will certainly be an impressive task and a necessity for a Gaia catalogue.

In addition Gaia must deal convincingly with the time dimension and the notion of multiple observations. The limits of 3D technology will need to be pushed to properly visualize the Gaia data. The Gaia catalogue will adhere to the IVOA (International Virtual Observatory Alliance) standards, meaning that certain tools will immediately be available for use. Such a rich source of information demands an impressive interface - probably more impressive than what the VO will have to offer. We must remember the VO is in a way restricted by a least common denominator problem in that it wishes to bring together existing data from multiple astronomical sources. In contrast, the Gaia science products will be both diverse and interdependent, and the eventual interface between user and catalogue will have to be able to accurately reflect this rich detail.

It is clear the Gaia catalogue will of necessity be distinctly electronic in nature and

---

[24] `http://skyserver.sdss.org`

not distributed in the traditional manner. Rather we need to provide an interface to the community - a portal - to access the data. The presentation of this catalogue is itself a challenge requiring significant manpower and thought. It is understood that the hub (ESAC) will host at at least one copy of such a portal as well as serving any intermediate data.

9 The Data Processing Centres

This section introduces the main data processing centers (DPC) involved in the Gaia data processing effort, describing the resources and expertise found in each.

9.1 Data processing centres in general

All of the Gaia DPCs face a formidable task which may be seen from the following perspectives:

- Accumulation of sufficient processing power and its application to the processing.
- Creation of a reliable robust processing system which can be maintained until 2019.
- Execution and maintenance during operation of the processing system.

9.1.1 Accumulation and application of processing power

Significant resources are needed for Gaia processing, estimates continue to be refined but remain at around 10^{21} FLOPs for complete Gaia processing. However simply having a computer or set of machines capable of producing so many FLOPs in an acceptable amount of time is only one part of the problem. We also have a relatively large volume of data which must be operated upon. The volume is too large to consider holding it in all in any form of shared memory so it will need to be repeatedly read from disk. The choice of hardware architecture is therefore quite important in that a super fast machine with bad I/O will probably not do the work. Indeed early work such as [GST$^+$02] show us that I/O is a significant problem for Gaia processing.

Gaia data naturally lends itself to highly distributed processing which lends a model for building DPCs. The architecture being considered for managing this distributed processing would consist of a set of heterogeneous clusters linked by a high speed network and a storage infrastructure which meets the project's requirements.

The deployment of Astrophysical Parameter extraction on Gaia Grid, GIBIS on the CNES Cluster and AGIS on the ESAC cluster provides valuable insights into the potential constraints and limits of this approach.

In addition to dedicated facilities, other complementary possibilities will be studied to take into account peak workloads whose execution is constrained by the timing required for the Gaia scientific data processing. These may involve:

- Implementing a “public computing” solution for certain types of processing.

- Using the partners' computing facilities via a grid type software solution.

In addition to this distributed approach we need a high level of optimisation of the software at every level as well as enough redundancy to meet a 95% availability target.

9.1.2 Reliable, robust and maintainable processing system

The implementation of a reliable and robust system made up of the software components developed by laboratories and observatories is arduous. Furthermore the complex algorithms are often developed by scientists and not by computing professionals. It will be the role of the DPC engineer in charge of coordinating each CU to provide the best solutions for this critical point. He will play an essential role in choosing the software architecture. The DPC shall provide support and expertise for quality assurance and optimisation to the CU.

For DPCs to support their CUs they must develop a hardware and software infrastructure. The DPC will be responsible for:

- optimising processing sequences
- managing the local database
- organising storage
- maintenance and operation of communications with the Main Database (See Fig. 63)

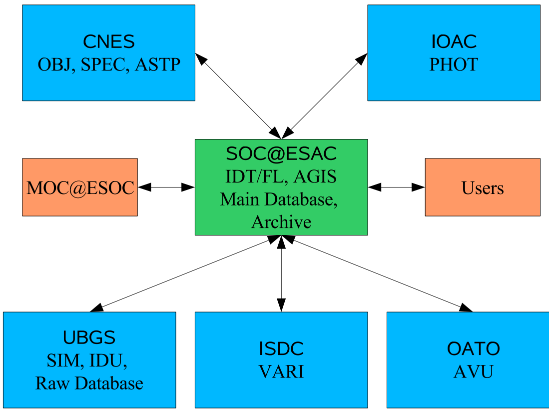


Figure 63: DPAC physical Architecture showing the ESAC hub with other DPCs and MOC on the spokes. The hub-and-spokes architecture minimizes interdependencies between DPCs.

Such an infrastructure is in itself a complex system which should be developed and validated in parallel with the sub-systems for scientific processing in the CUs. The complexity of this infrastructure merits early investigation.

CU1, in particular CNES, will assess the feasibility of developing common host infrastructure components. These components could then be customised and complemented for the needs of each DPC. The objective here is to reuse, whenever possible, common software components for the different CUs.

In order to avoid any duplication of effort at the DPAC level, CU1 will be coordinate the studies.

A key aspect of the processing approach consists in separating the algorithm software layer from the data organisation layer by means of a 'data access layer'. This concept has been implemented for Gaia by ESAC and will be enriched and extended for DPCs in general. It has two major advantages:

- Enabling the DPC to define the optimal data organisation without affecting the algorithms.
- Allowing development of software components by laboratories which may then be run in the laboratory environment or in the DPC environment, without code changes.

Availability of the system is part of this reliability - DPAC will strive for 99% availability to surpass the 95% availability required for operational systems.

9.1.3 Operation of the processing system

During the operations phase DPCs will have to maintain an operational system. Data will be transfered on a regular basis to and from the MDB system, so that the MDB processed data becomes available for further processings in the distributed DPCs. No direct exchange between satellite DPCs is foreseen in this structure since the MDB is a 2-way hub feeding all the processings and receiving their results. Additionally the DPCs will have to operate its CU software to perform the processing on the received data.

9.2 DPC-E: ESAC

The ESAC DPC is ESA's contribution to the DPAC, as specified by the SMP. The services and resources provided by this facility and its personel are available to any proposing consortium, and are presented here for completeness. A detailed accounting of these resources are given in the Appendix Sect. D.

9.2.1 Role of ESAC in the Gaia data processing

The European Space Astronomy Centre (ESAC) will host the Gaia Science Operations Centre (SOC) activities similar in nature to other ESA missions. In addition ESAC will be the DPC for CU1 hosting the Main Database (Sect. 7.3) and coordinating input to it from the other DPCs forming th hub of the processing system (as may be seen in Fig. 63). ESAC will provide a good deal of the effort for other CU1 activities also. Finally ESAC will be one of the core processing(CU3) DPCs running IDT and AGIS. ESA has seen these services as essential and wished to provide expertise and manpower in these complex areas. The services at ESAC are covered by ESA in the cost at completion of Gaia.

9.2.2 The ESAC team

The ESAC team has been building up since August 2005 with the notion of providing excellent technical support to the DPAC. The team has a strong background in scientific processing and key team members have been involved in Gaia since the white book[gai00] era. The ESAC team will support, and be involved in, the development of the systems which will ultimately run on hardware at ESAC. The team will be involved at all phases of the development and will provide guidance to the DPAC on technical issues.

9.2.2.1 Stakeholders and their roles at ESAC

The ESAC team is managed by the Science Operations Development Manager and the chair of CU1. The manager reports to the head of SCI-SD in the ESA hierarchy and to the DPACE through the DPAC hierarchy. Also on the ESA side the project Scientist has direct access to the development manager and may provide science requirements which need to be implemented in the Science Operations Centre (SOC). As depicted in Fig. 64 a team leader/manager will exist for each of the major development areas in ESAC namely:

- Core Processing : Production of The Global Astrometric Iterative Solution.
- Initial Data Treatment/First look : Integration of IDT/FL System
- Payload Management : Production of Payload Management software.
- Main Database : Production of MDB and transfer software.

Orthogonal to these development areas there are two other main roles:

- Configuration Manager : Maintain system integrity and software releases.

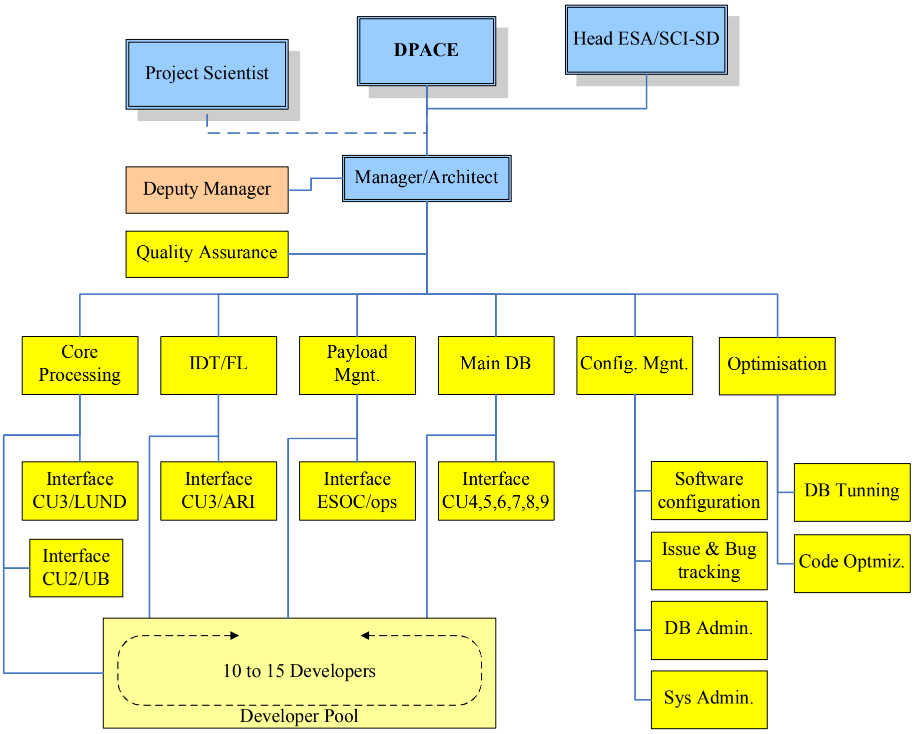


Figure 64: Organisation of the ESAC team.

- Optimisation : Ensure the best performance in terms of code and DBMS.

The development will be supported by a team of software engineers who form a dynamic developer pool. The notion here is that any developer can work on any system thus allowing easy allocation of manpower where needed to meet deadlines.

The Quality Assurance role is part time and supplied by TEC-Q thus providing independent quality assurance to all ESAC activities.

9.2.2.2 Skills and expertise at ESAC

Senior ESAC team members have many years of experience in producing high quality scientific software. The current team members collective space science experience spans numerous missions for example: Hipparcos, XMM, Envisat, Integral, Mars Express, NVO, SDSS, ScaRaB, GALEX. In addition some team members are ESOC veterans and have experience of SCOSII.

9.2.2.3 Experience and know-how at ESAC

In general ESAC is home to many missions spanning the entire spectrum, science operations and data dissemination for Integral, XMM, Herschel and Planck are on going or planned at ESAC. In addition ESAC is the archive for all ESA space science Data.

ESAC has been involved in the production of software for many ESA space science missions and is fully versed in the required software engineering standards. Moreover ESAC has experience in bringing these standards to the scientific community.

9.2.3 Facilities and services available at the ESAC

ESAC already has an 18 node dual processor cluster in use for Gaia. In addition 16 Terabytes of Disk are available. Backup facilities are already available. Many Grid compute nodes and systems are already available at ESAC. However considering the scope of the Gaia mission, budget has been identified to buy specific hardware for Gaia processing.

9.2.4 Critical points, risk evaluation and solution outline for ESAC

The primary area of concern for ESAC is timely delivery of software components from the Science Community. Work has already commenced on building the framework for slotting the science algorithms into. In Particular a testbed is being developed to explain and ease the development of algorithms. ESAC have suggested a series of six month development cycles to start the flow of software into the system and build it up over the next five years.

ESAC and CNES see ”community buy-in” to the engineering approach as important. We have already begun to organise workshops on Management, Java and Quality Assurance for the community.

There are several key areas of technology which ESAC needs to monitor, processor performance, network performance , disk space etc. In addition many algorithms have not yet been shown to be computationally possible - ESAC is following a rapid prototyping philosophy to prove initial concepts for the processing which will then be matured in to the final system

9.2.5 Preparation of the necessary infrastructure at ESAC

As mentioned above ESAC already has sufficient Hardware for the next year or so. In addition the software infrastructure has been developed to an initially usable level. This infrastructure code has been included in the GaiaTools library and will eventually form part of the overall ’Host Framework’. As this distributed processing infrastructure for Gaia matures ESAC will tailor it to the specific needs of the ESAC DPC. Although ESAC has proposed a fairly flat annual hardware budget we presume to purchase operational hardware in a few major chunks possibly in 2010, 2014, 2015. A gradual purchase of hardware could lead to a lot of inhomogeneity or purchase of expensive upgrades to possibly obsolete machines. Bulk purchases will normally get a better discount and allow for selection of the optimal hardware available when needed. However at this point in time it is impossible to set precise dates for major purchases. We wish to take advantage of the ever falling price of hardware and purchase as late as possible.

9.3 DPC-C: CNES

9.3.1 Role of CNES in the Gaia data processing

CNES will be the DPC for the three Coordination Units namely: Spectroscopic processing (CU6), Object processing (CU4) and Astrophysical parameters (CU8). Additionally CNES will be involved in the production of simulated data, which is indispensable for validating both the algorithms and the final hardware/software system. In fact CNES has already deployed the GIBIS simulator (see Sect. 6.5) in its computing facilities, and has set up a service to enable the development teams, through an Internet connection, to produce simulated data to meet their needs. CNES is determined to play a major and critical role in the scientific processing of the Gaia data.

9.3.2 The CNES team

The scientific and technical challenges are such that a DPC cannot simply be a provider of CPU power and storage media. CNES must qualify, integrate and operate software sub systems developed in the framework of CUs 4, 6, 8 and partially CU2. It is indispensable that these sub-systems be robust, reliable, efficient, maintainable and portable. The DPC will thus not only have to intervene during the operational phase but also at a very early stage in the CUs in order for the software to meet the performance requirements. This aspect was taken into account when deciding on the organisation and the preliminary work to be undertaken (a product assurance document recommending a set of rules for operations, organisation and software specifications). As a result four engineers from the CNES Gaia team will have to coordinate the technical aspects of these CUs so as to be able to guide, from the beginning, the development process of the scientific software.

9.3.2.1 Stakeholders and their roles

The CNES Gaia team reports to a manager who is in charge of: planning and coordinating work for the team, managing resources, interfacing with the CNES Directorates involved in Gaia and providing technical representation for CU4,6 and 8 within DPACE. This team is made up of a group of engineers assigned to the Gaia project and includes:

- the manager,

- a global system manager who is very active within the CU1, of which he is deputy chairman and who is responsible for coordinating the technical aspects for all of the CNES activity,
- a computing and computer network architect who is responsible for setting up the necessary infrastructures (computers, storage facilities, databases, network) both for the development and the operational phases. He plays an essential role in defining and managing the technical studies to be undertaken,
- four engineers who are respectively in charge of technical coordination of CU2, CU4, CU6 and CU8. These engineers will guide the specific architectural choices for each of the sub-systems developed by the CUs. They will also coordinate the work and contributions of the CNES experts, particularly for Quality Assurance and optimisation purposes,
- a software quality engineer

At a later date an engineer will be assigned to manage development of the host infrastructure. With the exception of the quality engineer who works part-time for the Gaia project, all the other engineers in the team will be assigned full-time to Gaia. They are all experienced computer engineers with experience in the processing of space data. Industrial teams financed by CNES are, and will remain, in charge of the preliminary studies, development work and operations. The CNES team are acting as supervisors.

9.3.2.2 Skills and expertise at CNES

The engineers in the CNES team have a great deal of technical experience in a wide range of specialised fields. Some experts are available to support them. These include:

- experts who are working for CNES and have been explicitly assigned to Gaia: this is the case for optimisation, which is already being done for GIBIS,
- experts who are made available according to needs and critical points which have been identified concerning storage, databases, networks, security, etc.
- experts who are available within companies with which CNES has signed contracts.

9.3.2.3 Experience and know-how of CNES

CNES has a proven know-how in developing and operating systems for processing space science data due to its deep involvement in many space missions over the years, covering fields as diverse as Earth observation, planetary exploration or space geodesy and astronomy. This experience over some thirty five years includes the full gamut of space science (Astronomy, Planetology, Sun research, heliosphere, magnetosphere, Fundamental physics, Life and material science) as well as Earth science. A lot of solid, pragmatic and high-quality knowledge has been acquired in the production of ground systems for the processing of scientific data for such missions.

In practical terms, cooperation with the astronomical community has led to the Corot mission (a French mission launched in 2006), as well as the ESA Herschel and Planck projects and Gaia. It was in the framework of this cooperation that CNES undertook the technical responsibility for reducing the Hipparcos data, developed the data management and command system and managed exploitation of this reduction for the FAST consortium. Among those CNES skills which are specific to the processing of space science data we highlight the following:

- the handling, storage, access and organisation of very large volumes of data contained in databases,
- the implementation of complex algorithms in high performance computing facilities,
- the transfer of large volumes of data between distant sites,
- the development of highly automated and robust processing systems adapted for continuous flow of data produced by on-board experiments,
- technical operations (system exploitation and surveillance) and the implementation of means for checking the scientific results,
- organisation of software maintenance in order to guarantee correction of anomalies and upgrading of software.

CNES's experience in space science field is a great advantage for meeting the challenging task of processing Gaia data.

9.3.3 Facilities and services available at CNES

Some general facilities and services are available at the CNES computing centre: (Note that the convention in this document is FLOPs for some number of operations and FLOP/s for some number of operations per second.

- Intensive computing facilities:
  - 200 GFLOP/s for the PC cluster (46 x 4.4 GFLOPS),
  - 270 GFLOP/s for the AIX machines:(32 x 7,6 + 4 * 6.8).

- Facilities for multi-mission development, qualification and exploitation (Linux, Solaris, Windows)
    - Service for housing secured Internet servers,
    - The STAF - file transfer and archiving service for long-term conservation of data. More than 500 terabytes of data are currently stored at STAF.
    - The SEM, which is a media exchange service for producing data on physical media (CDs, DVDs, etc.)
    - The SEF or file exchange service which provides a secure solution for all file exchanges via networks between CNES and the external network.

All of these services communicate with each other through a secure internal network. The internal CNES network is connected to the Renater research network by a 100-megabit link. This link will be extended to 1 gigabits at the beginning of 2008. On a more general level, the constant increase in computing power and the even more rapid increase of network capacity and storage media as well as the huge increase in volumes of data produced by science satellites over the next 10 years mean that facilities and services will be continuously upgraded to keep abreast of the evolving situation.

For Gaia, the CNES DPC will use both the dedicated Gaia computing and storage facilities and also existing general facilities: the GIBIS simulator is currently installed on a cluster of PCs which have been assigned for general use. For studies undertaken for Gaia, a dedicated Gaia configuration has been installed. This should make it possible to find satisfactory solutions for the different configurations required for Gaia: a study platform, a service for distributing simulation data, a development configuration, a system integration and test configuration, an operational configuration and a maintenance configuration.

9.3.4 Critical points risk evaluation and solution outline for CNES

As was pointed out in Sect. 9.1 and Sect. 9.2, taking charge of software applications developed by the scientific community has critical implications from two points of view: on the one hand the need for software to be reliable, portable, in conformity with interface specifications and fully optimised; on the other hand rigorous respect for the project schedule. In the CNES case, this latter scheduling aspect is all the more crucial with a resulting increase in risk, because CNES is responsible for receiving, integrating, validating and exploiting applications to be delivered by four CUs. In practical terms this equates to software developed by several dozen scientific institutions spread throughout Europe. Several efforts are underway to mitigate this risk, in particular:

- actions by CU1, defining the Product Assurance Plan and arrangements for software engineering aspects, training and communication activities concerning the plan, etc.
- the choice of an iterative development cycle for the entire DPAC,
- the high profile participation and significant contributions of CNES engineers in the CUs in which CNES is involved.

In addition, it can be said that taking charge of GIBIS and deploying it at CNES in 2006 in a multi-machine environment, and optimising and exploiting it from 2006 to 2012 also provide an excellent opportunity to find out the best ways of resolving potential difficulties.

Another cause for concern is the uncertainty that still exists about the precise evaluation of CPU resources. It is clear that this uncertainty will remain as long as no representative processing algorithms have been installed. The studies underway tend to suggest an approach based on an abstract conception of the application's hardware and software infrastructure that makes no suppositions about the computing resources that will be available for it. It will therefore be possible to leave the final choice of computing configuration until later (cluster or computing grid within the data processing centre, recourse to computing resources among the DPC's partners, etc.) without any consequences for the software that has already been developed.

9.3.5 Preparation of necessary infrastructure CNES

The GIBIS simulator is currently deployed on the multi-mission resources at the CNES computer centre: a PC cluster, a storage system and a secure Internet server. As far as integration, validation and exploitation of the CU4, 6 and 8 sub-systems are concerned, CNES intends to develop a specific software and hardware infrastructure and to equip it with the elements needed to take into account the unique features of any given CU. This infrastructure will make joint use of computing and storage resources reserved specifically for Gaia and also the available multi-mission resources. More specifically for Gaia, the following plan is in place:

- Mid-2006 - acquisition of an initial test configuration of three Sun Fire X4200 servers. This configuration has now been installed and is reserved for studies and experiments.
- 2009 - acquisition of more representative computing resources (in terms of computing and storage technologies) of the operational infrastructure.
- 2011 - acquisition of the operational computing infrastructure.
- 2011-2016 - incremental acquisition of storage disks and bays.

- 2015 - renewal of the computing infrastructure.

9.3.6 The challenge of the intense computing

The computing power required by the GAIA data processing is a challenge which is managed by the CNES team with the following actions or arguments :

- The today estimation is quite accurate about the data volumes. However the data throughputs shall have to be tightly controlled to avoid bottlenecks, this is today our major concern : CNES is leading studies and experiments to prepare an architectural design of adequate data repository infrastructure. The experiments will provide benchmark on :
  - cluster database servers (failover and load balancing mechanism allowing to scale easily)
  - different data organisation (full database approach or mix mode using database as a catalog and external files to store raw/observation data).
- The necessary computing power of 6000 GFLOP/s has been based on an estimation of the main algorithms (see Sect. 7.8 and [Per04]); it is only preliminary and subject to updates, but the author has taken a margin ratio of 10 (justified by the data access impact) which leaves some margins. As said before, CNES is investigating how to provide a performant data access, and is also looking for optimising the execution of data processing using multithreading approach in order to optimise the CPU use, while accessing data.

The estimation of the algorithms performance will be checked after every 6-months cycle, using the last algorithms benchmark: if any, the drifts will be detected ASAP; if the drift is greater than the x10 ratio, corrective actions (e.g. algorithms optimizations or simplifications) could be done in proper time. CNES is preparing a test and integration infrastructure with appropriate database servers in order to have representative benchmarks information.

- The estimated power of 6000 GFLOP/s should require in 2012, using the Moore's law, a reasonable set of computers; CNES is ready to procure, install and operate such a system dedicated to GAIA. Should this power be too weak, CNES should require additional computing power from its own computing center, or from other supercomputing centers. Another option should be to spread the processing in time, which should lead to some delay in the final data availability.
- The risk of underestimation of the processing power is already an identified risk in the CNES project management. It has been presented in the Phase A review

in CNES in nov 2006 (see [Mona]). It is today rated as ”medium risk”. This risk shall be monitored all along the project life, and preventive or corrective actions taken as soon as required.

For information, hereafter is a list of available technical notes issued from the tasks of the CNES DPC performance estimation : [Monb][Monc] [1]

9.4 DPC-I: IoA

9.4.1 Role of the IoA in the Gaia data processing

The Institute of Astronomy in Cambridge (IoA) will be responsible for the processing of the photometric data from the SM and AF broad-band intensities and the BP and RP dispersion spectra. The input to those processes will be the image parameters for the SM and AF transits and the raw spectra for BP and RP. During the mission, the products of the IoA DPC will be fully calibrated mean and epoch photometry (per field transit) for SM and AF, and fully calibrated mean spectral data derived from the BP and RP data streams. At the end of the mission, epoch photometry at CCD transit resolution can be released too. Reduced data will be released on a 6-monthly basis to ESAC for incorporation in other Gaia data processing activities.

9.4.2 The IoA team

The IoA team will be built at least in part from the development team, adjusted and supplemented according to the processing requirements. These requirements are in first instance estimated based on the size of the data stream and complexity of the data processing, the details of which will become more clear as developments are progressing.

The IoA team will be employed by Cambridge University through a post-launch support grant from the relevant research council (currently PPARC). This grant will cover all man power, hardware and travel costs foreseen for the data processing activities. The post-launch support grant will cover the period from launch till the agreed date of the official release of the Gaia data.

A wide range of skills concerning large-scale data processing is already present at the IoA, mainly within the Cambridge Astronomical Survey Unit (responsible amongst others for the VISTA data processing, led by Mike Irwin), the VO and grid-computing activities (led by Nic Walton), and satellite projects (Hipparcos, Planck, Gaia, led by Floor van Leeuwen). Within the UK Gaia collaboration there is additional expertise on XMM and various other relevant projects.

9.4.3 Facilities at the IoA

For the development phase a 128-node cluster will be put in place, with adequate disk space facilities. This will be used to test the data processing software, which has to be developed to run efficiently on large numbers of processors to be feasible. From the experience within the development period, and considering the availability and capacity of computer hardware in 4 to 5 years time, a data processing hardware system will be designed and obtained tuned to the needs of the tasks for which the IoA will be responsible.

9.4.4 Risk assessment

The main risk for creating a realistic IoA DPC comes from the possible complexity of the photometric calibration model, which can ultimately lead to very large numbers of model parameters. As processing-time requirements for such solutions are roughly proportional to the number of model parameters to the power three, excessively complex models can become prohibitive. Such models may be forced upon us through the complications introduced by CCD radiation damage and “curing” methods applied (for example: charge injections).

9.5 DPC-G: ObsGE/ISDC

9.5.1 Role of ObsGE/ISDC in the Gaia data processing

The ObsGE/ISDC will host the data processing centre for the variability processing under the responsibility of CU7. As such the ObsGE/ISDC will receive and process the Gaia data. The results of the variability processing will be fed back into the Main Database.

The main tasks are depicted below in Fig. 65.

In more detail the role of the ObsGE/ISDC in relation to CU7 is to provide:

1. Software Development Support

   The tasks of the ObsGE/ISDC include to evaluate the special software needs of CU7 for the variability processing and propose a coherent software environment for the data processing. The software tools and development support provided by the CU1 team to all coordination units will be carefully considered for implementation in CU7, and there will be a continuous dialogue with the CU1 team. In some cases, the CU1 recommendations may not be the best solution for CU7 and compromises will be worked out.

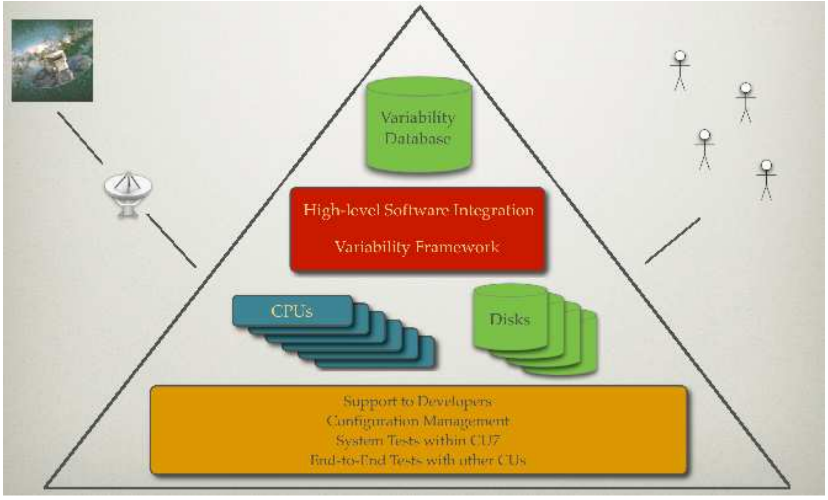


Figure 65: Schematic view of the main tasks of the CU7 Data Processing Center

All software development support solutions are also to be discussed and agreed with all CU7 partners contributing to the software development. The task includes to:

- select a single programming language.
  Currently, Java is the CU1 recommendation but C++ is also an option to be further evaluated;
- set up coding and testing standards
  including standard ways to handle parameters, errors, and logs;
- study and propose the physical data format and tools for data access.
  The data to be included in the Variability Database will be identified and specified through the data model. In general, many programs will need to read and write the same data files. Using common software components for data access allows to avoid duplication of effort. Preference is given to use existing tools and adapt them to the needs of CU7.
- evaluate and select common software libraries
  of general interest, such as astronomical and mathematical libraries, and a software development environment, incorporating a debugger and possibly a GUI designer.

2. Change and Configuration Control As per tailored [LJMD$^+$07].
3. Integration and System Tests As per tailored [LJMD$^+$07].
4. Variability Database setup and maintenance

   All data identified in the data model will eventually reside in the Variability Database. A robust and efficient technical solution will have to be designed and implemented for the database system in order to meet the data selection and access requirements. Here, database is used in a very general term not confined to a database management system in the strict sense.

5. Hardware and System Administration

   The hardware requirements for the operational phase are very important given the very large number of objects (10^9) to be processed. Realistic simulations (Eyer et al. 2005) have been performed running a prototype software program on a Grid environment. The size of the input and the variability database will increase gradually to reach the order of 500 Tbyte. The DPC will ensure that sufficient amount of disk space will be available. Care will be taken to provide sufficient backup capabilities as well.

   During the development phase, the initial requirements will not be as important, and only a very small number of powerful computers will be required. However, as early as the end of 2007 hardware will be needed to set up a software delivery, integration and testing environment. This environment needs to be upgraded in 2009 to support the first performance tests with a hardware capability of the order of a tenth of that required during operations.

   Finally, the operational hardware required for the beginning of the mission has to be set up in time for the first processing of Gaia data after launch.

6. Operational tasks

   During the operational phase, two different types of activities are anticipated. The first is the routine operation of the pipelines to process the Gaia data in a systematic way. The second is the validation of the results and a number of interactive analysis of peculiar cases, such as potential new object types to be listed in the Gaia catalogue, calibrations of odd features, etc. As a result of the second activity types, there will be a number of announcements to the community in agreement with the Gaia Science Team (GST) and the Gaia Project Scientist (PS).

   There will be a $\sim$6 months timescale in the operational work as it is currently foreseen to receive new input data every 6 months, although re-processing on a different timescale will also be supported as required.

   At the end of the development phase, a period of operations training will be organised, possibly combined with some of the end-to-end tests.

9.5.2 The ObsGE/ISDC team

The organisation of the ObsGE/ISDC team takes full advantage of the collocation of the Lead of the Coordination Unit and the Data Processing Centre for the Variability Processing.

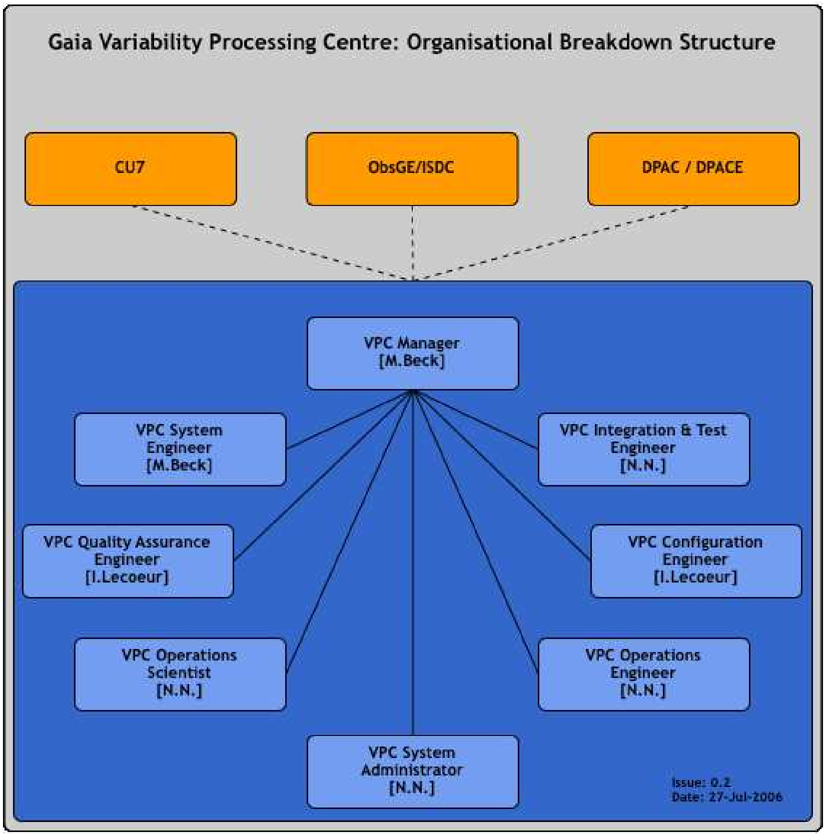


Figure 66: Organisational Breakdown of the Variability Data Processing Center

A joint team for the two activities is being set up. The relevant responsibilities are well identified and are assigned to the staff of the ObsGE/ISDC participating to the Gaia project. One of the tasks of the CU7 Technical Coordinator is the management of the DPC. The organisational breakdown is depicted in Fig. 66.

Stakeholders and their roles

- DPC Manager
  with overall responsibility for the development, installation and operation of the Gaia Variability DPC related software, hardware and other activities.

- DPC System Engineer
  with the responsibility for the overall architecture of the software system for the Gaia Variability Processing.

- DPC Integration and Test Engineer
  with the responsibility for the integration and test of the software system for the Gaia Variability Processing.

- DPC System Administrator
  with the responsibility for the procurement, installation and operation of the hardware system for the Gaia Variability Processing.

- DPC Configuration Engineer
  with the responsibility for the configuration management of the software and documentation as well as the release management.

- DPC Quality Assurance Engineer
  with the responsibility for the software quality assurance for the Gaia Variability Processing.

- DPC Operations Scientist
  with the responsibility for the scientific aspects of operations for the Gaia Variability Processing.

- DPC Operations Engineer
  with the responsibility for the operations of the software system for the Gaia Variability Processing.

Skills and expertise at ObsGE/ISDC

The ObsGE/ISDC staff involved in the Gaia activities have long-term experience in the domain of data processing for space science and in the scientific analysis of variable objects.

The experience covers all aspects from the very early requirements definition phase, the implementation and test phase as well as the multi-year operations phase.

All team members are used to working on a multi-cultural and multi-site project with external teams delivering software to the central place where integration, system tests and operations are done.

Several team members have proven experience in the scientific analysis of astronomical data, such as in the gamma ray and the optical energy ranges.

The staff members of the ObsGE have decades experiences in time series analysis in either photometry (Geneva photometry group) or in radial velocities (binary star and extrasolar planet search group) and has strong research groups in these fields as well as in stellar evolution.

Experience and know-how at ObsGE/ISDC

The ObsGE/ISDC is hosting the INTEGRAL Science Data Centre and is a member of the Planck LFI consortium.

For INTEGRAL, the ISDC plays the role of a coordination unit and the corresponding data processing centre. The ISDC interacts with all the INTEGRAL mission components: ESA's operations centre at ESOC, ESA's planning facility at ESAC, the instrument teams spread all over Europe and the world astronomical community. The tasks of the ISDC include the integration of the software required for processing INTEGRAL data, the processing of INTEGRAL data from the raw telemetry to final data products, the monitoring of gamma-ray bursts and transients, on time scales from few seconds to several hours or days, the monitoring of the instruments on board of INTEGRAL, the population and maintenance of the INTEGRAL archive, and the distribution of both data products and software to analyse them to the scientific community.

For Planck, the ISDC is providing the level 1 software for the LFI DPC in Trieste, Italy. The software processes telemetry and auxiliary data, providing the so-called 'Time Ordered Information' used in the higher level software for map building etc.

Additionally, the ISDC is providing software for the EURECA project and studying hardware for possible use on ESA's XEUS satellite.

The ObsGE has been heavily involved in the HIPPARCOS mission. First, as a member of the INCA consortium it participated to the preparation of the input catalogue. Then it made the variability analysis of the main mission photometry (in collaboration with members of the Royal Greenwich Observatory at Cambridge) and produced volumes 11 to 13 of the Hipparcos catalogue.

9.5.3 Facilities and services available at the ObsGE/ISDC

As mentioned above. the ObsGE/ISDC staff are working on many diverse projects. To support those activities computing infrastructure is available.

Some examples are:

- INTEGRAL operations network
  A SUN Sparc farm made from some 45 CPUs offering approximately 1 Tbyte of work space and currently some 10 Tbyte of disk space reserved for the INTEGRAL data archive.
- INTEGRAL analysis farm
  A PC Linux farm made from 24 CPUs offering some 10 Tbyte of storage space.
- Gravitor simulation farm
  A PC Linux farm made from 132 CPUs with fast interconnect for parallel processing.

9.5.4 Critical points, risk evaluation and solution outline for ObsGE/ISDC

- CU7 Data Model

The CU7 data model is not yet fully defined. The data model and its physical implementation need to be well known for the definition of the software infrastructure at the DPC. Efforts are now ongoing in CU7 to define the data model. This effort needs to be continued and closely monitored by the DPC. Even with a still evolving data model the definition of the infrastructure can be started.

- Input Data
  The input data is not yet fully defined. Specifically it needs to be clarified whether data from the MDB only are sufficient for the Variability Processing or whether the contents of the RAW Database needs to be available as well. This impacts on the way data is received at the DPC. If RAW data is to be accessed, continuous data transfer on a daily basis will be needed.

- Continuous CU7 involvement
  All the scientific knowledge that is encoded in the algorithms for the variability processing needs to be kept alive throughout the project duration. With time as the mission matures and even more important during post mission phase it might be more and more difficult to keep this scientific knowledge alive. Measures need to be taken to prevent this knowledge drain.

9.5.5 Preparation of the necessary infrastructure at ObsGE/ISDC

Software Infrastructure

The software framework to accommodate the variability processing software components will be defined together with the scientific components to be developed inside CU7.

Early and regular delivery and integration of CU7 developed software components into the framework are strongly encouraged. This will help to optimise the scientific components as well as the framework. To enable this, the overall approach for the DPAC will be based on cyclic development (Sect. 7.5.2).

Hardware Infrastructure

In 2006/2007 a new computer room will be established at the ObsGE/ISDC. This will provide sufficient rack space, power and air conditioning to house the expected hardware for the variability processing.

Up to the end of 2007 only a small number of processors and a relatively small amount of disk space will be needed. Afterwards, the hardware will be gradually scaled up to match the requirements of the respective development and operations phase. This is to take advantage of the ever-falling price of hardware. Additionally, with time the CPU and memory requirements of the scientific components will be better defined and the hardware can be chosen accordingly.

9.6 DPC-T: INAF-OATO

9.6.1 Role of INAF-OATO in the Gaia data processing

INAF-OATO will host the DPC for supporting AVU activities for CU3 and the Italian participation to the Gaia data processing task. A full accuracy verification of the astrometric experiment on-board Gaia must be a well-structured effort, focused on those data processing areas critical to mission success, and capable of gauging the degree of success throughout the mission (section 5.1.8). For these critical areas, independent procedures/models are designed, implemented, and operated , and results compared to the baseline processing pipelines.

INAF-OATO will provide software development support to AVU. More general support will be provided to the Italian participation to the activities to the other CUs (CU2, instrument model; CU4, extra-solar planets and solar system objects; CU5, absolute calibration; CU8, spectral library).

9.6.2 The INAF-OATO team

The INAF-OATO team is being set up, the relevant responsibilities are identified and assigned. Several of the Italian institutes participating in Gaia will contribute to the formation of the INAF-OATO DPC, each contributing their skills and know-how.

9.6.2.1 Stakeholders and their roles The INAF-OATO DPC team is managed by the AVU manager and its small and agile managing structure. As part of CU3 the AVU manager will report to the CU3 manager. The coordinating team will be the following members:

- DPC Manager with the responsibility of the designing,developing, implementing and operating of the DPC related software, hardware and other activities;
- DPC Operation scientist with responsibility for the scientific aspect of the operation phase and first evaluation of the comparisons results;
- DPC Quality Assurance Engineer with responsibility for software and documentation;
- a DPC software engineer;
- DPC System and DB Administrator, responsible for the procurement, installation and operation of the hardware and software system for the DPC.

9.6.2.2 Skills and expertise at INAF-OATO The staff of the INAF-OATO DPC team has a long-term experience in data processing for space missions, in managing large astronomical archives (Objectivity, ORACLE, MS-SQL Server), and in developing pipelines, analysis and interrogation software, in different languages(FORTRAN, C, C++, IDL, C#, Perl, ...) and platforms (Solaris, Linux, Windows, VMS).

9.6.2.3 Experience and know-how of INAF-OATO Many of DPC team members have experience in producing high quality scientific and system software for various projects and missions (Hipparcos, HST, GSC2, SOHO, GALEX, PLANCK, SOLARNET) and are involved in projects for the distribution of scientific data (Virtual Observatory).

9.6.3 Facilities and services available at INAF-OATO

At time of writing, the facilities and services available at the INAF-OATO computing centre and at other Italian institutes which are part of INAF-OATO DPC effort are as follow:

- 120 GFLOP/s (16 CPU Beowulf)
- 22 TB RAID arrays available
- (beginning 2007) 20 GFLOP/s Shared memory machine (SGI Altrix 450) with 100 GB of RAM to handle up to 1 million of well-behaved stars for the sphere solution
- Grid infrastructure to all the INAF institutes
- (end 2006) HPC machine with 80 nodes (300 cores) and 50TB of storage at INAF-Catania and Catania-INFN

9.6.4 Critical points risk evaluation and solution outline for INAF-OATO

The highly distributed processing for the Gaia data reduction requires an architectural model for building the DPCs which must give appropriate emphasis to the engineering aspects, e.g., hardware organization for connections to different sites, computing, data base operations, etc. Significant experience in management, programming (especially Java based) and QA procedures are key to the success of the project, and the Italian community is already in the process of organizing itself to respond to these needs.

As the Data Model and the ICDs between the MDB at ESAC and the different DPCs have not been finalized yet, the necessary resources for the INAF–OATO DPC cannot be completely defined at this time. The expectations are that any critical issues concerning the INAF-OATO DPC design and realization will be resolved by the end of 2009. This is considered as the minimum lead time necessary to our DPC for effectively responding to new needs before the overall test of the Gaia Ground System is exercised (end of 2010, beginning of 2011).

9.6.5 Preparation of necessary infrastructure at INAF-OATO

Computing power and storage needs available and soon to be available at the INAF-OATO DPC are deemed adequate for the proper operation of the experimental DPC up to the end of 2007.
At the begin of 2008, the hardware will be scaled up for the final experimental tests and preparation will commence for the acquisition of the final hardware configuration . It is anticipated that this phase will last until the end of 2008 beginning of 2009. The hardware for the operational DPC will be gradually acquired during the following 12/24 months.

9.7 DPC-B : BPC

9.7.1 The role of the Barcelona Processing Centre in the Gaia data processing

In the frame of the Gaia Data Processing activities, the Barcelona Processing Centre (BPC) will be used as a unique label to encompass the activities that are and will be performed in two separate institutions, namely:

- The Barcelona Supercomputing Centre (BSC-CNS);
- The Centre de Supercomputacio de Catalunya (CESCA).

The general goals, hardware and management of these institutions are very different but both are sharing the interest to contribute to the DPAC tasks. Their activities in the Gaia data processing are:

- The BSC will run the successive versions of the GASS simulator at least until the operational Gaia Data Processing System has been set up. It will host and run the GASS simulator producing simulated telemetry to feed the IDT and AGIS processes as well as intermediate data to experiment with other reduction algorithms. This task is performed depending on CU2, and in close collaboration with the CNES team and the BSC engineering group. This task has been running at the Mare Nostrum Supercomputer since the beginning of 2006 using Grid-superscalar to distribute the task on up to 250 processors.

- The BPC will host a Raw Database in order to allow the off-line running of the IDU process, one of the CU3 tasks. Good connectivity to ESAC will be required to update the Raw Data Base on a daily basis as ESAC will produce new raw data by running IDT, under CU3, each day.

- During the operational phase, the IDU process will be run at Mare Nostrum (actually its successors), thus interfacing with the MDB at ESAC and through the MDB interact with the other CUs.

- Before implementing and running the GASS simulator at BSC, testing is performed at CESCA, which in addition offers a large storage area (5 TB at present), where results from BSC are stored and analyzed before they are sent to ESAC. Permanent storage in a magnetic tape robot is also provided by CESCA.

- A proposal to run image stacking on BPC is now under evaluation.

- CESCA premises are used to implement and validate the successive versions of IDT. Once it has been validated through detailed checking with the data in the simulator files, IDT is implemented at ESAC. Since it evaluates the quality of the data of the IDT, which are the inputs for the reduction processes, this test plays an important role in all the data processing tasks.

- BPC will thus directly contribute to the tasks of CU1, CU2 and CU3 and provide hardware for CU2 and IDU, which is one of the more demanding processes in DPAC.

9.7.2 The Centres and the teams; the Gaia team

The Spanish Government, the local Catalan Government and the Polytechnic University of Catalonia (UPC) took the initiative of creating a National Supercomputing Centre in Barcelona in 2004. BSC is the National Supercomputer Facility in Spain. Funded in 2005, it has inherited all traditions of the well-known CEPBA Institute in Parallel Computing in Europe, incremented with the incorporation of Mare Nostrum, the leading Supercomputer in Europe.

The mission of BSC is to investigate, develop and manage information technology in order to facilitate scientific progress. BSC is defined as a research centre instead of simply a supercomputer facility. However, BSC does not want to become just a supercomputing research centre, but a research centre in other science areas where supercomputing is a must.

The BSC team is a very large one bringing together specialists in architecture, deep computing, performance tools, etc. (see http://www.bsc.es/ for more details). In addition they have commenced work in specific scientific areas dedicated to Life Sciences and Earth Sciences (the BPC started its activities late 2005).

Up to now a team dedicated to Gaia has not yet been built. An important collaboration has been established around the simulator and to study IDU problems. This will be the base for a larger and stable team as described below.

CESCA was created in the 1990's by the Local Government, the Catalan Foundation for Research and the Catalan Universities. It aims to provide supercomputing resources as well as communication tools between scientific institutions. CESCA has been one of the teams that designed, deployed and tested the first Gaia prototypes both for the Gaia System Simulator (GASS) and for the Global Iterative Solution (in GDAAS).

To run the GDAAS system, up to four different platforms were used at CESCA, from an old IBM SP2 to the Compaq Alpha Server HPC320 (8 ES40 nodes, 4EV68 processors each) used in 2005. Several experiments on portability and performance of the system were done at other systems (Beowulf Compaq with 8 DS10 nodes, and Grid systems). On the other hand, as the needs of storage increased, CESCA implemented a Storage Area Network with an EVA of 60 discs providing 5 TB of storage. Several DB engines (Objectivity, Oracle 9, Oracle 10) were used at CESCA to run GDAAS.

The Gaia-Grid was deployed from CESCA in 2004 to provide intermediate data of binary systems and to check the Grid approach in a network of 23 nodes distributed over 8 institutes in 5 European countries.

The team that supports Gaia activities at CESCA is, at present, rather small, although as explained above it is very well acquainted in the kind of problems that Gaia data treatment faces. At present the group is composed of two engineers devoted to Gaia a fraction of their time.

A Gaia Engineering Team will be built around the small cores already existing at BSC and CESCA plus the group at the UB. The National Plan of Space has recently confirmed and enlarged its support to the Gaia group at the University of Barcelona which acts as a core of the Gaia project in Spain. The approved project has important manpower resources in the engineering area. Hence, a team of five full-time engineers will be built in the coming months.

9.7.3 Facilities and services available at BPC

BSC and CESCA host several computers and services. Those with a relevant contribution to Gaia data processing are:

- CESCA
  - HP CP4000: 16 nodes DL145 G2 (2 AMD64 Opteron 275 dual-core, 2.2 GHz, 64 KB/1 MB each core), 256 GB of main memory, 4.56 TB of hard disk, with an Rpeak of 281.60 GFLOP/s and an estimated Rmax of 177.41 GFLOP/s, interconnected by 3 Gigabit Ethernet).

  - HP Proliant DL360-G4p, with 10 dual-processor Xeon nodes at 3.2 GHz, 360GB disk Ultra320 and 8 GB of memory per node.
  - SGI Altix 3700 Bx2: 128 processors Itanium2 (1.6 GHz, 16 KB/256 KB/6 MB of cache), 368 GB of main memory, 5.13 TB of hard disk, with an Rpeak of about 800 GFLOP/s and an estimated Rmax of about 700 GFLOP/s.
  - Storage Area Network (SAN) (EVA 2C6D-B) of 5 TB connected by Fiber Channel of 2 Gbps.
  - An automated tape library StorageTek TimberWolf 9740 containing 302 tapes of 9840 type with native capacity of 20 GB, and 2 transfer devices 9840, each of them with a transfer speed of 10 MB/s and a cartridge exchange rate of 350 per hour.
  - CESCA manages the network connection to the Spanish RedIris net using Gigabit Ethernet technology. There are two 2.5 Gbps links to Madrid. Through RedIris access is open to Géant2.
  - CESCA is a computer centre, so the normal way of operation is sharing computing time and resources. To overcome this problem, two of the Proliant nodes are exclusively devoted to Gaia, and depending on our computing needs computing nodes at CP4000 will be exclusively devoted to us. 5 TB SAN is exclusively used by the Gaia team.

- BSC

  - The main tool that BSC can offer to Gaia data processing is the Mare Nostrum supercomputer.
    Mare Nostrum is a supercomputer based on PowerPC processors, the BladeCentre architecture, a Linux system and a Myrinet interconnection. These four technologies configure the base of an architecture and design that will have a big impact in the future of supercomputing.

    A summary of the system is:

    - 4.812 2.2 GHz IBM Power PC 970FX processors (2406 dual 64-bit processor blade nodes)
    - Peak performance of 42.35 TFLOP/s
    - 9.6 TB of main memory
    - 236 TB of disk storage
    - Mare Nostrum has three interconnection networks: Myrinet, Gigabit Ethernet, and Ethernet 10/100.

    There is the commitment of the leading institutions of maintaining Mare Nostrum among the fastest computers in the world. An update of the computer has already been approved on a two-year horizon.

10 DPAC : Composition and Management

This section defines DPAC, its resources (people, funds and DPCs) and management, with a breakdown of the resources per CU.

10.1 Definition of the DPAC

10.1.1 Role of the DPAC

The Gaia scientific community is composed of a broad and extensive body of scientists who have strongly supported the Gaia mission before its selection and will eventually make use of the Gaia data products for their basic research. However, not every member of this extended community will participate actively in the actual data processing as it is described in this document. The subset of the scientific community comprising all the people seriously involved in the data processing (that is to say, taking part in the development phases and providing either management effort, algorithms, software or is member of one of the DPCs) has organised itself in the form an international consortium called DPAC (Data Processing and Analysis Consortium). The DPAC is the community structure responding the ESA Announcement of Opportunity. The DPAC has set up an internal organisation and management structure and secured the availability of the necessary hardware to carry out the tasks. It is formed around a set of Coordination Units (CU) each responsible for a key aspect of the data processing. CUs may be sub-divided into smaller groups called Development Units (DU) and in charge of one or more work packages. The CUs are supported by a set of Data Processing Centres (DPC) and the overall coordination is performed by the consortium executive (DPACE).

In short the DPAC is responsible for:

- preparation of the data analysis algorithms to reduce the astrometric, photometric, and spectroscopic data within a coherent and integrated processing framework, including special objects such as multiple stars and minor planets;
- generation and supply of simulated data to support the design, development and testing of the entire data processing system;
- the design, development, procurement and operation of all aspects of the hardware and software processing environment necessary to process the mission data throughout the simulation, mission operations and final catalogue production phases;
- the design, development and operation of the final Gaia database, which will contain the intermediate and final mission products of interest to the scientific community at large.

All the above items are made more explicit in the different sections of this document and should be sufficient to explain how the DPAC structure is optimised to cope with the very challenging task of the Gaia data processing.

10.1.2 Membership

As explained above, the DPAC is not equivalent to including every member of the scientific community having an interest in the Gaia mission and its results. This would be too loosely defined and would not provide a clear membership of who is actually in the Consortium. Likewise without a strict definition there would be no way to assess properly the effort involved and to provide the national identification of who is entitled to claim funding on behalf of DPAC activities.

For these reasons the membership criteria must be clearly established so that one can have at any time an official list of members that will be updated periodically. The membership will evolve during the DPAC lifetime with the arrival of new members (e.g. PhDs, new permanent positions, temporary positions created for a specific activities, etc.) or the departure of previous members for any reason. The membership must imply having real duties (tasks to be fulfilled with real constraints) towards fulfilling DPAC goals and responsibilities, and this will grant rights to the individuals listed as belonging to the Consortium.

It must be clear that only individuals are members of the DPAC, and not institutes or groups, whatever their implication in the Consortium. Much effort must be excercised so that all the committed people with identified responsibilities are included, together with the less visible individuals in their institute spending a significant fraction of their time in Gaia related activities, in case they are not directly associated to a CU. The principle adopted is that any DPAC member must be identified by his (her) involvement in one or several work packages within a particular CU.

In practice the DPAC membership is inherited from the participants recognised by each CU, supplemented when applicable by the few supporting individuals that might work closely with the DPACE chair or in a DPC, and not associated to a specific CU. The official membership is revised and formally adopted by the DPACE based on proposals made by the CU managers or the DPACE chair. An electronic version is currently maintained by the Project Scientist Support Team at ESTEC and accessible on the Gaia website, together with the associated distribution lists. Each CU manager (or a person appointed by him or her) is responsible for the maintenance (adding or deleting entries) of the membership list for that CU. This list is the official version (and the only one) of the membership, giving the identification and personal details of every member to which the permissions to access the Gaia electronic documentation (accessible via Livelink) and the GaiaWiki are granted. The different lists can be consulted by any member of the DPAC and interrogated to extract sub-lists matching particular criteria. Each member can update his/her personal details. The DPAC statistics as of the submission of this proposal is given in Tab. 19.

An additional distribution list will be set up to include the contacts in the funding agencies or the directors of the main collaborating institutes, so that they could receive the DPAC newsletter or any other relevant documentation designed for a wider circulation.

10.2 Human resources of the DPAC

As mentioned in the previous subsection, with few exceptions, the members of the DPAC are those individuals committed to specific workpackages. As such it will be a constantly changing list, especially as funding will allow work-specific contracts to be filled. It is expected that each CU will maintain it's own membership list and track changing commitments. Thus, the membership of the consortium is primarily inherited from the CU membership as communicated by each CU manager. Here we present the current membership as it stands based on commitments made to date.

Currently (as of 25 March, 2007) there are 325 members in DPAC. The breakdown by country and by CU is given in Table 19 and shown in Figure 67. The country is defined as the place where the position is currently funded, usually the country where the person is working (not their nationality). The statistics are based on the domain name of the email addresses and this leaves only a few cases not easily linked to a particular country. The individuals recorded in this table are those with identified tasks or responsibilities in the CUs, and thus may exclude for the moment IT people working part or full-time for Gaia within an institute if they are not listed as a member of a particular CU.

Also, it should be noted that ESA's contribution to the DPAC (14 people) is included and indicated in Table 19 and Figure 67. See also appendix D.

Obviously such statistics should be interpreted with care, as it does not reflect the various levels of commitment that people have made to DPAC, and this can vary significantly from person to person; a better measure of the human resources available to the DPAC is the *Effort Available*, taking into account the fraction of time that people have dedicated to the Consortium. We measure the Effort Available in units of Full Time Equivalent(FTE), whereas *Effort Required* will be measured in Man Months (MM). For scheduling purposes it should be noted that 1FTE = 10MM should be adopted to take into account vacation and a approximate 10% contingency factor.

Based on current secured funding levels, it is expected that both the membership and the FTE will increase significantly in 2007, as contracted positions are filled. Table 20 shows an estimate of the FTE that will be available, with the a breakdown by country and CU, showing a total of approximately 180FTE. Figure 68 shows the relative contribution to the total DPAC Available Effort for 2007, by country and by CU. **It should be noted that ESA's contribution to the DPAC is included and**

Table 19: Composition of the DPAC as of 25 March 2007. The column *People* gives the number of individuals currently supported in each country for the whole of DPAC, with subsequent columns showing the breakdown for each CU.

Country	People	CU1	CU2	CU3	CU4	CU5	CU6	CU7	CU8
France	82	8	22	7	27	1	29	5	18
Italy	62	1	13	20	7	19	1	10	12
Germany	28	1	3	20	3	1	2	1	4
UK	26	2	3	2		20	5	1	1
Belgium	23	1			11		7	10	4
Spain	21	3	10	8		9		1	3
ESA	14	13	1	9					1
Switzerland	13	1	1		2		3	10	
Sweden	10	1		2	0				8
Greece	9		7		2				7
Portugal	8	3	2					3	
Brasil	4		1	1	2	1			
Finland	4		1		3				
Denmark	3					1			2
Lithuania	3								3
Netherlands	3		1			3			
Slovenia	3						3		
others	9		2		1			5	1
total	**325**	**34**	**67**	**69**	**58**	**55**	**50**	**46**	**64**

indicated in Table 20 and Figure 68.

In Appendix A an estimate of the Effort Required is given for each top-level Work Package as a function of time. In comparing the Effort Available with the Effort Required it is important to note several important points. First, the estimate of the Effort Available (FTEs) for each year does not take into account when new positions will start: a person hired in 2007 and working full-time for DPAC is counted as 1 FTE, even if that person may not start until late in the year. A precise comparison to the Effort Required should take into account the start and end dates of the positions. It should also be pointed out that for some WPs, especially those that have a significant developement effort, it is quite difficult to estimate how much effort will in fact be required, thus the estimates for some of the WPs are quite approximate. Finally, as stated in section 2.1, the DPAC intends to go beyond the required goals of the data processing. This means the Effort Required for WPs that are part of core processing is not the same as the Effort Required for WPs dedicated primarily to producing additional data products that are not part of the primary science products; it is the

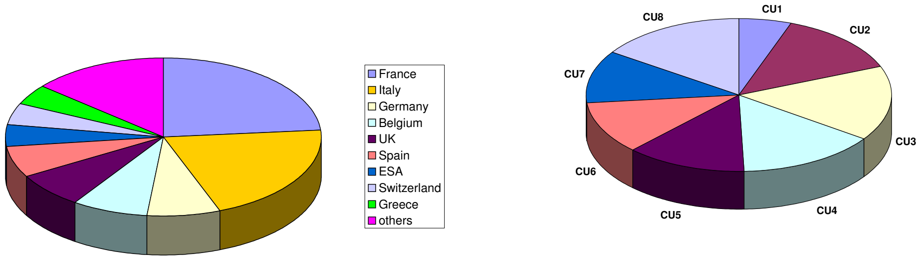


Figure 67: Relative distribution of DPAC members, by CU and nationality.

Table 20: Available Effort for 2007

Country	CU1	CU2	CU3	CU4	CU5	CU6	CU7	CU8	DPCs	total FTE
France	3.8	8.2	1.55	5.9	0.6	6.7	1	5.35	9.7	42.8
Italy		3.75	7.8	3.3	6.7	0.3	2.4	2.6	1.6	28.45
Germany		1.2	11.35	0.35	0.6	2.2	0.6	3.5		19.8
Belgium				6.2		1.7	3.65	1.9		13.45
UK		0.1			12.75	5.1	0.1		0	18.05
Spain		4.9	6.25		5.1		1.2	3.45	1.5	22.4
ESA	9		5.6						1.2	15.8
Switzerland				1.05		0.3	3.9		1.7	6.95
Greece		1.8		1.25				2.25		5.3
Sweden			1.5					2.3		3.8
Portugal	2	1.2	0.5				0.9	0.2		4.8
Finland				1.3						1.3
Lithuania								1.8		1.8
Netherlands					1.5					1.5
Slovenia						0.3				0.3
others			0.2	0.3			3.4	0.9		4.8
total	**14.85**	**21.15**	**34.75**	**19.6**	**27.25**	**16.6**	**17.15**	**24.25**	**15.7**	**191.3**

responsibility of each CU manager to appropriately prioritize the work and to allocate resources as needed.

It is worth noting the different forms that these human resources take, in the sense of the type of positions that people are actually in. Broadly speaking the positions

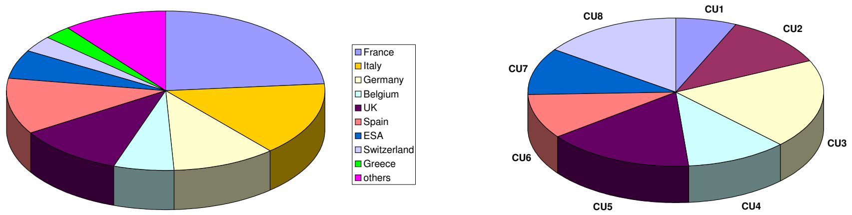


Figure 68: Relative distribution of Available Effort, by CU and nationality, for 2007.

can be considered as being of two different types. One is work specific contracts, where people have been selected for positions defined by the work to be done. These include post-doctoral fellowships at academic institutes to long-term contracts dedicated to a specific project. These positions exist only as long as the funding can be secured to support the positions, but they have the advantage that the amount of effort dedicated to the project is well defined. The other type is positions that are permanent, most often in the form of tenured academic positions. Here the position is not at the mercy of current funding, but the level of commitment is sometimes difficult to quantify and is not contractually assured over a fixed time. In practice people in such positions are involved in the project by choice; as such these people can be highly motivated and highly committed, especially if they have already invested a significant amount of time to the project, but newcomers may be quite uncertain in their commitment and capable of leaving the project.

Given these two different profiles, which often represents two different work cultures, one typical of industry the other of academia, effort must be made to maintain an effective communication network between the components of the consortium to maintain a healthy and fruitful collaboration. On a practical level membership in DPAC will, to some degree, be performance dependent: work assigned to those who do not meet their commitments will have to be reassigned to others; CU managers will have to track the progress of the work, setting periodic milestones and evaluating whether they are being met. Partially motivated by this the consortium has chosen the cyclical developement approach outlined in Sect. 7.5.2.

10.3 Funding of the DPAC activities

In this section we identify the main funding agencies involved in supporting the DPAC activities, and those which shall provide Letters of Commitment (see table 21). These agencies are nationally based funding agencies supporting the activities of the DPAC members from their respective countries. **In particular, the activity of each CU has**

Table 21: Funding agencies that have been asked to provide a Letter of Commitment

Agency	country
BELSPO	Belgium
CNES	France
INSU	France
DLR	Germany
ARI	Germany
Dresden	Germany
GSRT	Greece
Univ. Athens	Greece
INAF	Italy
ASI	Italy
MEC/PNE	Spain
SNSB	Sweden
Geneva University	Switzerland
SSO	Switzerland
PPARC	United Kingdom

multiple sources of support. Furthermore, as a consequence of the DPAC being an international scientific collaboration, WPs were not assigned to institutes, but rather individuals are committed to WPs without regard to nationality. As a consequence individual WPs (and their deliverables) cannot always be mapped to a single funding agency.

Many of the institutions in Table 21 have already been supporting the scientific community involved in Gaia in the past years of mission preparation, while for others Gaia is a relatively new project. Also, not all agencies involved in supporting DPAC are considered here, but only those making a significant contribution to the DPAC at the time this document was prepared.

It is worth noting that support from these agencies comes in two general forms: financial support, typically awarded to the PI's of funding proposals, and ”in-kind” support, such as dedicated infrastructure (including computing facilities) and human resources (such as permanent staff positions) attached to the supporting institution. While this later can be measured in units of currency, it is clearly not of the same nature as financial support which can be transformed into a variety of resources (HW, travel, contracts, etc.). Nevertheless, it is important to take into account this second form of support as it will be significant in some countries and the preferred

Table 22: Profile of Available Effort for 2007 per country

country	non-permanent FTE	permanent (staff) FTE	total FTE
Belgium	11.2	2.25	13.45
France	14.5	28.3	42.8
Germany	11	8.8	19.8
Italy	8	20.45	28.45
Spain	7	25.4	22.4
Switzerland	5.6	1.35	6.95
United Kingdom	17.1	0.95	18.05

form of support for many academic institutions.

Another factor which will vary from one funding agency to another will be the duration of the promised support. Many institutions are legally constrained to limit the duration of their funding awards and are not able, albeit they may be willing, to quantify their future support beyond a certain number of years. This is an important consideration for a long-term project like the DPAC, and means that those involved in DPAC activities will have to repeatedly secure funding to support their activities in the years ahead.

There are two notable exceptions to the national profile of the funding agencies supporting DPAC, which are not listed in Table 21. The first is the contribution of ESA itself, mostly through the activity of the ESAC DPC, whose efforts mainly falls under CU1, headed by the ESAC SOC Development Manager, and CU3. (This contribution is detailed further in Appendix. D.1.) The other exception is the support of the EC through the Marie Curie Research Training Network ELSA (European Leadership in Space Astrometry); post-doctoral and doctoral positions funded by this RTN will contribute only indirectly to DPAC, partly through the research carried out by those funded, which may influence final algorithm developement, and partly by contributing to the training and developement of young researchers who represent future DPAC members. The young researchers funded by the ELSA network will be affiliated members of the DPAC, giving them access to relevant documentation, but will not be assigned to DPAC workpackages.

10.4 Management structure

10.4.1 The DPACE

Our overall organization gives the CUs much autonomy in the way they handle their part of the data processing and the internal organization and management structures do not need to be uniform across CUs. However, there is a single goal shared by all the CUs which must follow a common schedule and adhere strictly to many interfaces so that the results produced by one group are available in a timely manner and may be used efficiently by other groups. The standards, the MDB (Main Database) content and structure, the processing cycles must be agreed collectively. Therefore, in addition to a local management of each "coordination unit", the overall DPAC is coordinated and managed by an Executive Committee, called DPACE for "Data Processing and Analysis Consortium Executive", in accordance with the SMP. This overall management structure of the Consortium deals with all the matters which are not specific to the internal management of a CU and is meant to make an efficient interaction between the CUs possible. The DPACE responsibilities are primarily coordination tasks although it will make important decisions to be implemented by all CUs which are akin to real management.

The DPACE will meet at least once per developement cycle (2 times per year).

10.4.1.1 Role of the Executive Committee

The list below indicates the major roles of the Executive in complement to the management structure set in each CU.

The Consortium Executive Committee (DPACE) :

- oversees the DPAC activities in consultation with the CU and DPC leaders.
- agrees upon and communicates to the DPAC the overall policy of the data processing.
- agrees on the DP cycles, schedules, MDB versioning.
- agrees and endorses standards, documentation, testing procedure common to all CUs.
- reviews and validates the periodical management reports issued by the CUs and DPCs.
- compiles progress reports to be presented to the GST and communicated to the community.
- proposes to and agrees with the GST on the schedule and content of the intermediate data release.

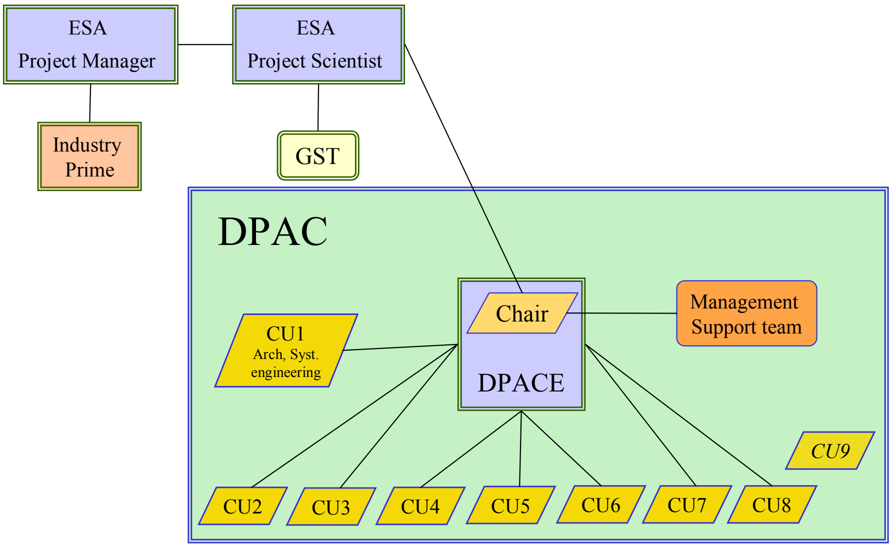


Figure 69: The upper level management structure of the DPAC and its relationship with the ESA management levels for the Gaia mission.

- advises the GST on data rights policy.
- acts as a communication point between the CUs and DPCs, when this is relevant.
- agrees and endorses the Q&A procedures which are common to all CUs as proposed by CU1.
- discusses with the CU the replacement of CU leader as appropriate, and eventually confirms the new CU leader nominated by the CU.
- takes the necessary steps to solve conflicts of any nature between CUs, such as:
  - unnecessary duplication of tasks between CUs
  - assignment within a CU of a new task
- establishes criteria for DPAC membership
- maintains overall Risk Register for DPAC
- monitors general resources that may have an impact on the DPAC like,
  - major change in the funding within a DPAC group
  - any modification of tasks or processing with a serious cost impact

10.4.1.2 Composition of the Executive Committee

The DPACE is primarily a technical committee discussing all the matters relevant to the DPAC objective, namely the achievement of the Gaia data processing from the data reception to the production of the final scientific database (together with the intermediate date release(s)) within a strict schedule. This is a daunting task in term of organization due to the technical complexity of the whole system (development, implementation and operations) combined with the number of different national organizations willing to participate and investing significant effort and resources into a cooperative activity and that for many years before the production of tangible scientific results. No doubt numerous problems will surface during the DPAC lifetime and not all them will be of pure technical nature and will have to be resolved outside the CUs. They will be handled by the DPACE, as soon as the issue and the decisions go beyond the management of a Coordination Unit or when it cannot be resolved by agreement between CUs. The DPACE composition must be such that it can cope with all these issues and potential problems. Our proposal comes with an initial composition that will very likely evolve in the future to meet the DPAC needs. It belongs to this initial committee to bring the necessary adjustments that may prove necessary to ensure a smooth running of the DPAC over the long term.

The nominal composition is as follows:

- A Consortium chair who is not the leader of a CU
- A deputy chair
- The CU leaders
- A representative of the CNES DPC providing technical representation for CU4, 6 and 8

The Gaia Project Scientist has the status of observer in the DPACE, meaning that he assumes no role for the community but is involved in all the discussions and aware of the decisions made by the DPACE regarding the data processing. It is anticipated that this coordination level between the ESA Project Scientist and the DPAC executive will resolve all possible conflicts relating to the scientific output of Gaia as defined by the processing undertaken by the DPAC. The Gaia PS has a standing invitation to all the DPACE meetings.

10.4.1.3 Reporting at DPAC level

Each CU will manage it's internal reporting as it best sees fit. Nevertheless, to maintain an global overview of the DPAC work it is necessary that each CU and DPC submits reports to the DPAC Chair. These reports will be given at each DPACE meeting, held at least once per developement cycle. The reports will include: Membership and funding status, milestones/objectives reached since the last report, milestones to-be-achieved for current/up-coming cycle, critical deliveries made or received, any late deliveries received and impact assessment of same, lessons learnt, critical unresolved issues.

The DPAC Chair Management Support Team (described below) will query the CU leaders for any necessary information to maintain the Master DPAC Schedule, as necessary. In any case, the CUs are encouraged to maintain a certain level of transparency of work in progress through use of the Gaia wiki pages.

10.4.2 The DPACE chair, deputy chair and management support team

The DPACE is led by a nominated representative from the scientific community. He/she has specific responsibilities to run the Consortium and to implement the decisions of the DPACE. The amount of work, let alone the potential conflict of interest, makes this position unsuitable for a CU leader. The chair is appointed for a renewable mandate of three years and is assisted in his task by a deputy chair, also a nominated member of the scientific community and with a similar three year renewable mandate.

The responsibilities of the chair of the DPACE are:

- chairs the Consortium Executive Committee
- prepares and convenes the DPACE meetings
- is in charge of the overall coordination of the Consortium, as delegated by the DPACE
- monitors the overall progress (Master Schedule) of the DPAC
- coordinates the preparation and submits the response of the community to the AO on behalf of the DPAC
- coordinates the preparation and submits the material for the DPAC wide reviews to the PS
- acts as the main contact point between the Consortium and ESA

In the case that a consensus can not be reached by the DPACE on an urgent issue, the Chair may call for a vote; in the case that there is not a majority, the vote of the Chair carries. The DPACE chair has special contacts with CU1 where most of the activities relevant to all CUs are taking place.

The responsibilities of the deputy chair are:

- compile the minutes of each DPACE meeting
- maintain the DPACE wiki pages
- assist the chair in fulfilling his/her duties

If for any reason the chair is not able to fulfill his duties, the deputy chair will be able to temporarily act in his place until the chair can return to active duty, or the DPACE confirms a new chair.

The DPACE Chair will be supported by a Management Support Team. In addition to the Deputy Chair, this will include a Project Controller, charged with monitoring the overall master schedule of the DPACE, including critical deliveries, and a System Engineer, charged with managing interfaces, especially those with ESA (requests and deliverables). The Project Controller and the System Engineer will report directly to the DPACE Chair.

The tasks assigned to these support positions are further detailed in WPs in appendix C.

10.5 DPAC Master Schedule

The schedules of each CU is maintained and monitored by each CU manager, who reports to the DPACE the achievement of milestones and reception of deliverables from other CU's. It is the responsibility of the DPACE Chair and Deputy Chair to monitor the overall DPAC Master Schedule using these CU Status and Progress Reports. (See also section 10.4.1.3.) As mentioned in the previous section, they will be assisted in this role by a Management Support Team, including a Project Controller (Work Package GWP-T-010-00000) with project management experience. This position has yet to filled, so a detailed Master Schedule for the entire DPAC cannot be presented at this time.

However, toward establishing the Master Schedule, the DPAC has identified the *critical deliverables* that must be made between the CUs: these deliverables are ”critical” in the sense that a late delivery on the part of the provider can potentially cause a slip in the delivery schedule of the receiving CU (the customer). These critical deliverables are listed in table 23, with the delivery schedule indicated by the notation ”TN”, indicating the end of the Nth developement cycle. As described in section 7.5.2, deliveries of data and SW, as well as integration and testing, takes place every 6 months – see this section for the end-dates of each cycle. This table does not include the transfer of telemetry and processed Gaia data; a complete mapping of the dependencies of the data products from each CU will be completed at a later date when each CU has completed it's detailed functional analysis, and is the responsibility of

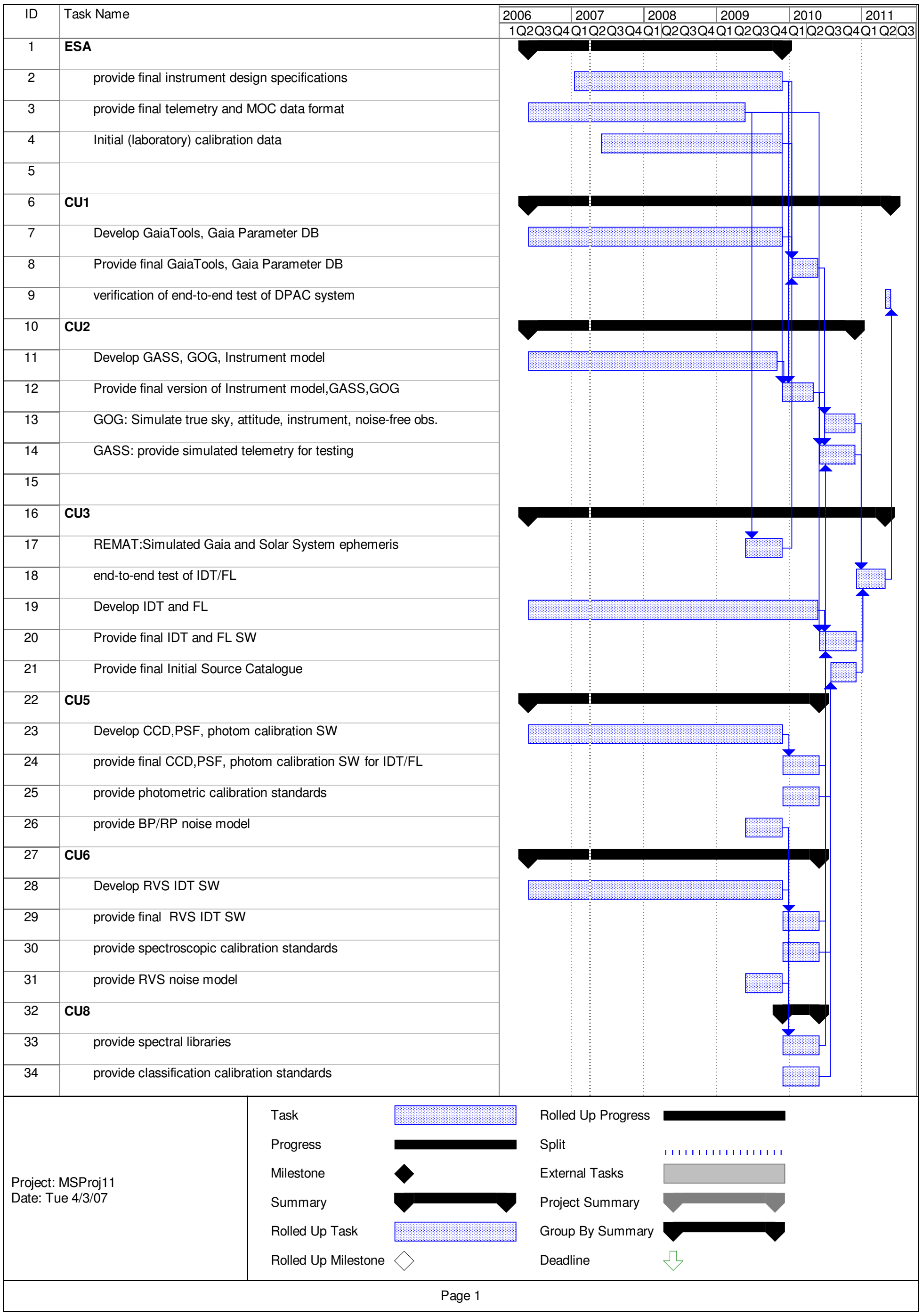
CU1. Most of the deliveries listed in table 23 are specifically for the developement and verification phases of the DPAC activities, ending with cycle 10 (T10). However, some deliverables will continue supporting data processing and science verification activities during and after operations: these deliverables have EoM (End of Mission) indicated as the last delivery date.

From the list of critical deliverables the detailed Master Schedule can be built. As just mentioned, the Master Schedule for DPAC cannot be presented at this time. However, for this proposal we present an important element of this schedule: the schedule for completing the IDT/FL end-to-end testing in cycle 10. We choose this element of the DPAC system because it appears to be launch critical, from the point of view of ESA; this element must be in place once the science telemetry begins. Other elements of the processing system will either not receive data until six months after the science telemetry begins to be received, or will be able to delay their processing as long as their host systems for receiving data transmissions from the MDB are in place.

Figure 70 shows the Gantt chart for the IDT/FL end-to-end testing. The minimum detail of the CU activities is shown here, just sufficient to show the source and endpoints of the inter-CU and ESA deliverables. Only the latest cycle for tasks making or receiving deliveries are shown, and SW development tasks are not resolved into individual cycles, but are compressed into single tasks. The schedule shows the importance of receiving the ESA deliverables ASAP, and that the ”last delivery” of several of the deliverables in table 23 will have to be adjusted to an earlier date: the preparation of the simulated data for the end-to-end testing begins 1.5 years before the tests begin.

Table 23: Critical Deliverables

Provider	Deliverable	Type	Customer	Delivery times	First delivery	Last delivery
ESA:	Initial calibration data (laboratory data)	data	CU2,3,5,6	T9,10	T9	T10
	Instrument design specifications	documents	CU2,3,5,6	TN-5mo	T2	T7
	Telemetry and MOC data format	document	CU1,2,3	TN	T2	T10
CU1:	GaiaTools, Gaia Parameter DB, MDB dictionary and access tools	SW	CU2–8	TN	T2	T10
	MDB ICD and updates	document	CU2–8	TN	T3	T6
CU2:	GOG:Simulated intermediate data	data	CU3–8	TN	T3	EoM
	GOG:Simulated true sky, attitude, instrument, noise-free obs.	data	CU3,5,6	TN	T2	EoM
	GASS:Simulated telemetry	data	CU3	TN	T2	T10
	GOG:simulated end-of-mission data	data	CU8	TN	T1	EoM
CU3-8:	Error models for DP products	model	CU2	TN	T2	T9
	Data Structures and Descriptions for MDB	document	CU1	TN	T2	T5
	Test Data according to MDB ICD	data	CU1	TN	T4	T10
CU3:	REMAT:Simulated Gaia & Solar System ephemeris	data	CU1	TN	T2	T10
	Reconstructed attitude interpretation	SW	CU5	TN	T3	T9
	Details of astrometric noise model	model	CU8	TN	T3	EoM
CU4:	Eclipsing binary light curve synthesis	model	CU2,7	TN	T3	T8
	Simulation specs for specific types of objects	document	CU2	TN	T2	T10
CU5:	CCD,PSF, photom calibration SW for IDT/FL	SW	CU3	T3,5,7,8	T3	T9
	Calibration standards	data	CU3	T3,6,9	T3	T9
	Details of BP/RP spectral combination	document	CU8	TN	T7	EoM
	Details of BP/RP noise model	model	CU8	TN	T3	EoM
CU6:	RVS IDT SW	SW	CU3	T3,5,7,9	T3	T9
	Calibration standards	data	CU3	T3,6,9	T3	T9
	Details of RVS spectral combination	document	CU8	TN	T7	EoM
	Details of RVS noise model	model	CU8	TN	T3	EoM
CU7:	Details of BP/RP variability metrics	document	CU8	TN	T4	EoM
	Simulation specs for specific types of objects	document	CU2	TN	T2	T10
CU8:	Final reference spectral library	data	CU6	T9	T9	T9
	Spectral libraries	data	CU2	TN	T2	T10
	Calibration standards	data	CU3	T3,6,9	T3	T9
	Simulation specs for specific types of objects	data	CU2	TN	T2	T10



ID	Task Name
1	**ESA**
2	provide final instrument design specifications
3	provide final telemetry and MOC data format
4	Initial (laboratory) calibration data
5	
6	**CU1**
7	Develop GaiaTools, Gaia Parameter DB
8	Provide final GaiaTools, Gaia Parameter DB
9	verification of end-to-end test of DPAC system
10	**CU2**
11	Develop GASS, GOG, Instrument model
12	Provide final version of Instrument model,GASS,GOG
13	GOG: Simulate true sky, attitude, instrument, noise-free obs.
14	GASS: provide simulated telemetry for testing
15	
16	**CU3**
17	REMAT:Simulated Gaia and Solar System ephemeris
18	end-to-end test of IDT/FL
19	Develop IDT and FL
20	Provide final IDT and FL SW
21	Provide final Initial Source Catalogue
22	**CU5**
23	Develop CCD,PSF, photom calibration SW
24	provide final CCD,PSF, photom calibration SW for IDT/FL
25	provide photometric calibration standards
26	provide BP/RP noise model
27	**CU6**
28	Develop RVS IDT SW
29	provide final RVS IDT SW
30	provide spectroscopic calibration standards
31	provide RVS noise model
32	**CU8**
33	provide spectral libraries
34	provide classification calibration standards

Project: MSProj11
Date: Tue 4/3/07

Page 1

Figure 70: Schedule for IDT/FL final end-to-end testing.

Part IV

Appendices

A DPAC Effort Required

Note the numbers in the following tables have been rounded.

Table 24: Total effort (staff months) required for each CU per year.

	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	**Total**
Total DPACE	12	38	38	38	38	38	38	38	38	38	38	38	38	**468**
Total CU1	110	130	128	134	124	100	65	71	68	72	74	61	51	**1186**
Total CU2	280	268	200	187	159	64	30	25	15	15	14	14	14	**1283**
Total CU3	434	458	473	557	538	468	384	374	288	288	195	194	194	**4845**
Total CU4	155	154	147	137	145	80	29	18	18	17	25	24	24	**972**
Total CU5	303	311	316	321	316	288	243	223	238	193	186	160	127	**3225**
Total CU6	152	198	207	204	200	120	64	58	58	57	57	56	55	**1486**
Total CU7	166	188	187	184	174	124	126	121	111	102	99	98	97	**1777**
Total CU8	220	278	288	288	283	132	174	146	160	159	159	167	125	**2579**
Total DPC-B	15	25	25	35	35	45	35	35	35	35	35	35	45	**435**
Total DPC-C	97	158	158	144	104	86	71	61	60	51	42	35	28	**1095**
Total DPC-E	15	16	23	34	45	47	31	31	31	33	34	32	32	**402**
Total DPC-G	18	20	15	12	11	17	26	22	16	16	16	13	13	**212**
Total DPC-I	0	0	0	0	8	60	60	60	60	60	60	60	60	**488**
Total DPC-T	16	16	16	29	29	27	19	19	19	19	19	19	19	**266**
Total DPAC	**1991**	**2256**	**2222**	**2303**	**2208**	**1694**	**1394**	**1302**	**1215**	**1154**	**1052**	**1005**	**921**	20718
Total ESA														1513
Total Community														**19205**

Table 25: Effort (staff months) required per work package for DPACE.

WP Number	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	**Total**
GWP-M-001	12	18	18	18	18	18	18	18	18	18	18	18	18	228
GWP-M-010	0	10	10	10	10	10	10	10	10	10	10	10	10	120
GWP-M-018	0	10	10	10	10	10	10	10	10	10	10	10	10	120
Total DPACE	**12**	**38**	**38**	**38**	**38**	**38**	**38**	**38**	**38**	**38**	**38**	**38**	**38**	**468**

Table 26: Effort (staff months) required per work package for CU1.

WP Number	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	**Total**
GWP-M-101	9	12	12	15	12	12	11	11	11	11	14	14	12	154
GWP-T-102	14	19	17	12	7	7	6	6	6	6	5	5	5	113
GWP-T-103	20	19	21	20	20	12	10	10	10	10	18	12	4	181
GWP-M-104	4	6	6	4	4	2	2	2	2	2	1	1	1	31
GWP-T-110	6	9	10	10	11	9	8	12	12	12	8	7	7	116
GWP-T-140	14	15	11	16	11	11	0	3	0	3	4	0	0	87
GWP-T-150	7	9	10	11	14	9	5	5	5	5	0	0	0	79
GWP-T-160	4	2	2	3	2	3	1	1	1	1	0	0	0	17
GWP-T-170	1	2	3	4	6	8	8	8	8	8	0	0	0	56
GWP-T-180	27	30	31	33	32	26	16	16	15	16	24	22	22	309
GWP-T-190	7	8	7	7	6	4	1	1	1	1	1	1	1	46
Total CU1	**110**	**130**	**128**	**134**	**124**	**100**	**65**	**71**	**68**	**72**	**74**	**61**	**51**	**1186**

Table 27: Effort (staff months) required per work package for CU2.

WP Number	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	**Total**
GWP-M-201	15	15	15	15	15	2	2	2	2	2	1	1	1	88
GWP-M-202	10	10	6	4	2	1	0	0	0	0	0	0	0	32
GWP-M-203	3	3	2	2	2	1	0	0	0	0	0	0	0	12
GWP-M-220	67	67	50	50	40	1	1	1	1	1	1	1	1	282
GWP-M-230	60	60	30	20	30	40	20	15	5	5	5	5	5	300
GWP-M-240	29	26	25	20	15	5	2	2	2	2	2	2	2	134
GWP-M-250	41	40	26	26	15	8	2	2	2	2	2	2	2	170
GWP-M-260	40	27	26	25	15	5	2	2	2	2	2	2	2	152
GWP-S-270	15	20	20	25	25	1	1	1	1	1	1	1	1	113
Total CU2	**280**	**268**	**200**	**187**	**159**	**64**	**30**	**25**	**15**	**15**	**14**	**14**	**14**	**1283**

Table 28: Effort (staff months) required per work package for CU3.

WP Number	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	**Total**
GWP-M-301	2	2	2	2	2	2	2	2	2	2	2	2	2	26
GWP-T-302	1	1	1	1	1	1	1	1	1	1	1	1	1	13
GWP-T-303	1	1	1	1	1	1	1	1	1	1	1	1	1	13
GWP-T-304	2	2	3	12	12	12	6	2	2	2	2	2	2	61

`GWP-M-320`	55	49	51	51	49	46	42	36	36	36	31	30	30	541
`GWP-M-330`	60	60	60	72	72	60	48	48	36	36	36	36	36	660
`GWP-M-335`	30	40	40	48	48	6	6	6	6	6	6	6	6	254
`GWP-M-340`	54	54	54	41	41	41	51	51	51	51	51	51	51	642
`GWP-M-345`	36	48	60	60	60	60	60	60	60	60	60	60	60	744
`GWP-M-350`	108	116	115	171	157	109	109	109	59	60	4	4	4	1125
`GWP-D-360`	84	83	83	95	93	128	56	56	32	32	0	0	0	742
`GWP-D-370`	1	2	3	3	3	2	2	2	2	1	1	1	1	24
Total CU3	**434**	**458**	**473**	**557**	**538**	**468**	**384**	**374**	**288**	**288**	**195**	**194**	**194**	**4845**

Table 29: Effort (staff months) required per work package for CU4.

WP Number	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	**Total**
`GWP-M-401`	0	0	0	0	0	0	0	0	0	0	0	0	0	0
`GWP-T-402`	10	10	10	10	8	4	0	0	0	0	0	0	0	52
`GWP-T-403`	3	5	8	9	7	4	0	0	0	0	0	0	0	35
`GWP-M-404`	1	1	1	2	10	7	0	0	0	0	0	0	0	21
`GWP-T-405`	14	16	18	21	25	15	0	0	0	0	0	0	0	108
`GWP-M-430`	2	2	3	2	2	2	1	1	1	1	1	1	1	20
`GWP-M-432`	12	12	12	12	12	11	11	0	0	0	0	0	0	82
`GWP-M-433`	2	2	2	4	10	10	10	10	10	10	10	9	9	98
`GWP-M-434`	2	2	2	2	2	2	0	0	0	0	0	0	0	12
`GWP-M-435`	1	1	1	1	1	1	0	0	0	0	0	0	0	6
`GWP-M-436`	25	25	17	1	1	1	1	1	1	1	1	1	1	77
`GWP-M-437`	22	22	22	22	22	2	2	2	2	2	10	10	10	150
`GWP-M-438`	3	3	2	2	2	0	0	0	0	0	0	0	0	12
`GWP-M-439`	0	0	0	2	2	0	0	0	0	0	0	0	2	6
`GWP-M-450`	2	2	2	2	2	2	1	1	1	1	1	1	1	19
`GWP-M-451`	1	1	1	1	1	0	0	0	0	0	0	0	0	6
`GWP-M-452`	1	1	1	1	1	0	0	0	0	0	0	0	0	6
`GWP-M-453`	6	6	6	6	6	0	0	0	0	0	0	0	0	30
`GWP-M-454`	2	2	2	2	2	2	0	0	0	0	0	0	0	9
`GWP-M-455`	2	2	2	2	2	0	0	0	0	0	0	0	0	8
`GWP-M-456`	19	10	10	8	7	2	0	0	0	0	0	0	0	56
`GWP-M-457`	12	13	9	9	5	3	0	0	0	0	0	0	0	50
`GWP-M-458`	8	8	8	8	8	7	0	0	0	0	0	0	0	47
`GWP-M-459`	0	3	3	3	3	3	3	3	3	2	2	2	0	30
`GWP-M-460`	3	3	3	3	3	0	0	0	0	0	0	0	0	15
`GWP-M-470`	3	3	3	3	3	2	0	0	0	0	0	0	0	17
Total CU4	**155**	**154**	**147**	**137**	**145**	**80**	**29**	**18**	**18**	**17**	**25**	**24**	**24**	**972**

Table 30: Effort (staff months) required per work package for CU5.

WP Number	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	**Total**
`GWP-M-501`	5	5	5	5	5	5	5	5	5	5	5	5	5	65
`GWP-T-502`	2	2	2	2	2	2	2	2	2	2	2	1	1	24
`GWP-T-503`	5	5	5	5	5	2	2	2	2	2	2	2	2	41
`GWP-M-504`	5	5	10	15	15	15	15	10	10	5	5	5	2	117
`GWP-M-505`	4	4	4	4	4	4	4	4	4	4	2	2	2	46
`GWP-M-510`	20	20	20	20	20	20	15	15	10	10	10	5	5	190
`GWP-M-511`	12	20	20	20	20	10	10	15	20	20	25	20	20	232
`GWP-M-512`	60	60	60	60	60	60	60	40	40	20	20	10	10	560
`GWP-M-513`	50	50	50	50	50	50	30	20	20	10	10	10	10	410
`GWP-M-514`	20	20	20	20	20	10	10	15	20	20	20	25	15	235
`GWP-M-515`	50	50	50	50	50	50	40	40	40	40	30	25	20	535
`GWP-M-516`	15	15	15	15	15	15	10	10	10	5	5	5	0	135
`GWP-M-517`	15	15	15	15	15	10	10	10	10	10	10	5	5	145
`GWP-M-518`	20	20	20	20	15	15	15	20	30	30	30	25	20	280
`GWP-M-519`	20	20	20	20	20	20	15	15	15	10	10	15	10	210
Total CU5	**303**	**311**	**316**	**321**	**316**	**288**	**243**	**223**	**238**	**193**	**186**	**160**	**127**	**3225**

Table 31: Effort (staff months) required per work package for CU6.

WP Number	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	**Total**
`GWP-M-601`	10	10	10	10	10	10	10	10	10	10	10	10	10	130
`GWP-M-602`	9	9	9	8	6	2	0	0	0	0	0	0	0	43
`GWP-M-603`	3	5	8	9	7	4	0	0	0	0	0	0	0	36
`GWP-M-604`	8	11	13	15	30	18	10	10	10	10	10	10	10	165
`GWP-M-610`	16	31	35	30	15	11	10	9	9	8	8	7	6	195
`GWP-M-620`	20	28	28	28	28	14	5	5	5	5	5	5	5	181
`GWP-M-630`	32	32	32	32	32	20	12	7	7	7	7	7	7	234
`GWP-M-640`	13	13	13	13	13	12	7	7	7	7	7	7	7	126
`GWP-S-650`	37	37	37	37	37	18	5	5	5	5	5	5	5	238
`GWP-S-660`	4	22	22	22	22	11	5	5	5	5	5	5	5	138
Total CU6	**152**	**198**	**207**	**204**	**200**	**120**	**64**	**58**	**58**	**57**	**57**	**56**	**55**	**1486**

Table 32: Effort (staff months) required per work package for CU7.

WP Number	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	**Total**
`GWP-M-701`	20	18	17	17	16	12	13	16	16	15	15	15	15	203
`GWP-T-702`	2	2	1	1	1	1	1	0	0	0	1	0	0	7
`GWP-T-703`	3	3	4	6	7	7	7	6	6	6	6	6	6	69
`GWP-M-704`	1	4	8	8	8	8	11	8	5	5	5	5	5	80
`GWP-M-710`	18	21	21	21	21	14	15	15	15	14	12	12	12	211
`GWP-M-711`	16	19	19	19	19	15	12	12	12	12	12	12	12	191
`GWP-M-712`	40	44	44	44	44	35	35	30	25	20	20	20	20	421
`GWP-M-720`	50	50	45	40	30	10	10	10	10	10	10	10	10	295
`GWP-M-721`	13	16	16	16	16	13	13	13	10	8	8	8	8	158
`GWP-M-730`	0	3	3	3	3	2	3	3	3	3	2	2	2	29
`GWP-M-731`	4	4	4	4	4	3	3	3	4	4	4	4	3	48
`GWP-M-732`	1	6	6	6	6	5	5	5	5	5	5	5	5	65
Total CU7	**166**	**188**	**187**	**184**	**174**	**124**	**126**	**121**	**111**	**102**	**99**	**98**	**97**	**1777**

Table 33: Effort (staff months) required per work package for CU8.

WP Number	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	**Total**
`GWP-M-801`	4	4	4	4	4	4	3	3	3	3	3	3	3	45
`GWP-T-802`	10	9	9	8	6	2	0	0	0	0	0	0	0	43
`GWP-T-803`	3	5	8	9	7	4	0	0	0	0	0	0	0	35
`GWP-T-804`	5	8	13	17	32	18	10	10	10	10	10	10	10	163
`GWP-T-805`	16	32	35	30	15	10	10	9	9	8	8	7	6	194
`GWP-T-806`	3	3	2	2	1	1	2	2	1	1	1	0	0	19
`GWP-S-811`	30	44	44	44	44	22	22	22	22	22	22	11	0	349
`GWP-S-812`	6	6	6	6	6	6	6	6	6	6	6	6	0	72
`GWP-S-821`	15	16	16	16	16	8	16	10	10	10	10	16	11	170
`GWP-S-822`	16	15	16	16	17	10	16	10	11	11	11	16	11	176
`GWP-S-823`	16	16	16	16	16	10	16	10	10	10	10	16	10	172
`GWP-S-824`	11	11	11	11	11	6	10	6	10	6	6	10	6	115
`GWP-S-825`	11	11	11	11	11	6	6	6	11	11	11	11	11	128
`GWP-S-831`	16	16	16	16	16	3	9	6	6	6	6	6	6	128
`GWP-S-832`	16	16	16	16	16	3	9	6	6	6	6	6	6	128
`GWP-S-833`	11	11	11	11	11	3	9	6	6	6	6	6	6	103
`GWP-S-834`	6	11	11	11	11	3	9	6	6	10	10	10	6	110
`GWP-S-835`	20	33	33	33	33	10	10	22	22	22	22	22	22	304
`GWP-S-836`	6	11	11	11	11	4	11	6	11	11	11	11	11	126
Total CU8	**220**	**278**	**288**	**288**	**283**	**132**	**174**	**146**	**160**	**159**	**159**	**167**	**125**	**2579**

Table 34: Effort (staff months) required per work package for DPC-B.

WP Number	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	**Total**
`GWP-T-B01`	5	5	5	5	5	5	5	5	5	5	5	5	5	65
`GWP-O-B10`	10	20	20	30	30	40	30	30	30	30	30	30	40	370
Total DPC-B	**15**	**25**	**25**	**35**	**35**	**45**	**35**	**35**	**35**	**35**	**35**	**35**	**45**	**435**

Table 35: Effort (staff months) required per work package for DPC-C.

WP Number	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	**Total**
`GWP-M-C01`	8	8	8	8	8	8	10	10	7	5	2	2	3	87
`GWP-T-C60`	89	150	150	136	96	78	0	0	0	0	0	0	0	699
`GWP-O-C10`	0	0	0	0	0	0	61	51	53	46	40	33	25	309
Total DPC-C	**97**	**158**	**158**	**144**	**104**	**86**	**71**	**61**	**60**	**51**	**42**	**35**	**28**	**1095**

Table 36: Effort (staff months) required per work package for DPC-E.

WP Number	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	**Total**
`GWP-T-E01`	3	3	3	5	7	7	4	4	4	4	5	5	5	56
`GWP-T-E02`	11	12	18	23	23	23	16	16	16	18	16	16	16	221
`GWP-O-E10`	1	1	2	6	15	17	12	12	12	12	14	12	12	126
Total DPC-E	**15**	**16**	**23**	**34**	**45**	**47**	**31**	**31**	**31**	**33**	**34**	**32**	**32**	**402**

Table 37: Effort (staff months) required per work package for DPC-G.

WP Number	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	**Total**
`GWP-M-G01`	0	2	3	3	4	5	5	3	3	3	3	3	3	40
`GWP-T-G02`	1	1	2	2	2	2	2	1	1	1	1	1	1	17
`GWP-M-G05`	17	17	11	8	5	5	4	3	2	2	2	1	1	75
`GWP-O-G10`	0	0	0	0	0	5	15	15	10	10	10	8	8	81
Total DPC-G	**18**	**20**	**15**	**12**	**11**	**17**	**26**	**22**	**16**	**16**	**16**	**13**	**13**	**212**

Table 38: Effort (staff months) required per work package for DPC-I.

WP Number	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	**Total**
`GWP-M-I01`	0	0	0	0	8	10	10	10	10	10	10	10	10	88
`GWP-O-I10`	0	0	0	0	0	50	50	50	50	50	50	50	50	400
Total DPC-I	**0**	**0**	**0**	**0**	**8**	**60**	**60**	**60**	**60**	**60**	**60**	**60**	**60**	**488**

Table 39: Effort (staff months) required per work package for DPC-T.

WP Number	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	**Total**
`GWP-T-T01`	3	3	3	5	5	3	3	3	3	3	3	3	3	43
`GWP-T-T02`	11	11	11	17	17	17	9	9	9	9	9	9	9	147
`GWP-O-T10`	2	2	2	7	7	7	7	7	7	7	7	7	7	76
Total DPC-T	**16**	**16**	**16**	**29**	**29**	**27**	**19**	**19**	**19**	**19**	**19**	**19**	**19**	**266**

B Summary of top-level work packages

WP Number	Description	Manager
GWP-M-001-00000	Coordination of the DPAC activities	F. Mignard, R. Drimmel
GWP-T-010-00000	Project Control of DPAC Activities	R. Drimmel (DPACE Deputy Chair)
GWP-T-020-00000	DPAC-ESA technical interfacing.	F. Mignard
GWP-M-101-00000	Management and scientific coordination of CU1	W. O'Mullane
GWP-T-102-00000	Architecture and technical coordination CU1	W. O'Mullane
GWP-T-103-00000	Quality assurance and config management for CU1	T. Levoir
GWP-M-104-00000	Integration, Validation and Operation of CU1 systems	J. Hoar
GWP-T-110-00000	Coordination common software resources	F. de Angeli
GWP-T-140-00000	Technology Trend Monitoring	W. O'Mullane
GWP-T-150-00000	End-to-end system testing	J. Hoar
GWP-T-160-00000	Host Framework	X. Passot
GWP-T-170-00000	Payload Managment	J. Hoar
GWP-T-180-00000	Main Database Design/Code/Operate	J. Hernandez
GWP-T-190-00000	Gaia Transfer System Design/Code/Operate	J. Hernandez
GWP-M-201-00000	Management and scientific coordination of CU2	X. Luri, C. Babusiaux, F. Mignard
GWP-T-202-00000	Architecture and technical coordination of CU2	J.M. Wallut
GWP-T-203-00000	Configuration Management and quality assurance for CU2	J.M. Wallut
GWP-M-220-0000	Management and implementation of the Universe Model	A. Robin, C. Reylé
GWP-M-230-00000	Management & implementation of the Instrument Model	M. Gai, J. Rebordão
GWP-M-240-00000	Management and implementation of GASS	E. Masana
GWP-M-250-0000	Management and implementation of GIBIS	C. Babusiaux
GWP-M-260-00000	Management and implementation of GOG	X. Luri
GWP-S-270-0000	Scientific Quality Assurance and Validation of CU2 simulations	D. Egret
GWP-M-301-00000	Management and scientific coordination of CU3	U. Bastian; M. G. Lattanzi, J. Torra
GWP-T-302-00000	Architecture and technical coordination CU3	W. O'Mullane
GWP-T-303-00000	Configuration Management and quality assurance for CU3	J. Hoar
GWP-M-304-00000	Integration, Validation and Operation of CU3 systems	W. O'Mullane
GWP-M-320-00000	Management and implementation of the AGIS (framework and algorithms)	L. Lindegren, U. Lammers
GWP-M-330-00000	REMAT: Relativistic Models and Tests	S. Klioner
GWP-M-335-00000	Auxiliary data: definition and acquisition	U. Bastian
GWP-M-340-00000	Astrometric Verification Unit (AVU), BAM and WFS processing	M. G. Lattanzi
GWP-M-345-00000	IDU: coordination, framework, modules and operations	J. Torra
GWP-M-350-00000	IDT: Management, Implementation, Operation	J. Torra, J. Hoar
GWP-M-360-0000	FL: management, implementation, operation	S. Jordan, J. Hoar
GWP-M-370-00000	Manage and Implement IDT/FL Database	J. Hoar, J. Hernandez
GWP-M-401-00000	Management and scientific coordination	D. Pourbaix
GWP-T-402-00000	Architecture and technical coordination CU4	T. Levoir
GWP-T-403-00000	Quality assurance and config management for CU4	T. Levoir
GWP-T-404-00000	Integration, Validation and Operation of CU4 system	T. Levoir
GWP-T-405-00000	CU4 Host Software Framework development	T. Levoir
GWP-M-430-00000	Management and implementation of Non Single Star processing	D. Pourbaix
GWP-M-432-00000	Astrometric Binaries	J.-L. Halbwachs
GWP-M-433-00000	Resolved Multiples	D. Pourbaix
GWP-M-434-00000	Spectroscopic Binaries	E. Gosset
GWP-M-435-00000	Photometric Analysis of NSS	O. Malkov
GWP-M-436-00000	Eclipsing Systems	C. Siopis
GWP-M-437-00000	Extrasolar Planets	A. Sozzetti
GWP-M-438-00000	Simulated Test data for NSS processing	F. Arenou
GWP-M-439-00000	NSS solution combiner	D. Pourbaix
GWP-M-450-0000	Management and implementation of Solar System Object processing	P. Tanga
GWP-M-451-00000	Auxiliary data for SSO processing	J. Berthier
GWP-M-452-00000	Solar System objects cross matching	F. Mignard
GWP-M-453-00000	CCD processing for SSO observations	A. Dell'Oro
GWP-M-454-00000	Astrometric Reduction for SSO	J.-E. Arlot

GWP-M-455-00000	Threading of SSO	J.-M. Petit
GWP-M-456-00000	Orbital inversion for SSO	K. Muinonen
GWP-M-457-00000	Global Effects on Solar System Dynamics	D. Hestroffer
GWP-M-458-00000	SSO physical parameters	A. Cellino
GWP-M-459-00000	Ground based observations	W. Thuillot
GWP-M-460-00000	Simulated Test Data for SSO processing	F. Mignard
GWP-D-470-000000	Extended Objects	C. Ducourant
GWP-M-501-00000	Planning, management, and coordination of CU5 activities	F. van Leeuwen
GWP-T-502-00000	Architecture and technical coordination of CU5	P. S. Bunclark
GWP-T-503-00000	Quality assurance and configuration management for CU5	P. J. Richards
GWP-T-504-00000	Integration, validation and operation of CU5	F. De Angeli
GWP-T-505-00000	Technical support	F. De Angeli
GWP-T-510-00000	PSF and LSF calibration	M. A. Barstow
GWP-M-511-00000	BP/RP flux extraction and initial data treatment	A. Brown
GWP-T-512-00000	Photometric calibration models for G and BP/RP	C. Jordi
GWP-T-513-00000	Instrument absolute response characterisation: ground-based preparation	E. Pancino
GWP-T-514-00000	Instrument absolute response characterisation: definition and application	C. Cacciari
GWP-T-515-00000	Internal photometric calibration and its application	D. W. Evans
GWP-M-516-00000	Selection of internal reference sources	C. Jordi
GWP-T-517-00000	Flux and classification-based science alerts	S. Hodgkin
GWP-M-518-00000	2D image restoration	A. Brown
GWP-T-519-00000	Photometric data base and archive	N. Hambly
GWP-M-601-00000	Management and Scientific Coordination of CU6	D. Katz
GWP-M-602-00000	Architecture and Technical Coordination of CU6	A. JeanAntoine
GWP-M-603-00000	Quality Assurance and Configuration Management for CU6	T. Levoir
GWP-M-604-00000	Integration, Validation and Operation of CU6 System	A. JeanAntoine
GWP-T-610-00000	CU6 Host Software Framework development	A. JeanAntoine
GWP-S-620-00000	Spectra extraction	S. Rosen
GWP-S-630-00000	Calibration of the spectroscopic instrument	S. Rosen
GWP-S-640-00000	Radial Velocity zero point	G. Jasniewicz
GWP-S-650-00000	Single transit analysis	Y. Viala
GWP-S-660-00000	Multiple transits analysis	S. Rosen
GWP-M-701-00000	Management and Scientific Coordination of CU7	L. Eyer, D.W. Evans, P. Dubath
GWP-M-702-00000	Architecture and Technical Coordination of CU7	M. Beck
GWP-M-703-00000	Quality Assurance and Configuration Control for CU7	I. Lecoeur
GWP-M-704-00000	Integration, Validation and Operation of the CU7 System	M. Beck
GWP-M-705-00000	Host Framework for CU7	M. Beck
GWP-M-710-00000	Special Variability Detection & Analysis	A. Lanzafame
GWP-M-711-00000	Variability Characterisation	J. Cuypers
GWP-M-712-00000	Classification	C. Aerts
GWP-M-720-00000	Specific Object Studies	N. Mowlavi
GWP-M-721-00000	Global Variability Studies	L. Sarro
GWP-M-731-00000	Analysis of Impact on Astrometry	A. Jorissen
GWP-M-732-00000	Supplementary Observations	G. Clementini
GWP-M-801-00000	Management and scientific coordination	C.A.L. Bailer-Jones
GWP-T-802-00000	Architecture and technical coordination CU8	A.M. Janotto
GWP-T-803-00000	Quality assurance and config management for CU8	T. Levoir
GWP-T-804-00000	Integration, Validation and Operation of CU8 system	A.M. Janotto
GWP-T-805-00000	CU8 Host Software Framework development	A.M. Janotto
GWP-T-806-00000	Data model and utility library	C. Tiede
GWP-S-811-00000	Provide training and testing data	F. Thévenin
GWP-S-812-00000	Interstellar extinction	R. Drimmel
GWP-S-821-00000	Discrete Source Classifier	C.A.L. Bailer-Jones
GWP-S-822-00000	Generalized Stellar Parametrizer - Photometry (GSP-phot)	C.A.L. Bailer-Jones
GWP-S-823-00000	Generalized Stellar Parametrizer - Spectroscopy (GSP-spec)	A. Recio-Blanco
GWP-S-824-00000	Object Clustering Analysis	L.M. Sarro
GWP-S-825-00000	Final Luminosity, Age and Mass Estimator (FLAME)	Y. Lebreton
GWP-S-831-00000	Quasar classifier	J.-F. Claeskens
GWP-S-832-00000	Unresolved galaxy classifier	M. Kontizas
GWP-S-833-00000	Solar system object classifier	C.-I. Lagerkvist
GWP-S-834-00000	Multiple Star Classifier (MSC)	C.A.L. Bailer-Jones

`GWP-S-835-00000`	Extended Stellar Parametrizer (ESP)	Y. Frémat
`GWP-S-836-00000`	Outlier Analysis	M. Manteiga
`GWP-M-B01-00000`	Management of Gaia at the BPC	S.Girona
`GWP-O-B10-00000`	Software deployement and operation at BPC	S. Girona
`GWP-M-C01-00000`	Management of the CNES DPC	X.Passot
`GWP-O-C10-00000`	Operation of DPC systems	X. Passot
`GWP-T-C60-00000`	CNES DPC Host Framework development	X. Passot
`GWP-M-E01-00000`	Management of the ESAC DPC	W. O'Mullane
`GWP-O-E10-00000`	Operation of ESAC (DPC-E) systems	W. O'Mullane
`GWP-T-E02-00000`	Architecture and technical coordination DPC-E	W. O'Mullane
`GWP-M-G01-00000`	Management of DPG	L. Eyer, M. Beck
`GWP-O-G10-00000`	Operation of the DPC-G systems	M. Beck
`GWP-T-G02-00000`	Architecture and Technical Coordination of DPG	M. Beck
`GWP-T-G05-00000`	Host Framework for CU7	M. Beck
`GWP-M-I01-00000`	Management of the IoA DPC	F. van Leeuwen
`GWP-O-I10-00000`	Operation of DPE systems	F. van Leeuwen
`GWP-M-T01-00000`	Management of the INAF-OATo DPC	A. Volpicelli
`GWP-O-T10-00000`	Operation of INAF-OATo systems	R.Morbidelli
`GWP-T-T02-00000`	Architecture and technical coordination INAF-OATo DPC	A. Volpicelli

C Detailed top-level work packages descriptions

Note the project phases for DPAC are defined in Sect. 7.5.1.

C.1 Top-level Work Packages of DPACE

Gaia DPAC WP:		GWP–M-001-00000
Title: Coordination of the DPAC activities		
Provider(s): F. Mignard, R. Drimmel,XX,YY		
Manager(s): F. Mignard, R. Drimmel		
Start: Phase B	**End**: Phase F	**Total Effort**: 228 MM
Objective: Overall coordination of the activities of the DPAC, of the DPACE and management the Management Support Team.		

Tasks:
The responsibilities of the DPAC Chair are listed in 10.4.2. Consistent with these, the tasks of the Chair are:

1. Coordinates and monitors the overall activities of the DPAC
2. Prepare, convene, organise and chair the meetings of the DPACE
3. Implement the decisions of the DPACE
4. Receives and reviews CU and DPC status reports
5. Monitor the master schedule of the DPAC activities
6. Monitor the critical interfaces between CUs
7. Interface with ESA (esp. PS and PM), managing all deliveries/requests between DPAC and ESA
8. Manages and prepares for the ESA reviews of the DPAC activities
9. Initiate and organise the plenary meetings of the DPAC
10. Maintain the current and historical statistics of the membership
11. Prepare reports for the DPACE, the GST and the Steering Committee
12. Supervise the preparation of the intermediate data release

In addition, the Chair directs the work of the Management Support Team, which includes the Deputy Chair, the Project Controller (GWP-010) and the System Engineer (GWP-020).
The responsibilities of the DPAC Deputy Chair are listed in 10.4.2. Consistent with these, the tasks of the Deputy Chair are:

1. Prepare the minutes of the meetings for Livelink
2. Maintains the DPACE wiki pages.
3. Assists the Chair in fullfilling his/her duties, listed above.

Input:
Reports from GWP-010 and GWP-020

Output:

Deliverables: Minutes of the DPACE, status reports of the DPAC, reports for the GST and steering committee, documents for the DPAC reviews
Dependencies:
Interfaces: CU leaders, Gaia Project Scientist and his support team, GST, Funding agencies representatives
Remarks: The *XX* name refers to the assistant to the chair that will be appointed by CNRS in 2007, while YY will be a support person to the deputy chair, provided by ASI funding.

Gaia DPAC WP:		GWP–T-010-00000
Title: Project Control of DPAC Activities		
Provider(s): F. Mignard, R. Drimmel,Project Controller		
Manager(s): R. Drimmel (DPACE Deputy Chair)		
Start: Phase B	**End**: Phase F	**Total Effort**: 120 MM
Objective: To monitor the overall progress (master schedule) of the DPAC activities. Reports directly to the DPAC Chair.		
Tasks: 1. Establish and maintain the Master Schedule of the DPAC; 2. Monitor the critical deliveries between CUs; 3. Monitor the achievment of CU and DPAC milestones; 4. Monitor the critical deliveries received and delivered to/from ESA; 5. Analysis of schedules and resource deployment to assess feasibility and determine critical areas requiring special attention; 6. Liason with CU and DPC managers, as needed; 7. Prepares regular DPAC progress reports; 8. Prepares impact analysis on DPAC master schedule of any late deliveries; 9. Issue warnings if critical deliveries are delayed or if master schedule risks being compromised; 10. Support the DPAC Chair in preparing reports.		
Input: CU and DPC Status and Progress reports		
Output:		
Deliverables: Progress Reports		

Dependencies:
Interfaces: DPAC Chair and Deputy Chair, DPAC System Engineer
Remarks: The Deputy Chair assumes the responsibility of monitoring the master schedule until the position of Project Controller is filled. Appointment should be by 2008. Person filling this position should have schedule engineering and/or project management experience, and be familiar with MS Project.

Gaia DPAC WP:		GWP–T-020-00000
Title: DPAC-ESA technical interfacing.		
Provider(s): F. Mignard, R. Drimmel,System Engineer		
Manager(s): F. Mignard		
Start: Phase B	**End**: Phase F	**Total Effort**: 120 MM

Objective:
To handle technical interfacing between the DPAC and ESA. Reports to the DPAC Chair.

Tasks:
Manage the technical interface between DPAC and ESA, excluding the interface between ESAC (SOC) and MOC. Interface between ESAC and CUs (eg. MDB data transfers, etc) are considered to be internal to DPAC and therefore not handled by the DPAC System Engineer.

1. Establish and maintain any ICDs defining deliveries between DPAC and Gaia Project Team or PS Support Team.
2. Formulate technical requests received from CUs and forward to ESA interface.
3. Receive technical requests from ESA and forward, as directed by DPAC Chair, to appropriate provider.
4. Manage technical deliverables received from or delivered to ESA.
5. Support to scheduling and reporting.
6. Maintain top-level technical documentation, where appropriate.

Input:
Requests/Deliveries from CU managers.

Output:

Deliverables:

Dependencies:

Interfaces: DPAC Chair and Deputy Chair, DPAC Project Controller
Remarks: The DPAC Chair assumes these repsonsibilities until the position of DPAC System Engineer is filled. Appointment should be in 2008. Person filling this position should have technical engineering experience. Previous experience with Gaia highly desirable.

C.2 Top-level Work Packages of CU1

Gaia DPAC WP:	GWP–M-101-00000

Title: Management and scientific coordination of CU1		
Provider(s): W. O'Mullane, U. Lammers, T. Levoir		
Manager(s): W. O'Mullane		
Start: Phase B	**End**: Phase F	**Total Effort**: 219 MM
Objective: Sect. 7.6.3.1		
Tasks: Tasks and sub tasks of GWP-M-101-00000 As in Sect. 7.6.3.1.		
Input:		
Output:		
Deliverables:		
Dependencies:		
Interfaces:		
Remarks: We assume to have a reasonable overhead in ESA for reporting etc. Also we assume the ESAC group will interact with all other groups i.e. a high communication overhead.		

Gaia DPAC WP:		GWP-T-102-00000
Title: Architecture and technical coordination CU1		
Provider(s): W. O'Mullane, J. Hoar, T. Levoir		
Manager(s): W. O'Mullane		
Start: Phase B	**End**: Phase F	**Total Effort**: 343 MM
Objective: Sect. 7.6.3.2 In addition for CU1 the overall System architecture needs to be defined it is the task of CU1 to coordinate all inputs to the Ground Segment Design Definition.		
Tasks: Tasks and sub tasks of GWP-T-102-00000 Sect. 7.6.3.2 Define Overall System Architecture for Processing (GSDD).		
Input:		
Output:		
Deliverables:		
Dependencies:		
Interfaces: This WP interfaces with all other CUs for the definition of the overall system architecture.		
Remarks:		

Gaia DPAC WP:	GWP-T-103-00000

Title: Quality assurance and config management for CU1		
Provider(s): T. Lock, T. Levoir, W. O'Mullane		
Manager(s): T. Levoir		
Start: Phase B	**End**: Phase F	**Total Effort**: 213 MM
Objective: Sect. 7.6.3.3.		
Tasks: Tasks and sub tasks of GWP-T-103-00000 Sect. 7.6.3.3.		
Input: Sect. 7.6.3.3.		
Output:		
Deliverables: Software Quality Assurance Report		
Dependencies: GWP-T-102-0000 - Management Plan		
Interfaces: This work package contributes to the overall QA activities of the DPAC which falls to CU1 e.g. production of [LJMD$^+$07] falls in this packages.		
Remarks:		

Gaia DPAC WP: GWP-M-104-00000

Title: Integration, Validation and Operation of CU1 systems

Provider(s): J. Hoar

Manager(s): J. Hoar

Start: Phase B	**End**: Phase F	**Total Effort**: 31 MM

Objective:
Sect. 7.6.3.4 This WP covers work performed by the Operations Team of the Gaia SOC

Tasks:
The tasks of this work package are:
Sect. 7.6.3.4
These tasks are described below and form distinct sub-work packages. The CU1 systems applicable to this work package are:

1. Main Database (MAINDB)
2. Gaia Transfer System (GTS)

The tasks in the work package are performed by the Operations Team at the SOC, led by the Science Operations Manager (SCOM). Specific operations workpagages under GWP-O-E10 cover operations ofthe ESAC systems in general. Most of the operator effort appears in that workpackage.
GaiaTools is not considered in this work package, since it is not operational system per se, rather a component of other operations systems. There is no integration activity to be performed and no acceptance testing is suitable, since testing at Unit level only is appropriate for the particular case of GaiaTools.

Tasks:
Integration and validation The CU1 DPC is ESAC.

Input:

Output:

Deliverables:

Sect. 7.6.3.4

Dependencies: This workpackage depends on the development work packages of CU1 software systems as part of their overall lifecycle
Interfaces: The operation of CU1 systems are core to the overall functioning of the Gaia Data Processing Ground Segment and represent the primary interface between the SOC and DPCs.
Remarks:

Gaia DPAC WP:		GWP-T-110-00000
Title: Coordination common software resources		
Provider(s): F. de Angeli, J. Hoar, F. Jocteur		
Manager(s): F. de Angeli		
Start: Phase B	**End**: Phase F	**Total Effort**: 143 MM
Objective: Provide a set of software tools for the deployment of the DPAC activities. This includes common software libraries and display facilities as well as programming support.		
Tasks: Tasks and sub tasks of GWP-T-110-00000 1. Common toolbox and libraries for Java 2. Gaia Parameter Database 3. Common display facilities 4. General DPAC assistance (a) Java06 course 5. Select and document environment and tools		
Input: Java coding standards and guidelines and requirements from all the CUs forming the DPAC		
Output: Common software resources		
Deliverables: Software library distributions with associated documentation		
Dependencies:		
Interfaces: Each CU will have at least one librarian with access to the common software.		

Remarks:

Gaia DPAC WP: GWP-T-140-00000

Title: Technology Trend Monitoring		
Provider(s): W. O'Mullane, T. Levoir, Portugal		
Manager(s): W. O'Mullane		
Start: Phase B	**End**: Phase F	**Total Effort**: 89 MM

Objective:
Keep on top of the latest trends in Hardware/Software. Occasionally do some testing and report to DPAC. Make recommendations to DPAC on hardware/software. We foresee in the next phase to test multiple database systems as well as Root, for I/O performance. In the future there will be other tests needed - this package sets aside time for such small testing projects or for the initialisation of study contracts.

Tasks:
Tasks and sub tasks of GWP-T-140-00000

- Processor Testing/ Monitoring
- Database Technology Testing/Monitoring
- Language Testing/ Monitoring
- Network Technology Testing/Monitoring

Input:
Performance requirements from any systems which have them in a good state.

Output:
Infrequent study reports on key technologies and techniques

Deliverables:

Dependencies:

Interfaces:

Remarks:
We could see this as finished next year and our decisions cast in stone. Rather it is probably better to remain ambivalent to changing hardware and software and maintain our current flexible stance to take advantage of new developments. Hence this package is seen extending right into final catalogue production. CNES are running studies in this area and will provide effort.

Gaia DPAC WP: GWP-T-150-00000

Title: End-to-end system testing

Provider(s): J. Hoar, T. Levoir

Manager(s): J. Hoar

Start: Phase B	**End**: Phase D	**Total Effort**: 82 MM

Objective:
Organise the large scale test with the DPAC. If the Dataflow holds up interaction should in fact be a matter of transferring files and checking the contents and formats. Yet if we are to try to integrate a new Main DB every six months or so the mechanisms for these transfers need to be very well understood and the Interface Control Documents(ICD) need to be very well adhered to. Each CU should have a package for participating in these tests. These tests need to be defined, planned and executed. The results need to be carefully examined and recommendations made.

Tasks:
Tasks and sub tasks of GWP-T-150-00000

1. Strategy for the overall system testing
2. Planning of the large scale tests
3. Execution and monitoring large scale tests results
4. Analysis of large scale tests results (Scientifically and Technically)

Input:
Test plans and readiness reports from all CUs

Output:
Test Plan, Test Reports

Deliverables:

Dependencies:

Interfaces:
All CUs

Remarks:
Remarks - free text

Gaia DPAC WP:		GWP-T-160-00000
Title: Host Framework		
Provider(s): CNES, ESAC		
Manager(s): X. Passot		
Start: Phase B	**End**: Phase B	**Total Effort**: 36 MM
Objective: Investigate the possibility of a common host framework for DPC usage. This include Scheduling, Data Access, and distribution. The AGIS system at ESAC has its own Framework built in for this with the RunManager, Whiteboard and DAL - some of this may be reused. The grid software may provide some possibilities for this also.		
Tasks: Tasks and sub tasks of GWP-T-160-00000 1. Gather Requirements 2. Initiate Study 3. Design System 4. Test System		
Input: AGIS Framework documentation and Common Toolbox.		
Output: Requirements, Study Report		
Deliverables: Host Framework Components in the Common Toolbox		
Dependencies:		
Interfaces: Potentially each CU software unit will interface to this software		

Remarks:
This is complex and needs to be approached with care. Use of Interfaces will be essential to give some independence from the frame work. For testing also it will be essential to have a simple version of the framework. CNES will have major input and effort in this area.

Gaia DPAC WP:		GWP-T-170-00000
Title: Payload Managment		
Provider(s): ESAC		
Manager(s): J. Hoar		
Start: Phase B	**End**: Phase E	**Total Effort**: 56 MM

Objective:
Define the science operations of Gaia and develop software to support such operations. This includes interaction with ESOC both for definition and later for operational reasons.

Tasks:
Tasks and sub tasks of GWP-T-170-00000

1. Definition of ESOC ICD(with ESOC) and Mission Operations Concept
2. Definition of Payload Managment Requirements
3. Implementation/Testing/Maintenance of Payload Managment System
4. Implementation/Testing/Maintenance of ESOC Interface
5. Implementation/Testing/Maintenance of Payload Management System
6. Science Operations, new calib upload etc.
7. Operations Interactions with ESOC
8. SVT Participation

Input:
MRD

Output:
Updates to time line and calibrations for MOC

Deliverables:
MOC/SOC IRD, ICD and Software, Mission Ops. Concept, Payload Ops. Docs/Software

Dependencies:

Interfaces: CU3 for DFL outputs and Ground Based Observations
Remarks: This will become clearer as we understand the uplink/calibration possibilities better

Gaia DPAC WP:		GWP-T-180-00000
Title: Main Database Design/Code/Operate		
Provider(s): J. Hernandez, M. ter Linden, others as needed		
Manager(s): J. Hernandez		
Start: Phase B	**End**: Phase F	**Total Effort**: 370 MM

Objective:

The purpose of the Main Database is to support the Gaia Data Processing by taking care of the data storage and management.
The MDB will be described by a data dictionary in an electronic format. Tool shall be made available to gather inputs from the CUs to be included in the dictionary. Furthermore the ICD and MDB schema's shall be electronically derived from the data dictionary.
The Main Database will hold several types of Data (Raw, Intermediate, Reduced Data), the data contents will be agreed by all the CUs based on what is the data they need for their subsystems and what is the data they provide, this is captured in the MDB ICD.
The Main Database will be hosted at ESAC, data from the DPCs will be received at the end of each Data Reduction Cycle and ingested into the MDB. After consolidation of the data an extract from the MDB will be pushed to the DPCs to start a new Data Reduction Cycle.
This work package covers the data dictionary, MDB, MDB ICD, extraction and ingestion software. It includes writing the requirements specifications, design implementation, and test of these subsystems.

Tasks: Tasks and sub tasks of GWP-T-180-00000 1. Write and maintain ICD, gather inputs 2. Gather requirements specifications for the MDB 3. Design System 4. Document System 5. Implement MDB Dictionary Tool 6. Test MDB Dictionary Tool 7. Generate ICD and MDB 8. Implement Ingestor 9. Implement Extractor 10. Full subsystem tests
Input: Inputs from the CUs to the MDB ICD and to the MDB Dictionary tool
Output: MDB System and associated Tools
Deliverables: MDB ICD, SW distribution for the dictionary tool, Ingestion, Extraction, Consolidation subsystems, Documentation, Test reports
Dependencies: GaiaTools
Interfaces: Interface with all CUs for the ICD definition and generation
Remarks: This work is provided almost entirely by ESA as a service to DPAC.

Gaia DPAC WP:		GWP-T-190-00000
Title: Gaia Transfer System Design/Code/Operate		
Provider(s): J. Hernandez, M. ter Linden, others as needed		
Manager(s): J. Hernandez		
Start: Phase B	**End**: Phase F	**Total Effort**: 48 MM

Objective:
A system called the Gaia Transfer System (GTS) shall be developed in order to drive and prepare all the data transfers. This system shall be deployed and executed in all the data sites (including ESAC). It will have a web interface so that the transfer can be monitored and followed by all parties. The URD of the GTS should cover in detail all the functional aspects of the data transfer which are mentioned in this document. The main functions of the GTS will be:

- To allow the definition by the processing centre of the storage areas and computers where the data will be received and sent.
- To check before a data transfer begins if the necessary resources (disk space, network bandwidth, firewalls, ...) are available.
- To monitor the data transfers and provide information of its state and progress, also to flag any error that may occur.
- To check the completeness of the data before the transfer is initiated.
- To check the integrity of the data after the transfer has concluded.
- To notify the sender and the receiver about the success/failure of the transfer.

Tasks: Tasks and sub tasks of GWP-T-190-00000 1. Gather Requirements 2. Design System 3. Document System 4. Test System 5. Deploy System 6. Maintain System
Input: MDB ICD, requirements from all CUs
Output: Gaia Transfer System
Deliverables: SW distribution for the GTS, Documentation, User's manual, Test reports, ...
Dependencies:
Interfaces: Interface with all DPCs
Remarks:

C.3 Top-level Work Packages of CU2

Gaia DPAC WP:		GWP-M-201-00000
Title: Management and scientific coordination of CU2		
Provider(s): X. Luri, C. Babusiaux, F. Mignard; D. Egret, M. Gai, E. Masana, A. Robin, J.M. Wallut		
Manager(s): X. Luri, C. Babusiaux, F. Mignard		
Start: Phase B	**End**: Phase F	**Total Effort**: 88 MM
Objective: To provide the overall management and coordination activities needed for the proper functioning of CU2. See Sect. 7.6.3.1 for details.		
Tasks: Tasks and sub tasks of GWP-M-201-00000 (Sect. 7.6.3.1) with some additional tasks for specific interface with other CUs.		
Input:		
Output:		
Deliverables: Sect. 7.6.3.1.		
Dependencies:		
Interfaces:		
Remarks: The providers of this WP are the CU2 managers and the managers of the DUs therein		

Gaia DPAC WP:		GWP-T-202-00000
Title: Architecture and technical coordination of CU2		
Provider(s): CNES		
Manager(s): J.M. Wallut		
Start: Phase B	**End**: Phase D2	**Total Effort**: 32 MM
Objective: Tasks directly related to management and coordination of the software development. See Sect. 7.6.3.2 for details.		
Tasks: Tasks and sub tasks of GWP-T-202-00000 Sect. 7.6.3.2		
Input:		
Output:		
Deliverables:		
Dependencies:		
Interfaces:		
Remarks: The main provider of this WP is CNES and thus it is part of the overall technical coordination effort agreed by this institution		

Gaia DPAC WP:		GWP-T-203-00000
Title: Configuration Management and quality assurance for CU2		
Provider(s): CNES		
Manager(s): J.M. Wallut		
Start: Phase B	**End**: Phase D2	**Total Effort**: 12 MM
Objective: Configuration management and product assurance tasks for CU2. See Sect. 7.6.3.3 for details.		
Tasks: Tasks of GWP-T-203-00000: Sect. 7.6.3.3		
Input:		
Output:		
Deliverables:		
Dependencies: CU2 will take advantage of the common CM and QA tools provided by CU1 (as defined in the GWP-T-103).		
Interfaces:		
Remarks: The main provider of this WP is CNES and thus it is part of the overall technical coordination effort agreed by this institution.		

Gaia DPAC WP:		GWP-M-220-0000
Title: Management and implementation of the Universe Model		
Provider(s): Obs. Besançon and other DU3 institutes		
Manager(s): A. Robin, C. Reylé		
Start: Phase B	**End**: Phase F	**Total Effort**: 282 MM

Objective:

Management and coordination of the development of the Universe model necessary for the simulations of the Gaia mission

Tasks:

1. Coordinate the requirements and implementations of the Universe model
    (a) Define the requirements of simulations of a realistic Universe model for the Gaia simulator
    (b) Check for reliability of the defined simulations with regards to the requirements from other DU of CU2 (GIBIS, GASS, GOG)
    (c) Check the reliability of the defined simulations with regards to the requirements from other CU
    (d) Discuss new needs or specific requirements from other CU
2. Check feasibility of the implementation
3. Check for manpower. Find man power for every task in due time
4. Schedule the needs in time, decide priorities if necessary

Tasks:

1. Management and planning of the components
2. Define requirements, following prescriptions from other CUs
3. Produce design of the algorithms
4. Produce Junits for tests of the components of the Universe model
5. Integrate system, produce validation tests and bug tracking
6. optimize system
7. Insure maintenance of the system
8. Produce documents describing the structure, the algorithms, the tests

Tasks:

1. Object astrometric models coordination : check that every simulated objects have the necessary inputs for simulating realistic astrometry
2. Object photometric models coordination : check that every simulated objects have the necessary inputs for simulating realistic photometry
3. Object spectroscopic models coordination : check that every simulated objects have the necessary inputs for simulating realistic spectroscopy

Input:
Simulation requirements from all other CUs and from GIBIS, GASS, GOG. Spectral libraries from CU8.

Output:

Deliverables:
Software module producing catalogues of simulated objects in any direction of the sky with their intrinsic characteristics

Dependencies:
Depends on requirements for simulations from other CUs and requirements from GASS, GIBIS and GOG. Spectral libraries provided by CU8.

Interfaces:
Interfaces for requirement definitions from CU3 (core processing), CU4 (object processing), CU5 (photometry), CU6 (spectroscopy), CU7 (variability), CU8 (astrophysical parameters). Interface with CU8 for the spectral libraries. Interfaces with GASS, GIBIS and GOG development teams.

Remarks:
The effort includes a lot of scientific studies which add up to the development tasks. That explains the large amount of FTE necessary before launch (see effort table). After launch, a small amount of man power may be required for maintenance and checking. If the mission gives unexpected results, more simulations would be needed with revised parameters.

Gaia DPAC WP:		GWP-M-230-00000
Title: Management & implementation of the Instrument Model		
Provider(s): INAF-OATo and other DU4 institutes		
Manager(s): M. Gai, J. Rebordão		
Start: Phase B	**End**: Phase F	**Total Effort**: 300 MM

Objective:
The instrument model has the goal of implementing a set of tools for the simulation of Gaia astrometric, photometric and spectroscopic data, in support to the development of the data reduction software and its subsequent usage throughout the operation phase. Instrument response variation is unavoidable and significant, at the level of sensitivity targeted by Gaia. The goal is to include in the measurement model a detailed description of the instrument response, as a function of the hardware and operation parameters (nominal values in the Gaia Parameter Database), allowing the generation of a realistic representation of the science data, including all known contributions which might affect the performance, in terms of both noise level and systematic error. During development, the objective is the generation of realistic data sets for implementation, optimisation and validation of the data reduction algorithms. During operation and data reduction, the instrument model supports the sanity check by comparison with the data, contributing to identification of effects induced by variation of critical parameters. In particular, the data reduction algorithms may be verified, and possibly further optimised, with respect to the real data statistics.

Tasks:

1. Planning and scheduling
2. Scientific coordination
3. Architecture and technical coordination
4. Reporting (to CU2 Manager, DPACE, ...)
5. Communications (within CU2 structure and outside)

Input:

1. Requirements from ESAC; DPAC; CU2
2. Instrument specifications and component data sheets from ESA / GST
3. Documentation to be circulated among the work packages

Output:

Deliverables:

1. Periodic reports to ESAC, CU2 Manager and DUs within CU2, DPACE, and in case to other CUs and GST;
2. ICD from/to ESAC (to ESA standard);
3. Interface documents to/from other DUs within CU2 and, possibly, outside (to ESA standard).

Dependencies:

Interfaces:

1. CU2 Manager; DPACE; GST; ESA;
2. Internal WPs.

Remarks:
Progress of activity based on timely availability of technical information on the spacecraft and instrument, updated throughout the detailed design, construction, integration, and in-orbit commissioning.

Gaia DPAC WP: GWP-M-240-00000

Title: Management and implementation of GASS

Provider(s): E. Masana, Y. Isasi (UB) + BPC

Manager(s): E. Masana

Start: Phase B	**End**: Phase F	**Total Effort**: 134 MM

Objective:
GASS is the Gaia Telemetry simulator. The objectives of this packages are:

- The overall management and coordination of the GASS tasks, including software production, deployment, coordination with other CU, ...
- The definition of the system environment to develop, integrate, maintain and operate GASS
- The development, implementation and validation of the code in GASS
- The description of the GASS system administration

From 31/12/2011 the main task will be the maintenance of the system.

Tasks:
Tasks and sub tasks of GWP-M-240-00000

1. Management and coordination (1.2 MM/year)
2. Produce software system for GASS (4.8 MM/year)
3. Development of components and algorithms (21.6 MM/year)
4. System administration (6.0 MM/year)

Input:
Requirements from CUs

Output:
GASS telemetry and other auxiliary data generated by the simulator. Documentation.

Deliverables:
ASCII files containing the simulated data (TM, sources, attitude, auxiliary data, ...)

Dependencies: Dependencies with the deliveries of the Universe Model algorithms
Interfaces:
Remarks: CNES will be responsible of system optimization. BPC will collaborate in the deployment of GASS

Gaia DPAC WP:		GWP-M-250-0000
Title: Management and implementation of GIBIS		
Provider(s): C. Babusiaux + OPM and CNES		
Manager(s): C. Babusiaux		
Start: Phase B	**End**: Phase F	**Total Effort**: 170 MM

Objective:
Management and implementation of the pixel-level simulator GIBIS.

Tasks:

1. Review the CUs requirements on pixel-level data; review priorities and manpower; agree development plan with DU1 and DPACE.
2. Implement the universe model by using the models developed by DU3
3. Implement the instrument model by using the models developed by DU4
4. Implement user access: define and implement input parameters and output formats
5. Validate internally the simulation results
6. Deploy and maintain GIBIS simulator access to the DPACE through a web interface
7. Document

Input:
Simulation requirements from all other CUs, Universe models from DU3, Instrument models from DU4

Output:
Detailed pixel-level simulated data covering the DPAC needs.

Deliverables:

1. Development plan
2. Software simulating pixel-level data
3. Access to the simulator through a web interface
4. Documentation and user help

Dependencies:
Depend on inputs to be provided (DU1, DU3, DU4), plus software engineering guidelines from DU2.

Interfaces:
All other WPs of CU2, other CUs for requirements

Remarks:

Gaia DPAC WP: GWP-M-260-00000

Title: Management and implementation of GOG

Provider(s): X. Luri, C. Babusiaux + UB and OPM

Manager(s): X. Luri

Start: Phase B	**End**: Phase F	**Total Effort**: 152 MM

Objective:
GOG is the Intermediate Data Simulator of the Gaia mission. The objectives of this packages are:

- The overall management and coordination of the GOG tasks, including software production, deployment, coordination with other CU, ...
- The definition of the system environment to develop, integrate, maintain and operate GOG
- The development, implementation and validation of the code in GOG
- The description of the GOG system administration

From 31/12/2011 the main task will be the maintenance of the system.

Tasks:
Tasks and sub tasks of GWP-M-260-00000

1. Management and coordination (1.2 MM/year)
2. Produce software system for GOG (2.4 MM/year)
3. Development of components and algorithms (9.6 MM/year)
4. System administration (4.8 MM/year)

Input:
Requirements from CUs

Output:
GOG files with data generated by the simulator. Documentation.

Deliverables:
ASCII files containing the simulated data

Dependencies: Dependencies with the deliveries of the Universe Model algorithms
Interfaces:
Remarks: CNES will be responsible of system optimization. BPC will collaborate in the deployment of GOG

Gaia DPAC WP: GWP-S-270-0000

Title: Scientific Quality Assurance and Validation of CU2 simulations

Provider(s): D. Egret + OPM

Manager(s): D. Egret

Start: Phase B	**End**: Phase F	**Total Effort**: 113 MM

Objective:
The role of DU8 is to organize the scientific validation of the outcome of the simulators.

Tasks:

1. Internal consistency
2. External consistency (comparison with observational catalogues)
3. Processing of required validation tests.

Input:
Inputs of the work package are :

1. Results from DU5, DU6 and DU7 : simulators GASS, GIBIS and GOG
2. Requirements of validation, agreed by DPAC

Output:
Analysis of quality assessment and scientific validation of simulators.

Deliverables:
Reports of quality assessment of simulators.

Dependencies:
Depends on achievement of working versions of simulators produced by DU5, DU6 and DU7. Requirements (alerts) are expected from all users of simulators.

Interfaces:
Main interfaces : DU5, DU6, DU7

Remarks:
Remarks : Detailed organisation of this work package is still to be defined. Activity on this WP is not expected to start before July 2007.

C.4 Top-level Work Packages of CU3

Gaia DPAC WP:		GWP–M-301-00000
Title: Management and scientific coordination of CU3		
Provider(s): U. Bastian, W. O'Mullane, J. Torra, M. Lattanzi, S. Klioner, S. Jordan, L. Lindegren		
Manager(s): U. Bastian; M. G. Lattanzi, J. Torra		
Start: Phase B	**End**: Phase F	**Total Effort**: 26 MM
Objective: Sect. 7.6.3.1		
Tasks: Tasks and sub tasks of GWP-M-301-00000: Sect. 7.6.3.1.		
Input:		
Output:		
Deliverables: Sect. 7.6.3.1.		
Dependencies:		
Interfaces:		
Remarks: The providers of this WP are the members of the CU3 Steering Committee		

Gaia DPAC WP:		GWP-T-302-00000
Title: Architecture and technical coordination CU3		
Provider(s):		
Manager(s): W. O'Mullane		
Start: Phase B	**End**: Phase F	**Total Effort**: 13 MM
Objective: Sect. 7.6.3.2		
Tasks: Tasks and sub tasks of GWP-T-302-00000 Sect. 7.6.3.2		
Input:		
Output:		
Deliverables:		
Dependencies:		
Interfaces:		
Remarks: Much of the effort in this WP has been pushed into specific sub-work packages for the CU3 software components. The technical roles for these have been devolved to CU3 team members in agreement with the CU leader. Hence this is a light management package for the CU technical leader.		

Gaia DPAC WP:		GWP-T-303-00000
Title: Configuration Management and quality assurance for CU3		
Provider(s): ESAC		
Manager(s): J. Hoar		
Start: Phase B	**End**: Phase F	**Total Effort**: 13 MM
Objective: Sect. 7.6.3.3		
Tasks: Tasks of GWP-T-303-00000: Sect. 7.6.3.3		
Input:		
Output:		
Deliverables:		
Dependencies: CU3 will take advantage of the common CM and QA tools provided by CU1 (as defined in the GWP-T-103).		
Interfaces:		
Remarks: The Hardware configuration management activities will be described in sub-work packages within this WP for each DPC (ESAC, BPC and INAF-OATo).		

Gaia DPAC WP:		GWP-M-304-00000
Title: Integration, Validation and Operation of CU3 systems		
Provider(s): J. Hoar, A. Volpicelli, BPC		
Manager(s): W. O'Mullane		
Start: Phase B	**End:** Phase F	**Total Effort:** 61 MM

Objective:
As in Sect. 7.6.3.4. This WP covers work performed by the Operations Teams of the Gaia SOC at ESAC as well as INAF-OATO and BPC.

Tasks:
The tasks of this work package are: as in Sect. 7.6.3.4
More specifically this WP covers operation of:

1. Initial Data treatment (IDT)
2. First Look (FL)
3. Astrometric Global Iterative Solution (AGIS).
4. Intermediate Data Update (IDU)
5. Astrometric Verification (AVU)
6. End-to-end test participation (see GWP-T-150-00000)

Input:

Output:

Deliverables:

Sect. 7.6.3.4

Dependencies:

Interfaces: Mainly between CU3 (ARI, Lund, Barcelona) and ESAC as DPC for IDT/FL/AGIS. Also between CU3 and INAF-OATO as DPC for AVU. Between CU3 (ARI, Barcelona) and BPC as DPC for IDU. OATO and BPC data centres need to interface with ESAC for data transfer.
Remarks:

Gaia DPAC WP: GWP-M-320-00000

Title: Management and implementation of the AGIS (framework and algorithms)

Provider(s): LUND/ESAC

Manager(s): L. Lindegren, U. Lammers

Start: Phase B	**End**: Phase F	**Total Effort**: 541 MM

Objective:
The main objective of the work package is to plan, design, implement, operate and maintain the Astrometric Global Iterative Solution (AGIS) system. AGIS is the sole, central framework that will generate fundamental astrometric mission products for Gaia. These are:

- astrometric parameters for a subset of 'primary' sources to an accuracy level compliant with the overall astrometric mission accuracies defined in MRD SCI-250, and provisional values of the same parameters for all 'secondary' sources
- the three-axis attitude of the Astro instrument axes as function of time
- the geometric calibration of the Astro instrument as function of time and of the photometric characteristics of the sources
- an optional set of 'global' parameters describing the transformation from astrometric parameters and time to observed source directions.

These data must be determined in an internally consistent celestial reference frame, consistent with the ICRS, and providing the basis for the further astrometric analysis of secondary sources. The subset of primary sources is iteratively selected as part of the AGIS processing. The selection is based on the observed astrometric, photometric, spectroscopic and imaging characteristics of the sources, as well as the need for an adequate sampling versus position, magnitude and colour. Sources for which there is some indication that they may be non-single or otherwise deviating from the standard astrometric model will be excluded from the primary set. AGIS should be able to handle at least 10^8 primary sources.

Tasks:
Tasks may be broadly broken down into the two categories ‘Technical’ and ‘Scientific’. Only a top-level outline of each one is given here whilst a more detailed description is deferred to the corresponding subordinate packages GWP-T-320 and GWP-S-320 respectively.
Technical tasks:

- System definition
  The task encompasses all activities needed to formally define the system as a major and complex software product and to start the development process. This includes: requirements gathering, choosing and outlining the architectural design, software development tools and practices, quality assurance and software project management approach.

- Data ingestion and export of products
  Identify, design, and implement all processes needed to setup and organise the data needed for an AGIS cycle. This will include the extraction of all AGIS-relevant data from the MDB and their ingestion into an AGIS-specific store. At the end of an AGIS cycle the resulting output data need to be generated in an MDB-ICD compliant form.

- Data storage and access
  The task shall identify an optimal store mechanism for all data needed to run AGIS. From this store, data shall be read and passed on to the scientific algorithms, and computed data shall be written back. This could be a relational DB. The task shall encompass the design and implementation of the storage mechanism itself and all data access software.

- Algorithm execution framework
  AGIS is intrinsically amenable to parallelisation. The task shall identify, design and implement a framework in which the AGIS algorithms can be deployed and run optimally in a multi-processor environment. Also, the overall algorithmic flow control shall be designed and implemented as part of this task, i.e., the GIS loop control, triggering of final update calculations in the algorithms, etc.

Tasks:

- Monitoring tools
  For a running AGIS cycle it shall be possible to assess the status of the processing at any point in time. This shall include information about performance and other technical aspects but also the monitoring from a scientific and numerical point of view. From this it shall be possible to judge the numerical stability and convergence behavior of the system to detect and counteract problems quickly. The task shall identify, design, and implement all tools and components to support this concept. This will probably entail the real-time generation of scientific and technical graphs and their proper visualisation (e.g. through a Web interface).

- Algorithm integration
  Algorithms developed and unit-level tested outside the ESAC AGIS build environment need to be integrated into the system. This process shall be eased by providing a portable development environment to be used by the algorithm provider. This shall include a complete test data set that the algorithm can be sensibly tested against before delivery. The actual integration process will possibly comprise the usage of the algorithm as part of either existing or specific higher-level system integration tests.

  The task may also entail the writing of algorithms from scratch that are sufficiently detailed by technical notes or design documents.

- Algorithm optimisation
  Integrated algorithms that are functional in the baseline operational system need to be optimised to minimise overall runtimes and memory requirements. The task shall entail the usage of performance optimisation tools to generate detailed execution profiles, and re-writing or re-organisation of time-critical parts of the codes and/or any used components.

- System validation and testing
  The system needs to be thoroughly validated and tested with simulated data to prove both numerical correctness and computational feasibility at all times. These tests shall be performed on a regular basis up to the beginning of the operational phase (and possibly beyond) on an ever increasing volume of simulated data. Before launch these tests should show that the system is capable of processing the full amount of data in the envisaged way, i.e. reaching the targeted astrometric accuracies.

Tasks:

- Operation
  After the commencing of the operational phase the pre-launch-validated system needs to be run on a regular basis in line with the overall DPAC processing schedule. For each processing cycle, this activity will comprise the retrieval of all intermediate observation and initial astrometric, attitude, calibration, and global data needed to start a cycle. The cycle needs to be started and then monitored from a technical and scientific viewpoint. Potential problems during a run need to be identified, analysed, resolved, and a new cycle started. After convergence the output data need to be generated in an ICD-compliant form for re-ingestion into the MDB.

Scientific tasks:

- Algorithm definition
  This task shall identify and define the mathematical models and numerical methods needed to determine the source, attitude, calibration and global parameters from the observed centroid positions at CCD level. This includes procedures to compute the observed directions of sources in the framework of General Relativity, part of the spacecraft/instrument modelling (attitude perturbations, geometric calibration including chromaticity, CTI and gating), the link to the extragalactic reference frame, outlier management, weight balancing, and the identification of quality indicators and convergence criteria.

- Algorithm implementation
  This task covers the development of code for all identified algorithms either directly within the AGIS processing framework, or within a testbed environment to a state that allows safe and straightforward integration into the framework. Software units will be equipped with stand-alone unit test cases. Each major algorithm will in addition be equipped with sub-system test cases that can be run, for example, in the testbed environment.

Tasks:

- Test planning
  A sequence of tests will be identified both at sub-system and system level, whose successful execution will guarantee that the algorithms achieve the targeted accuracy and level of complexity at specific milestones. The tests will typically comprise the use of noise-free and noisy data as well as updating starting from offset parameter values. Since the milestones represent a progression in the complexity of source/attitude/instrument modelling, the characteristics of the test data must progress in a parallel fashion.

- Interpretation of results
  The results obtained during the large-scale system tests and later during the operation phase need to be interpreted from the scientific viewpoint in order to quickly identify and remedy possible deficiencies, e.g., in the adopted instrument and attitude models. The interpretation will to a large extent use the standard monitoring tools, but will also require the development of ad hoc analysis software. This task will ensure that sufficient resources and time are set aside for these activities, and that additional off-line analysis tools (including, for example, suitable statistical and graphical packages) are made available and/or developed.

Input:
List of inputs to the package:

- ECSS-tailored QA document detailing requirement documents and development guidelines
- Overall System Architecture Document from CU1
- Definition of software development environment from CU1
- Provision of Java language coding style document from CU1
- Provision of common Java toolbox
- Provision of Gaia Parameter Database
- Common calibration model
- Main Database Interface Control Document defining the Main Database input to AGIS
- Simulator Interface Control document(s)
- Provision of adequate bug/issue tracking tool

Output:
The WP shall produce an operational AGIS capable of meeting the objectives (see above)

Deliverables:
The following deliveries are foreseen:

- All source code representing the operational AGIS - intermediate deliveries (according to milestones) reflecting increasing algorithmic complexities will be done up to launch (and probably beyond)
- Regular progress reports
- Technical and scientific reports on executed tests (before operations)
- Technical and scientific reports on completed operational cycles (in operations)
- Input to the Main Database ICD defining the output of AGIS
- ECSS-compliant documentation (Requirements + Design documents, User documentation)

Dependencies:
Most of the testing aspects are closely linked to the timely availability of corresponding simulation data from CU2 via the MDB. There is also a close link to the IDT activities. Scientific effects can only be considered in AGIS if they are likewise implemented in the simulator.

Interfaces:
Nothing in particular — AGIS will interface with the rest of the DPAC system via the data received from and delivered to the MDB in the ICD-compliant format.

Remarks:
The required effort for this WP, 541 man-months, corresponds to an average of 4.2 FTE over the period 2007–2019. These 4.2 FTE are divided among the following subtasks (second-level WPs):
GWP-T-320-10000: Management and planning
GWP-T-320-11000: Gather requirements
GWP-T-320-12000: Design system
GWP-T-320-13000: Document system
GWP-T-320-14000: Test system
GWP-T-320-15000: Optimize system
GWP-O-320-16000: Maintain system
GWP-O-320-17000: Handover/Train operations staff
GWP-O-320-18000: Install, operate, monitor running system
GWP-D-320-20000: Implement data processing framework
GWP-D-320-21000: Implement/integrate core AGIS algorithms
GWP-D-320-22000: Implement diagnostic/monitoring framework
GWP-S-320-31000: Algorithm definition
GWP-S-320-32000: Algorithm implementation
GWP-S-320-33000: Test planning
GWP-S-320-34000: Analysis of test results
GWP-S-320-35000: Analysis of operational results
GWP-S-320-36000: System orientation
Thus just 0.23 FTE is assigned to each of these on average.

Gaia DPAC WP: GWP-M-330-00000

Title: REMAT: Relativistic Models and Tests

Provider(s): LO, INAF-OATo, UNIPD, UB and OPM

Manager(s): S. Klioner

Start: Phase B	**End**: Phase F	**Total Effort**: 660 MM

Objective:

The aim of this work package is to provide the required relativistic support for Gaia. This includes two major components: data modelling and relativistic tests with Gaia data.

- Fully consistent relativistic data treatment is the only way to guarantee that the astrometric products of Gaia are physically meaningful. This implies
  - establishing and monitoring of the relativistic consistency of various parts of the data processing chain and auxiliary data, and
  - formulation, optimization, testing, implementation and maintenance of the general-relativistic model for astrometric observations with microarcsecond accuracy.
- Realization of the Gaia potential as an experimental tool of gravitational physics should be achieved by planning, designing and optimizing a broad spectrum of relativistic tests with Gaia data.

This work package includes also software implementation of the corresponding algorithms as well as operational support and maintenance.

Tasks:
Only the top-level structure is shown here. Detailed description is deferred to the corresponding subpackages.

1. Monitoring of the consistency of various data processing components and auxiliary data in the relativistic framework

   Main subtasks are:

   - Relativistic background of the solar system ephemeris for Gaia
   - Relativistic background of Gaia orbital data
   - Relativistic formulation of orbit determination of minor bodies
   - Relativistic background of astronomical constants
   - Relativistic background of synchronization OBT–UTC
   - Relativistic background of the Gaia reference frame

2. Formulation, optimization, testing and implementation of the general-relativistic model for Gaia

   This task includes:

   - Two independently-developed relativistic models (GREM, Dresden and RAMOD, INAF-OATo and Padova)
   - Theoretical comparison of the two models
   - Numerical comparison of the two models
   - Formulation of the operational model
   - Software implementation for the Gaia simulator
   - Software implementation for the AGIS software
   - Maintenance of the operational model
   - Monitoring of the result consistency coming from different instances of the relativistic model
   - Relativistic aspects of binary-star models

Tasks:

3. Design, implementation and operation of the DU-specific simulation tools

   Several tools are planned: testbed for comparisons of the relativistic models, special modules for Gaia AGIS software aimed at specific relativistic simulations (that are not included in the main AGIS development chain), software for global parameter fits from a simulated final astrometric results (e.g., expansion of the proper motion field into vector spherical harmonics).

4. Global tests of relativity theory

   Global tests are the kind of the tests which involve all observational data available from Gaia (or at least as much data as possible). The planned global experiments are:

   - Tests of gravitational red shift with Gaia's clock
   - Tests of Robertson-Mansouri-Sexl parameters from aberration
   - Standard determination of the PPN parameter γ
   - Separate determination of the PPN parameter γ from the Sun, Jupiter, Saturn and the Earth
   - Alternative deflection patterns
   - Pattern matching in time-dependent individual positions and proper motions (this includes various tests: acceleration of the solar system barycenter relative to the quasars, stochastic background of primordial gravity waves, stability checks of the solutions etc.).

Tasks:

6. Local tests of relativity theory

   Local tests are the kind of the tests which involve some restricted amount of specially selected Gaia data. The planned local experiments are:

   - Differential astrometric solutions for the data close to the giant planets (mainly, independent measurements of several components of the light deflection due to major planets: the stationary (Schwarzschild) deflection, the quadrupole deflection and the gravitomagnetic deflection due to translational motion of the planets; investigations of the differential astrometric models, identification of the noise sources, numerical simulations, implementation and operation of the tests)
   - Relativistic tests with asteroids (large scale tests of the perihelion precessions of the asteroids due to the Schwarzschild field of the Sun, test of the equivalence principle with the Trojan and other resonance asteroids, tests of the relativistic N-body (non-Schwarzschild) effects)
   - Relativistic tests with the solar system ephemeris improved using Gaia data for minor and major planets and natural satellites (identification of the Gaia results useful for solar system ephemerides, improvement of Solar system ephemeris using Gaia data for minor and major planets and natural satellites, and execution of the standard set of the relativistic tests with the new improved ephemeris)

   The list of the global and local relativistic experiments will be refined as a result of investigations in the framework of this work package.

Input:

1. Various documents describing auxiliary data (solar system ephemeris and astronomical constants, Gaia orbit determination, OBT–UTC synchronization)
2. Software interfaces for the relativity model modules for the AGIS software and for the Gaia simulator
3. Documents describing various expected characteristics (e.g. noise) in various astrometric products of the mission
4. AGIS software implementation
5. Documents describing the rule for creating shell task software components

Output:
The whole set of scientific and technical reports on the various relativity-related aspects of the mission and data processing. The whole set of software for the operational relativistic model for AGIS and simulations, and the RAMOD implementation for AVU. Software for the relativistic tests (as part of AGIS and as stand-alone shell tasks).

Deliverables:
The following deliveries are foreseen:

- Scientific and technical reports on the relativistic interpretation of the auxiliary data
- Scientific and technical reports on the operational relativistic model and on the results of the comparisons of the alternative models
- Software module for the full operational general-relativistic model for AGIS
- Software module for the full operational general-relativistic model for the Gaia simulator(s)
- Software module for an extended relativistic model for the planned simulations of the relativistic tests
- Scientific and technical reports on the relativistic tests, their physical merit, expected accuracy and technical details of their implementation within Gaia
- The whole set of software necessary for the relativistic tests (as part of AGIS and as stand-alone (shell) applications)

Dependencies:
This work package depends on very many other parts of the project. In particular on AGIS, on CU2 (simulated data are necessary for optimizing the relativistic tests), on CU4 (solar system object processing for local relativistic tests). The software development of AGIS and the Gaia simulators (CU2) depends on the modules for the operational relativistic model to be delivered by REMAT. The software development of AVU depends on the implementation of RAMOD also delivered by REMAT.

Interfaces:
REMAT interfaces to

- AGIS (for the operational relativistic model and global relativistic tests within GIS)
- Gaia simulator (specifically the GASS and GOG data generators) for testing purposes
- AVU (for the REMAT implementation of RAMOD for GSR)
- The solar system objects data processing chain of CU4 for local tests involving solar system objects

Remarks:
The required effort for this WP, 660 man-months, corresponds to an average of 5.1 FTE over the period 2007–2019. These 5.1 FTE are divided among the following subtasks (second-level WPs):
GWP-T-330-10000: Management and planning
GWP-T-330-20000: Relativistic consistency of various data processing components and auxiliary data
GWP-T-330-30000: Basic relativistic model
GWP-T-330-40000: Special simulation tools
GWP-T-330-50000: Global tests of General Relativity
GWP-T-330-60000: Local tests of General Relativity
Thus just 0.8 FTE is assigned to each of these on average. Each contains a development part and an operations/maintenance part.

Gaia DPAC WP:		GWP-M-335-00000
Title: Auxiliary data: definition and acquisition		
Provider(s): ARI, INAF-OATo, ZAH, IMCCE, Porto		
Manager(s): U. Bastian		
Start: Phase B	**End**: Phase E2	**Total Effort**: 254 MM

Objective:

This work package initiates, coordinates and implements the definition and acquisition of all sorts of auxiliary data needed for the core processing. The only exceptions are data sets that naturally fall into the responsibility of other CUs, even though they are in practice used within some IDT/FL process. One example for this is the photometric standard star catalogue to be used in the one-day photometric calibration within the FL processing.

The responsibilities for the definition and provision of the various auxiliary data sets are fairly scattered, according to the scattered expertise needed. There is thus no well-defined, localized development unit associated with this work package, although an attempt has been made to concentrate the work package at the ARI. The presently foreseen managers of the various data sets are indicated in the 'tasks' section below. Each of the tasks may in turn involve several providers in different locations. The managers' responsibilities are mainly the detailed definition of the data sets, the initiation and coordination of the necessary activities on the side of the providers and the compilation of the inputs into a homogeneous data set in an agreed form.

All the resulting data sets are to be stored either in the Gaia main database (MDB) or in the Gaia parameters database (PDB).

Tasks:

- Orbit data:
  The orbit of Gaia with respect to the solar-system barycentre (location and velocity as function of time) is needed for parallax, light bending and aberration corrections in astrometry, as well as for light travel time and radial velocity corrections in the photometry and RVS data reduction, respectively.
  Manager: M. Biermann (ZAH). Providers: Gaia ground segment (ESOC, Darmstadt) for the data contents; Gaia SOC (ESAC, Villafranca) for the storage and access.

- Ground-based optical spacecraft tracking:
  The usual orbit determination procedures for interplanetary spacecraft will not be sufficient to reach the precision needed for Gaia. This problem can be solved by daily ground-based optical astrometry of the Gaia satellite. It is necessary to negotiate observation programs at suitable observing facilities, to initiate test campaigns and to prepare the routine observations and data reductions. Observed topocentric Gaia positions will be delivered to the ESOC flight dynamics department. All activity is to start in 2008.
  Manager: B. Voss (ARI). Providers: tbd ground-based astronomical observatories

- OBT/UTC relation:
  The instrumental clock of Gaia must be very precisely correlated with a well-defined physical time coordinate (specifically: TCB). This is done via an OBT/UTC relation provided by the ground segment, and a standard UTC-to-TCB transformation (using also the Gaia orbit data).
  Manager: M. Biermann (ZAH). Providers: Gaia ground segment (ESOC) for the data contents; Gaia SOC for the storage and access.

- Initial calibration data:
  The Gaia data reductions (specifically the IDT and FL tasks) need a very wide variety of instrument and spacecraft calibration data to start the processing.
  Manager: M. Biermann (ZAH). Providers: EADS/Astrium (for nominal and pre-launch laboratory values); Gaia ground segment, EADS/Astrium and First Look (for initial in-orbit calibration data); Gaia SOC for the storage and access.

Tasks:

- Initial Gaia source list:
  Although Gaia in principle produces a completely independent sky survey, it is in practice very useful to start the cross-matching process with some pre-existing, almost complete ground-based sky inventory. This is the initial Gaia source list. It will be derived from existing astronomical data. The main such source will be the GSC 2 Data Base, supplemented by other all-sky inventories like the USNO-B, 2MASS, and others which might become available in the coming years, appended by special-object lists like e.g. a quasar catalogue.
  Manager: R. Smart (INAF-OATo). Providers: INAF-OATo (compilation); public astronomical data archives.

- Attitude star catalogue:
  A reference star catalogue for attitude determination before the main IDT cross-matching process must be constructed before launch. The main data source will probably be the UCAC or a successor of that catalogue (e.g. a preliminary URAT catalogue). No dedicated observations will be needed. Activity is to start in 2009.
  Manager: R. Smart (INAF-OATo). Providers: INAF-OATo and H. Lenhardt (compilation); public astronomical data archives.

- Solar-system data I — ephemerides for major planets and the moon:
  The AGIS and the fundamental-physics goals of Gaia need a relativistically consistent high-precision model for the masses and motions of the major solar-system bodies, corresponding to very strict specifications and requirements in term of accuracy, time-scale and reference system. This will be based on the state-of-the-art planetary theories fitted to observations (classical astrometry, range measurements on space probes, . . .).
  Manager: F. Mignard. Provider: A. Fienga (IMCCE, Paris)

Tasks:

- Solar-system data II — ephemerides and physical data for minor planets: Orbital and physical data on minor solar-system bodies are needed for the cross-identification and reduction of the relevant Gaia observations, for relativistic parameters and for other purposes. An ephemeris with an optimised access tool must be provided. This is primarily a S/W with numerical integration (or any other representation) giving access to position and velocity of a minor planets at any time, the accuracy being determined by that of the available osculating elements. This package will be regularly updated (once or twice a year) until the end of the mission, as more orbital elements will become available on ground.
  Manager: F. Mignard. Provider: J. Berthier (IMCCE, Paris)
- Ecliptic-poles star catalogue:
  For the sake of the initial in-orbit calibration and verification of Gaia it is intended to produce special star catalogues for one or two small sky fields (about 2 degrees) around the ecliptic poles. For this task it is both needed to screen existing astronomical catalogue archives and to initiate dedicated ground-based observation campaigns. One goal is to prepare a star list down to about magnitude 23 (in R) which is as complete as possible. Another goal is to provide precise observational data for as many stars as possible down to magnitude 20 (photometry, astrometry) or magnitude 17 (radial velocities), respectively.
  Manager: B. Voss (ARI). Providers: B. Voss; tbd ground-based astronomical observatories; public astronomical data archives.

Tasks:

- Initial QSO catalogue:
  For the inertial de-rotation of an AGIS celestial sphere it is needed to have a list of confirmed QSOs. For the definition of the coordinate origins it is needed to have the list of QSOs defining the radio ICRS (at the time of the Gaia mission), critically revisited for the purposes of Gaia.
  Manager: J. Osorio (Porto). Providers: Porto, INAF-OATo; public astronomical data archives.

Input: • Requests and requirements from other CU3 WPs • MDB Interface Control Document • Public astronomical catalogues • Dedicated ground-based observations
Output: The WP shall produce the auxiliary data sets specified in the list of tasks.
Deliverables: The following deliverables are foreseen: • Brief requirements documents for the data sets • The auxiliary data files • Detailed descriptions of the data (contents, format and construction)
Dependencies: see list of tasks
Interfaces: see list of tasks

Remarks:
The required effort for this WP, 254 man-months, corresponds to an average of 1.9 FTE over the period 2007–2019. These 1.9 FTE are divided among the following subtasks (second-level WPs):
GWP-S-335-11000: Initial Gaia source list
GWP-S-335-12000 Attitude star catalogue
GWP-S-335-13000 Initial QSO catalogue
GWP-S-335-14000 Ecliptic-poles star catalogue
GWP-S-335-21000 Solar-system data I: ephemerides for major planets and the moon
GWP-S-335-22000 Solar-system data II: ephemerides and physical data for minor bodies
GWP-M-335-31000 Initial calibration data
GWP-M-335-41000 Orbit data
GWP-M-335-42000 OBT/UTC relation
GWP-M-335-43000 Ground-based optical spacecraft tracking
Thus just 0.2 FTE is assigned to each of these on average. Some contain a development part and an operations/maintenance part. The biggest tasks are GWP-S-335-110, GWP-S-335-13000 and GWP-M-335-43000 who together require two thirds of the entire effort.

Gaia DPAC WP:		GWP-M-340-00000
Title: Astrometric Verification Unit (AVU), BAM and WFS processing		
Provider(s): INAF-OATo, University of Padova		
Manager(s): M. G. Lattanzi		
Start: Phase B	**End**: Phase F	**Total Effort**: 910 MM

Objective:

The Astrometric Verification Unit (AVU) is, within the Coordination Unit 3, the Development Unit dedicated to the full accuracy verification of the astrometric measurements of Gaia.
The following tasks are all part of the verification activities:

i) verification of the performance of specific parts, those identified as being critical to the astrometric error budget, of the IDT pipeline;

ii) verification of the pipeline core astrometric solution (AGIS);

iii) verification of short and long time scale behaviour of the Basic Angle;

iv) astrometric instrument model maintenance and operation;

v) calibrations of the SM and AF parts of the focal plane modelling;

vi) processing, and interpretation of WFS data if available.

This WP takes responsibility for providing all the necessary resources for the timely development, implementation, test, and successful operation of AVU within the Italian DPC. In particular, it provides scientific and technical coordination, develops and updates the AVU Implementation Plan, develops and maintains an efficient scheduling system. Also, it guarantees that proper coordination is in place with the other DUs of CU3 and within the sub WPs comprising the its WBS.
[See also GAIA-C3-TN-INAF-ML-001.]

Tasks:

1. Development, implementation, and operation of an independent version of the global sphere reconstruction named GSR.
2. Verification of the performance of specific parts, those identified as being critical to the astrometric error budget, of the IDT/IDU pipelines. This task provides the framework for testing different algorithms so to verify the performance of specific parts of the main IDT/IDU pipelines. The overall structure must be at any step consistent with the current IDT/IDU, apart for the section currently investigated. Specific modules will encode algorithms alternative to the corresponding IDT baseline, and verified with respect to reliability, efficiency and other aspects. This activity will contribute to the integration of the alternative algorithms in the baseline IDT/IDU pipeline if needed.
3. Verification of the baseline core astrometric solution (AGIS). This is done through direct comparisons to GSR results.
4. Verification of short and long term behavior of the Basic Angle, i.e., monitoring and modeling of the data from the BAM device (this includes data reception, archiving, and routine processing at the INAF-OATo DPC), and dissemination to the relevant CU3 DUs.
5. Astrometric instrument model maintenance and operation. The astrometric part of the instrument model developed for the DU4 of CU2 will be used during ground-based activities and later during science operations for the (bottom-up) monitoring of the astrometric error budget and for establishing and executing calibration procedures for the SM and the astrometric (AF) section of the focal plane.
6. Modeling, processing, and interpretation of WFS data.

Input:

1. Description of AGIS procedures.
2. Data for AGIS from ESAC, according to ICD.
3. Results from AGIS from ESAC, according to ICD.
4. Description of IDT, IDU procedures.
5. Data from BAM and WFS.
6. Relativistic model from the RAMOD part of the REMAT DU for GSR.

Output:

1. Data and characterization of BAM temporal variations.
2. Data and characterization of WFS data.
3. Data and characterization of the AGIS-GSR comparisons.
4. GSR results.
5. Calibration data for the SM and AF sections of the payload and the two FOVs.

Deliverables:

1. Reports on GSR activities.
2. Reports on comparisons between AGIS and GSR sphere reconstruction.
3. Reports on comparisons between results of critical IDT procedures and alternative IDT procedures.
4. Computer code of the physical model of the BA device and its transformation into astrometric quantities (lines of sight differences) for BAM activities.
5. Updates on the astrometric instrument model.
6. Reports on calibration procedures for SM and AF.

Dependencies:
Effective development and utilization of the astrometric instrument model, and proper interpretation of the BAM (WFS) data will require, from ESA, the technical details of the astrometric payload and of the monitoring/calibration devices. In particular, given its strategic importance for the final astrometric error budget, both for the instrument model and BA monitoring activities ESA is expected to provide:

a) Ground-based development and testing programs,

b) updates of laboratory measurements,

c) the relevant FEM and thermal models and their updates (for example, these are necessary for the realization of the model of the BAM that will be used for the astrometric monitoring by DPAC).

Also, suitable and competent technical interfaces within the Agency and/or Industry will have to be guaranteed during all the phases of mission implementation (device development, construction, ground testing and calibration, in-flight commissioning, science operations).

Interfaces:
Main interfaces are coordination with IDT, IDU implementation and development, AGIS development (see Inputs and Deliverables), REMAT, and with CU3 management for reporting purposes (see Deliverables).

Remarks:
Activities of this unit will take place at the INAF-OATo DPC. [See GAIA-C3-TN-INAFML-001 for general description and details of the AVU.] The required effort for this WP, 910 man-months, corresponds to an average of 7.0 FTE over the period 2007–2019. These 7 FTE are divided among the following subtasks (second-level WPs):
GWP-M-340-10000: Management and coordination
GWP-M-340-20000: Verification of different IDT algorithms
GWP-M-340-30000: GSR development and maintenance
GWP-M-340-40000: Design and maintenance of comparison tools
GWP-S-340-50000: Astrometric instrument model maintenance
GWP-S-340-60000: BAM model and monitoring
GWP-M-340-90000: Operations
Thus just 1.0 FTE is assigned to each of these on average.

Gaia DPAC WP: GWP-M-345-00000

Title: IDU: coordination, framework, modules and operations		
Provider(s): UB		
Manager(s): J. Torra		
Start: Phase B	**End**: Phase F	**Total Effort**: 744 MM

Objective:

This work package aims to provide algorithms, framework and processing environment, implementation and operation for the Intermediate Data Updating (IDU) process.

The treatment of raw data by the IDT process results in the production of elementary data. Raw data are daily stored in the Raw Data Base and the elementaries are stored in the MDB. The IDT process makes use, besides of raw data, of current values for calibration, attitude, satellite orbit, etc. Thus every time that the AGIS process (every six months) produces a new calibration and refined attitude, the raw data must be treated again in order to produce new elementaries based on better calibration parameters. In addition to the AGIS results, information coming from the treatment of objects (photometry, radial velocity, others) should be incorporated in the redetermination of elementaries. Other processes like cross-matching should be rerun with better positions and photometric data.

IDU will repeat some of the IDT processes. On the one hand it can be simpler than the IDT because some of the variables have been already fixed (e.g. the attitude is much better), while on the other hand it will be complicated by the fact that detailed information on calibration etc. must be integrated. In any case, IDU is a very demanding process, both in terms of storage and computing capabilities that must be run every six months. A computer like the Mare Nostrum at the BPC will be appropriate for this task, and the right environment has to be designed to implement the IDU.

Tasks:

- IDU technical system definition
  This workpackage will perform a detailed study of the computational resources needed for IDU and for the local storage of the raw database. A second point will be the design of the system taking into account the available resources (in principle the Mare Nostrum at BPC, that will be upgraded in a couple of years).

- IDU software design
  The task result will be the design of the full IDU process and its interfaces taking into account the Overall System Architecture as well as other conditions set by the scientific and technical requirements for the core processes.

- Coding of modules
  The development of the IDU will require the coding of several modules performing the functionalities identified in the phase of design. It is expected that some of the IDT modules can be reused. Some of the modules will be produced by CU3 and some other by external people. The aim of this package is to produce the first ones and coordinate the tasks for the external ones.

- Raw database
  The aim is to study the design and implementation of the raw database as well as its management and interface with the IDT and MDB

- Integration and Validation of the IDU
  The IDU will be implemented in the BPC facilities according to the results of the tasks described above and to the Overall System Architecture.

- IDU modules testing
  IDU scientific modules must be individually tested before integration in the whole IDU system. A testbed providing input data produced by other IDT modules or by the Gaia simulator must be provided.

- IDU system testing
  The IDU algorithms should be fully tested from the scientific point of view. In addition, stress tests should be performed in order to optimize the task distribution in the available hardware resources.

Tasks:

- Maintenance and operation
  The completed system has to be run in roughly half-year intervals, and maintained throughout the entire data reduction phase.

Input:
List of inputs to the package as for GWP-M-320-00000, plus:

- IDT modules provided by other tasks

Output:
The WP result will be an operational IDU system capable of storing all the raw database and producing every six months an updated version of all the elementary data available, taking into account the results from AGIS and other processes.

Deliverables:
The following deliveries are foreseen:

- IDU design and implementation documents
- A system and framework to run the IDU
- Code modules to be implemented in successive versions of the operational IDU
- Regular progress and test reports
- Test Plan document
- Test results (before mission operations)
- Mission processing results (during and after mission operations)

Dependencies:

- The development and testing depends on the availability of simulated input data
- The IDU design and implementation is heavily dependent on the operational data inputs to be considered, i.e. on the MDB ICD, raw data ICD etc.
- The IDU development must follow AGIS and other CUs developments
- The IDU development depends on modules provided by other CUs and by the IDT task.

Interfaces:
The IDU interfaces with:

- The Local Raw Database (LRD) for inputs of raw data
- The MDB for inputs of elementaries, attitude etc.
- The MDB has to store its results, after validation and integration processes.
- For testing purposes IDU interfaces with GASS and other sources of simulated data (e.g. IDT and MDB prototype versions)

Remarks:
The required effort for this WP, 744 man-months, corresponds to an average of 5.7 FTE over the period 2007–2019. These 5.7 FTE are divided among about a dozen subtasks (second-level WPs) providing individual software modules and their operation/maintenance and assessment of results. Thus just 0.5 FTE is assigned to each of these on average.

Gaia DPAC WP:		GWP-M-350-00000
Title: IDT: Management, Implementation, Operation		
Provider(s): UB/ESAC		
Manager(s): J. Torra, J. Hoar		
Start: Phase B	**End**: Phase E2	**Total Effort**: 1125 MM

Objective:

The objective of this work package is the design, implementation, delivery and operation of the Initial Data Treatment (IDT) system. The IDT system will carry on the initial reduction of science telemetry to produce Raw and Intermediate data suitable for processing by the main scientific processing tasks located at the DPCs, and by the First Look.

IDT must process data from all Gaia instruments, combined with auxiliary data such as attitude and orbit data, instrumental calibration data and other data items. This requires a diverse collection of algorithms which must be able to generate a consolidated science dataset from data received at the SOC under potentially non-nominal on-board collection and transfer scenarios (including reception of data unordered in time).

IDT being a classical pipeline process, the scientific algorithms will be contributed by developers distributed across Europe. Therefore the key design decisions focus on an architecture which is both robust in operation and allows developers to implement their algorithms independently with a stream-lined integration process.

An important part of this work package is the coordination of the scientific tasks of design, development, testing and provision of the corresponding modules needed to perform the Initial Data Treatment (IDT). Several specific processes (controiding, cross-matching, etc.) involved in the IDT need inputs from processes created by other CUs, like fluxes or background. All these algorithms must be organized, and links and interfaces must be established in order to get the desired results.

The IDT software will be implemented following an evolutive approach. Starting from a simple model performing the basic and more relevant operations (i.e centroiding, cross-matching), successive versions will incorporate detailed and more sophisticated algorithms (i.e moving objects, SM centroiding, etc). Each IDT version must be carefully tested with data coming from the GASS simulator to ensure its correctness.

Tasks:
This work package covers both the technical aspects of IDT implementation, and the corresponding scientific aspects. The specific tasks are outlined below. The closely connected IDT/FL database is treated in a separate workpackage.

- System Definition
  This task encompasses all activities needed to formally define the large and complex software product that is the IDT system and start the development process. This includes: requirements gathering, choosing and outlining the architectural design, software development tools and practices, quality assurance and software project management approach. The task result will be the design of the full IDT process for the three instruments SM/AF, BP/RP and RVS as well as their required interfaces, taking into account the Overall System Architecture as well as other conditions set by the telemetry downloading process and MOC Operations and by the scientific requirements for the mission.

- MOC Interface
  Implementation of the IDT interface to the MOC. This interface will be responsible of accessing the telemetry and auxiliary data stored at the MOC and populating the IDT Database with incoming data in a form suitable for use by the IDT Framework.

- IDT processing framework
  The implementation of the IDT processing framework which will host the scientific algorithms. This implementation should foresee provision of a testbed system which allows algorithm developers to test the implementation of their algorithms independently of the full, operational IDT system.

- Integration and Validation
  This can be divided into two separate tasks: the integration of delivered scientific algorithms into the IDT framework and the overall integration of the IDT system. Thorough validation of the system as a whole is required with prompt feedback of non-conformances to the algorithm developers through the supplied issue tracking system.

Tasks:

- Operation
  Operation of the IDT system during the development and operations phases, in order to supply other data processing systems (via the Main Database). Simulated data will be used during development and real satellite data during the operations phase.
- Maintenance
  Maintenance of the IDT system through the entire mission lifetime, incorporating fixes to non-conformances and additional functionality required by software change requests.
- Coding of modules
  The implementation of the IDT will require the coding of several modules performing the functionalities identified in the phase of design. Some of the modules will be produced by CU3 and some other by external people. The aim of this package is to produce the first ones and coordinate the tasks for the external ones.
- IDT - GASS tests
  The IDT algorithms should match the implementation of the instruments and mission operation as set in the GASS simulations. Then it is very important that in-depth tests are performed in order to check its conformity.
- IDT modules testing
  IDT scientific modules must be individually tested before integrating in the whole IDT system. A testbed providing input data produced by other IDT modules or by the Gaia simulator must be provided.

Input:
List of inputs to the packages for GWP-M-320-00000, plus:

- IDT Interface Control Document
- MOC-SOC Interface Control Document
- Telemetry specification and model
- Requirements from other WPs in CU3 (WP-M-320, WP-M-345, WP-M-360)
- Requirements from CU4, CU5 and CU6

Output:
The WP shall produce an operational IDT capable of meeting the objectives (see above).

Deliverables:
The following deliverables are foreseen:

- All source code representing the operational IDT and IDT testbed — intermediate deliveries (according to milestones) reflecting increasing algorithmic complexities will be made up to and beyond launch
- Regular progress reports
- ECSS-compliant documentation (Requirements + Design documents, User documentation)
- A data model for SM and AF raw and intermediate data
- A data model for BP/RP and RVS raw and intermediate data
- Test plan document
- Test results

Dependencies:

- Being the interface between the satellite and the rest of the data processing systems, IDT depends on the details of the space segment and requires an accurate description of its characteristics and features. During the development phase, the IDT task is completely dependent on simulated data from CU2.
- The IDT relies on the telemetry data downloaded and feeds the rest of the Gaia scientific processing
- The IDT development must be matched by the GASS simulator development
- The IDT development depends on the modules provided by other CUs

Interfaces:
The IDT is the primary interface of the DPAC to the MOC and populates the Raw Database and Main Database (MDB) with new data on a regular basis, as defined by the Main Database ICD.
The interface to the MDB is organized through a dedicated IDT/FL database, where IDT stores its results.
Further interfaces exist to the FL processes, to CU5 and CU6, and to GASS.

Remarks:
The required effort for this WP, 1125 man-months, corresponds to an average of 8.6 FTE over the period 2007–2019. These 8.6 FTE are divided among more than a dozen subtasks (second-level WPs) providing individual software modules or system components and their operation/maintenance and assessment of results. Thus just about 0.5 FTE is assigned to each of these on average.

Gaia DPAC WP:		GWP-M-360-0000
Title: FL: management, implementation, operation		
Provider(s): ZAH/ESAC		
Manager(s): S. Jordan, J. Hoar		
Start: Phase B	**End**: Phase E2	**Total Effort**: 742 MM

Objective:
To plan, design, implement, deliver, maintain, and operate a First Look (FL) system for Gaia which provides payload monitoring and calibration functions. The main goal of First Look is to perform a judgment of the quality of the Gaia data on a daily basis in order to assure that the specified accuracy of the Gaia mission can be achieved, in particular for the astrometric measurements. During the commissioning phase and the main mission the FL will provide tools to optimise the overall performance of the instruments. The FL will perform detailed diagnoses on representative samples of the Gaia images.
Moreover, the FL will provide

- a daily calibration of the Astro instrument with a high (μas) accuracy in along-scan direction and a varying, lower accuracy across-scan.
- a daily high-precision calibration of the photometry, PSF, radial velocity, and CCD properties of all Gaia instruments.
- an update of the on-ground attitude (OGA1) used for the initial cross-matching. This three-axis attitude (OGA2) has an internal consistency of about 10 μas along-scan and about 0.5 mas in across-scan direction.
- a catalogue with relative source positions much better (0.1 mas along-scan, 1 mas across-scan) than any input catalogue at the time of the Gaia measurements.
- a daily assessment of the overall performance of Gaia's payload.
- trend analysis of the performance of Gaia's payload over time.
- in particular a support for the calibration of radiation effects in Gaia's CCDs which is extremely crucial for the whole mission.

These deliverables can be used as a "clean" input for the global astrometric solution.

Tasks:
This work package covers the management, implementation, and the operation of the First Look. It also includes the contact of the FL to the SOC.
Technical tasks:

- System Definition
  The task encompasses all activities needed to formally define the system as a major and complex software product and start the development process. This includes: requirements gathering, choosing and outlining the architectural design, software development tools and practices, quality assurance and software project management approach.

- Technical requirements
  In the early phase the FL supports the definition of requirements for the on-board software in order to be able to perform diagnostics which cannot be derived from the science data alone.

- IDT/FL Database Interface
  Define the interface to the IDT/FL Database, which provides a temporary shared storage area for use by the IDT and First Look systems during their processing activities. The FL will need a selection of windows for producing diagnostics based on image parameters as well as elementaries for sources smoothly distributed in time (several 10^6 transits per day).

- Operation
  This consists of the operation of the FL components located at ESAC during mission operations. During this period the FL must routinely process scientific data treated by the IDT system to produce diagnostic and calibration data. The Ground Segment contact is performed through the SOC.

- Maintenance
  Maintenance of the First Look system through the entire mission lifetime, incorporating fixes to non-conformance and additional functionality required by software change requests.

Tasks:
Scientific tasks:

- Algorithm definition
  This task shall define the mathematical models and numerical methods needed to judge the health of the scientific instruments. This includes the development of algorithms to judge the performance of the instruments on the level of CCD windows and of a One-day Astrometric Solution (ODAS) which will allow to judge the Gaia data on the μas level. A detailed analysis of the error budget will be calculated and used for the comparison with the residuals from the ODAS. The baseline for an ODAS is a direct astrometric solution (Ring Solution) which simultaneously solves for the source positions, the attitude, and the geometric calibration for the data of about one day.

- Non-astrometric algorithms
  This task will coordinate the overall FL task. For this purpose algorithms are needed from CU5 and CU6. These comprise the daily high-precision calibrations of the photometry, PSF, radial velocity, and CCD properties of all Gaia instruments. All algorithms will be integrated into one software package.

- Software design
  This task covers the design of the different FL software packages within the framework of the overall Gaia data processing. The whole design will aim at a maximum compatibility with the AGIS software so that many routines can be re-used for the ODAS.

- Algorithm development
  This task covers the development of a prototype software in which the functionality can be locally tested. The software will then be integrated into the data processing chain at the SOC, with major revisions foreseen about every half year. During the commissioning phase and during the main mission, updates will be performed when needed.

Tasks:

- Development of a FL monitor/evaluator
  The results of the FL algorithms have to be interpreted in order to judge the functioning of the instruments. Strong departures from the expected values should be detected and information to the First Look Scientist sent automatically. Smaller deviations should be investigated in order to better understand the instruments' performance. For these purposes software tools will be developed with a graphical user interface and real-time generated scientific and technical plots.
- Development of strategies to optimise the operation
  The FL diagnostics should also be used in order to optimise Gaia's performance by changing on-board parameters that can be adjusted from ground. Therefore, strategies will be developed in order to derive the optimum parameters for Gaia during the mission. Software will be developed in order to assist this process.
- Specification of Simulations
  This task covers the specification of the needs for simulations which should be coordinated with the needs of AGIS so that the same data sets can be used. Additionally some tailor made simulations for larger deviations from the expected performance of the instruments will be requested in order to test the contingency cases.
- Development of software tests
  In order to test the software packages the different modules will be tested individually before they will be integrated into the overall processing system.
- Support for the calibration of radiation effects in Gaia's CCDs
  One of the major obstacles for the Gaia mission is the difficulty to correct for the radiation effects of the CCDs on a short time basis. It will be necessary to model in detail centroid shifts (biases) as a function of signal level and time and other systematic effects induced by the strong radiation close to solar maximum. Since the ODAS will provide a daily astrometric calibration with a high along-scan precision, this task may help to correct for those defects as well as to optimise the strategy to technically minimize these shifts by charge injections.

Tasks:

- Algorithm tests
  The verification of the software on integration level will always be performed with noise-free simulated input data which should always result in sub-μas updates. In order to perform realistic test cases we will use noisy and systematically altered input data. These tests will be coordinated with similar tests of the AGIS.
- The FL will during the commissioning phase provide a team of First Look Scientists to run and judge the FL procedures and to communicate problems and suggestions for optimisation to the SOC operation team.

Input:
List of inputs to the package as for GWP-M-320-00000, plus:

- Main Database Interface Control Document defining the Main Database input to FL
- Common calibration model
- Detailed industrial documents
- Detailed studies of the behavior of the CCDs with respect to radiation damages and charge injection strategies
- Algorithms and software packages for the calibration of the photometry, the PSF, the radial velocities, and the CCDs
- Every day during the mission fast access to a sufficient, representative and timely well distributed data sample from the previous 24 hours.

Output:
The WP shall produce an operational FL capable of meeting the objectives (see above) - planned deliveries are listed below

Deliverables:
The following deliveries are foreseen:

- Source code for the FL to be installed at ESAC.
- Test case for the FL software
- Input to the Main Database ICD defining the output of FL
- The software is foreseen to be updated twice a year.
- Regular progress reports
- ECSS-compliant documentation (Requirements + Design documents, User documentation)

Dependencies:
The FL very much depends on the functioning of the IDT and the rest of the data processing systems. It also depends on simulation data from CU2 and many software routines from AGIS. A FL is also not possible without a very detailed description of the Gaia instruments and the on-board parameters that have influence on the performance. During the mission the FL strongly depends on a prioritization scheme for the measured data which provides fast access to a 24 hour sample to run the ODAS and other diagnostics.

Interfaces:
The FL will populate the Main Database with the daily processed data (via the IDT/FL Database), which can afterwards be used for the global astrometric solution.
The FL uses data provided by IDT.
The FL will proposes operation requests for the MOC to the Payload Manager at the SOC.

Remarks:
The required effort for this WP, 742 man-months, corresponds to an average of 5.7 FTE over the period 2007–2019. These 5.7 FTE are divided among the following subtasks (second-level WPs):
GWP-T-360-10000: Management and planning
GWP-T-360-11000: System Definition
GWP-T-360-12000: Technical requirements
GWP-T-360-13000: IDT Interface
GWP-T-360-14000: MDB Interface
GWP-T-360-15000: Maintenance
GWP-T-360-21000: Astrometric Algorithm definition
GWP-T-360-22000: Coordination of the non-astrometric algorithm definition
GWP-T-360-23000: Specification of Simulations
GWP-T-360-24000: Software design
GWP-T-360-31000: Development of ODAS algorithm
GWP-T-360-32000: Development of a FL monitor/evaluator
GWP-T-360-33000: Development of strategies to optimise Gaia's performance
GWP-T-360-34000: Software module tests
GWP-T-360-35000: Software integration tests
GWP-T-360-41000: ODAS support for AGIS
GWP-T-360-42000: Support for the calibration of radiation effects in Gaia's CCDs
GWP-T-360-52000: Running the FL Evaluator
GWP-T-360-53000: Support to Comissioning Phase
Thus just 0.30 FTE is assigned to each of these on average. Some contain a development part and an operations/maintenance part.

Gaia DPAC WP:		GWP-M-370-00000
Title: Manage and Implement IDT/FL Database		
Provider(s): ESAC		
Manager(s): J. Hoar, J. Hernandez		
Start: Phase B	**End**: Phase E2	**Total Effort**: 24 MM

Objective:
The purpose of the IDT/FL Database is to provide short and medium term storage of the input and output datasets of the IDT and FL systems.
The IDT/FL Database stores data used on a day-by-day basis by the IDT and FL systems in operation. Data from this database will be extracted on a daily basis for ingestion into the Gaia Main Database, Raw Database and sent to the DPCs. The IDT/FL Database also stores reference data necessary for the IDT and FL systems.
The IDT/FL Database will be located at ESAC.

Tasks:

- System Definition
  The task encompasses all activities needed to formally define the IDT/FL Database and start the development process. This includes: requirements gathering, choosing and outlining the architectural design, software development tools and practices, quality assurance and software project management approach.
- IDT/FL Database
  Design and implementation of the IDT/FL Database, providing a temporary shared storage area for use by the IDT and First Look systems. This shared storage area may be a combination of file system and database technology.
- Management Tasks
  This includes a) a task to manage the reference data in the database, retrieving new data from the Main Database as needed and removing obsolete data and b) a task to extract newly processed data according to the Main Database ICD ready for distribution to the DPCs and for ingestion into the Gaia Main Database.
- Operation
  Operation of the IDT/FL Database in support of its client systems during the development and operations phases.
- Maintenance
  Maintenance of the IDT/FL Database system through the entire mission lifetime, incorporating fixes to non-conformances and additional functionality required by software change requests.

Input:

- Main Database ICD
- First Look ICD

Output:
IDT/FL Database System and Software

Deliverables: The following deliverables are foreseen: • All source code representing the operational IDT/FL Database system • Regular progress reports • ECSS-compliant documentation (Requirements + Design documents, User documentation)
Dependencies: Main Database and Raw Database data model
Interfaces: Main Database, Raw Database, IDT and FL systems

C.5 Top-level Work Packages of CU4

Gaia DPAC WP:		GWP-M-401-00000
Title: Management and scientific coordination		
Provider(s): ULB, CNES		
Manager(s): D. Pourbaix		
Start: Phase B	**End**: Phase F	**Total Effort**: 27.1 MM
Objective: Manage the Coordination Unit, define work packages, set schedules, follow up on tasks, interface to other coordinating units and to the the DPACE.		
Tasks: 1. Define the requirements for CU4 of Gaia and set priorities 2. Define the work breakdown structure, i.e. the work packages 3. Allocate groups from the community to the work packages 4. Set milestones and schedules 5. Monitor progress 6. Oversee test runs and define analysis/interpretation protocols 7. Interface to other CUs 8. Report back to the DPACE		
Input:		
Output:		
Deliverables: progress reports on CU4		
Dependencies:		
Interfaces: to all other CUs		

Remarks:

Gaia DPAC WP: GWP-T-402-00000

Title: Architecture and technical coordination CU4

Provider(s): CNES

Manager(s): T. Levoir

Start: Phase B	**End**: Phase D2	**Total Effort**: 47 MM

Objective:
This WP has 2 main objectives:

- Define System architecture :
  The Define System architecture activities includes a definition phase and a design phase. It is dedicated both to the architecture of the software to be developed to fulfill all the requirements and to the data model to be managed at DPC level. The CU4 system software is structured in components, with identification of the dependencies between them, the common software,. . .
  This design phase also identifies the components developed by laboratories to be delivered to (and operated by) the DPC and the components to be developed by the DPC (components of the host framework).
- System administration : Install and maintain hardware (for development, integration, and operation phases) and development tools.

Tasks:
The main tasks of this WP are:

1. system definition.
2. System design.
3. System administration.

Input:
CU1 activities results (Overall system architecture, operation schedule, assurance quality, system ICD,. . .).
Estimations of computer and data storage resources (provided by each development unit in charge of a component delivered to DPC).

Output:
- Functional analysis document
- Software System Specifications
- Interface Control Documents (or inputs to CU1 ICD describing interfaces between MDB and CU4 system)
- Software Design Document
- Dedicated technical notes (studies results, choices justifications,. . .).

Deliverables:
See above - output

Dependencies:
Overall system definition

Interfaces:
With CU1 and CU3 activities and choices.

Remarks:

Gaia DPAC WP:		GWP-T-403-00000
Title: Quality assurance and config management for CU4		
Provider(s): CNES		
Manager(s): T. Levoir		
Start: Phase B	**End**: Phase D2	**Total Effort**: 31.9 MM

Objective:
- Define and organize the software quality assurance on behalf to development units
- Define and organize the configuration management on behalf to development units

Tasks:
The main tasks of this WP are:

1. Derive a product assurance plan for the CU4 from the general product assurance plan provided by CU1
2. Organize and manage reviews during the development cycle (phases B, C,D) on documentation and software
3. Support labs to apply quality assurance rules and recommendations during scientific software development and control their application.
4. Define the configuration management plan in conformance with the rules defined in the product assurance plan
5. Manage configuration and releases in the framework (to be defined with CU1)
6. Manage issue, anomaly, change with the support of a tool (to be defined with CU1)
7. Manage and organize Configuration Control Board activities and contribute for technical sides.

Input:
product assurance plan from CU1

Output:
CU4 product assurance plan, review reports

Deliverables: Task output - see above
Dependencies: With all DUs
Interfaces: With CU1
Remarks:

Gaia DPAC WP: GWP-T-404-00000

Title: Integration, Validation and Operation of CU4 system		
Provider(s): CNES		
Manager(s): T. Levoir		
Start: Phase B	**End**: Phase F	**Total Effort**: 163 MM

Objective:

- Simulation and test data management
- CU4 System Integration and Validation at DPC
- Optimise CU4 system
- CU4 system operations and monitoring
- Maintain the CU4 system

Tasks:
The main tasks of this WP are:

1. Define simulation data requirements (in conjunction with CU2) and test, verify received data
2. Define test plans and carry out subsystem integration and validation at the DPC according to test plans.
3. Optimize the sub system during the software development, integration and validation phases.
4. Operate and monitor the subsystem at CNES DPC.
5. Maintain all software components.

Input:

Output:

Deliverables: Task output - see above
Dependencies:
Interfaces: With DU managers and developers for software corrections /changes. With CU1 for end-to-end test at the GAIA system level.
Remarks: Lab efforts in test definition and testing will be identified in dedicated DUs WPs.

Gaia DPAC WP:		GWP-T-405-00000
Title: CU4 Host Software Framework development		
Provider(s): CNES		
Manager(s): T. Levoir		
Start: Phase B	**End**: Phase F	**Total Effort**: 195 MM

Objective:
Develop and validate the Host Software Framework

Tasks:
The host software framework is required :
- to monitor activations of each scientific software (provided by the DUs), managing their dependencies into workflows ;
- to manage their input/output, including interfaces with the Gaia Main Database (at ESAC)... for that, a data access layer will be implemented ;
- to manage resource allocation (CPU, network,...).
The main tasks of this WP are:

1. Gather detailed software requirements (functions, performance, quality, tests,...).
2. Follow up sub contracted development (including detailed design, component development and integration phases) and software reviews.
3. Integrate and validate the framework on a reference platform at CNES (Toulouse).

Input:
for Task 1 above: GWP-T-402-10000 output.
for Task 2 above: ITT document, software requirement specifications for the host software framework.
for Task 3 above: Framework software product, purchased reference platform, scientific software units in test, simulated data.

Output:
from Task 1 above: Software Requirement Specifications for the host software framework.
from Task 2 above: Product documentation (Software design document, Integration plan,. . .), Host software framework product.
from Task 3 above: Tests results, non conformance reports, change requests,. . .

Deliverables:
Progress reports on the framework development

Dependencies:
With GWP-T-402-10000

Interfaces:
TBD

Remarks:
- Framework development process could be incremental to define as soon as possible the framework design for its main functions.
- Host software framework product should be common for all the systems hosted by CNES DPC (spectroscopic processing, object processing, astrophysical parameters).

Gaia DPAC WP:		GWP-M-430-00000
Title: Management and implementation of Non Single Star processing		
Provider(s):		
Manager(s): D. Pourbaix		
Start: Phase B	**End**: Phase F	**Total Effort**: 21.0 MM
Objective: Coordinate the activity concerning Non Single Star data reduction by supervising the structure of the dataflow, task and subtask definition. Manage the interactions with the CU4 manager and the other CUs, maintain an up-to-date documentation of the software produced to accomplish the different tasks, and define the standards allowing the optimization of software interactions (data exchange format, common tools, conventions, etc.).		
Tasks: M-430-00100 Task and subtask definition M-430-00200 Interaction between tasks M-430-00300 Schedule Management I-430-00400 Interaction with CU4 manager and other CUs M-430-00500 Revision and integration of the documentation M-430-00600 Definition of standards for software development (in relation with CU1)		
Input:		
Output:		
Deliverables:		
Dependencies:		
Interfaces:		

Remarks:

Gaia DPAC WP:		GWP-M-432-00000
Title: Astrometric Binaries		
Provider(s): H.H. Bernstein, J.-L. Halbwachs and D. Pourbaix		
Manager(s): J.-L. Halbwachs		
Start: Phase B	**End**: Phase F	**Total Effort**: 94 MM

Objective:
This DU concerns the objects having individual transits similar to those of single stars, that are not moving fast like solar system objects, and for which the single star astrometric solution is not accepted. It calculates astrometric solutions assuming that the star has a companion, using various types of two-body models, which are tried one after one until an acceptable solution is obtained. When no two-body solution is accepted, a stochastic solution is provided.

Tasks:
The work packages of the DU are listed hereafter.

- **M-432-00100: Preprocessing.** All the quantities (scalars and vectors) subsequently required by any WP of the DU is computed here. The calculation of the μ_r (the secular parallax) is achieved in this WP, a single star solution with μ_r accounted for is therefore evaluated. In case no radial velocity is available, this operation is bypassed.
- **M-432-00300 Acceleration Solutions.** Astrometric solutions are calculated adding acceleration terms. This includes two processes:
- The 2-terms acceleration solution, which includes the first derivatives of the proper motion.
- The 4-terms acceleration solution, where the second derivatives of the proper motion are added.
- **M-432-00400: Astrometric orbits.** Astrometric orbits may be derived when the period is less than about twice the duration of the scientific mission of the satellite. The 7 parameters of the astrometric orbits are added to those of the single-star astrometric solution.
- **M-432-00500: Variability-Induced Movers.** It is assumed that each object is an unresolved binary with a photometrically variable component. The photocentre of the system is then moving between the components, in relation with the total magnitude. Several cases are considered:
- The position of the variable star with respect to the other is fixed (VIMF), or is only affected by a linear motion (VIML).
- **Variability-induced movers with acceleration (VIMA).** The time derivatives of the 2-D relative motion of the variable star are added. As for the acceleration solutions, the first and, when necessary, the second derivatives are considered. The model is different from that of the acceleration solution, since it includes a term, called ”h” hereafter, which is related to the mass ratio of the system and to the magnitude of the non-variable star.
- **Variability-induced movers with orbits (VIMO).** The model includes ”h” and the orbital elements of the motion of the variable star around the other one.
- **M-432-00600: Stochastic solutions.** When no solution is accepted in the previous steps, the star is analysed with a suspected stochastic solution with the least squares collocation method. The classical parameter estimation is combined with filtering of a probably coloured noise part in the data (caused by a double star).

Input:

1. The transits of the astrometric field of view, including the epochs, the residues (with respect to the single star solution fitted by CU3), their errors and the partial derivatives of the single star model.
2. A flag indicating that the μ_r parameter must be calculated, and the average radial velocity of the star.
3. A flag indicating that the star is variable, the variability amplitude, and, for each astrometric transit, the G magnitude and the epoch.
4. A flag indicating that a spectroscopic orbit was derived, the elements of this orbit, and the radial velocity measurements.

Output:
The outputs depend on the process which provides the accepted solution. Among all the parameters listed hereafter, only those in the first item are always present.

1. The 5 or 6 fundamental astrometric parameters (6 when μ_r is included), and the goodness-of-fit of the complete accepted solution. A flag indicates the type of the solution and specifies when these parameters refer to the photocentre of the system, and when they refer to the barycentre.
2. The 2 or 4 acceleration terms provided by the acceleration solutions or by the VIMA solutions. These terms always refer to the photocentre of the system.
3. The 7 orbital parameters of the photocentre or of the variable component around the barycentre. The flag in item (1) is used to differentiate these two cases.
4. For VIMF or VIMA solutions: the reference magnitude of the photocentre, the 2-D relative position and the 2-D relative linear motion of the variable component.
5. For VIMA or VIMO solutions: the ”h” parameter.
6.

Additionally to the parameters above, the outputs include their errors and the covariance matrix of the parameters that are coming from a linear system.

Deliverables:
Delivered documents:

- Reduction of the astrometric binaries, Pourbaix D. and Jancart S., DMS-PJ-01
- Astrometric binaries with a variable component, Halbwachs J.L. and Pourbaix D., 2005, ESA SP-576, pg 575
- Gaia treatment of astrometric binaries with a variable component: VIM, VIMA, VIMO, J.L. Halbwachs, D. Pourbaix, IAU-Symposium 240 (in press)

A detailed description of the calculation methods concerning the binaries with a variable component is in preparation.

Dependencies:
A star is injected in the DU when CU3 concludes that the single-star solution is not acceptable. CU3 indicates also if μ_r must be taken into account or not. The VIM process can start when the G magnitudes of the transits are available, and when it is known that the star is variable.

Interfaces:
From CU3: The single star solution (i.e. the five astrometric parameters) together with its residues and some quality indicators (e.g. goodness of fit).
From CU6: Some indication of the variability of the radial velocity. If positive, the result of GWP-M434-0000 (SB orbital elements).
From CU7: The variability amplitude of the variable stars, the G magnitudes of the transits, with the transit epochs.

Remarks:

Gaia DPAC WP:		GWP-M-433-00000
Title: Resolved Multiples		
Provider(s): D. Pourbaix		
Manager(s): D. Pourbaix		
Start: Phase B	**End**: Phase F	**Total Effort**: 98.0 MM

Objective:
To determine the astrometric, photometric and relative astrometric parameters or any source classified as a resolved double or multiple star. It is assumed that approximate positions and magnitudes are known from SM data handling WP, Image analysis WP, CU5 and CU8 analysis.

Tasks:

M-433-00100	Detected multiples, SM Data Handling
M-433-00200	Undetected Resolved multiples
M-433-00300	Relative motions in resolved multiples, Common PM
M-433-00600	Variable multiples
M-433-00700	Close Trapezium System
M-433-00800	Astrometric and/or R.V. orbits in Multiples

Input:

- times, observed fields, sample count from RP, BP and AF1-9, attitudes, PSF calibration data, astrometric calibration data, . . .
- approximate positions and magnitudes for the components (from the SM data handling WP or the Image analysis WP from CU5)
- number of components (n) to be fitted

Output:

- Source model (physical/optical; fixed/linear/curved motions, astrometric parameters of each sources).
- Quality of the fit and, for each component, a set astrometric and photometric parameters with covariance matrix and all relevant statistical indicators.

Deliverables:
Java codes with detailed descriptions of the methods and documentation (how to use them).

Dependencies:

Interfaces:

Remarks:

Gaia DPAC WP: GWP-M-434-00000

Title: Spectroscopic Binaries

Provider(s): G. Rauw, E. Gosset

Manager(s): E. Gosset

Start: Phase B	**End**: Phase D2	**Total Effort**: 13.2 MM

Objective:
This workpackage deals with composite spectra (basically due to multiplicity) of the stars recognized as non-single: spectroscopic binaries essentially. The workpackage is intended to treat RVS data, to output orbital solutions of variable single-transit spectra and information on composite multi-transit spectra.

Tasks:

M-434-00100	Spectrum Binaries
M-434-00200	SB1 orbits
M-434-00300	SB2 orbits

Input:
For GWP-M-434-00200/00300, input is one (SB1) or two (SB2) sets of epoch radial velocities (with error estimates) in barycentric reference frame. They will come from GWP-S-650-12000 where we are also involved. Each observation will correspond to a single transit. For GWP-M-434-00100, we will need the composite barycentric multi-transit spectrum, also in provenance from CU06.

Output:
For GWP-M-434-00200/00300, output will consist in the derivation of adapted SB1 or SB2 orbital solutions. These solutions will be provided under the form of, respectively, 6 or 7 parameters (P, v_0, K_1, e, omega, T_0, [K_2]) fully specifying the radial velocity orbit. The parameters will be delivered with estimated errors and parameter to parameter correlations. Various indicators of the goodness of fit will also be provided as well as the results of an inspection of the residuals for additional periodicities (SB3 and SB4 system candidates). For GWP-M-434-00100, we will designate the best pair of fitting templates, their respective shifts in RV and the deduced luminosity ratio along with the necessary statistical information.

Deliverables:
Each WP will result in a code implementing the best possible method identified after extensive tests.

Dependencies: None recognized
Interfaces: Interface with CU6.
Remarks: Time permitting, other methods working directly on the spectra could be envisaged.

Gaia DPAC WP: GWP-M-435-00000

Title: Photometric Analysis of NSS		
Provider(s): O. Malkov		
Manager(s): O. Malkov		
Start: Phase B	**End**: Phase D2	**Total Effort**: 8.0 MM
Objective: Detection a composite flux in RP and BP and, when possible, contribution to the parametrisation of the components of binary/multiple systems		
Tasks: M-435-00100 Photometric Binaries		
Input: MBP data		
Output: List of detected photometric binaries/multiplies, physical parameters		
Deliverables: Computer files containing the above cited data sets		
Dependencies:		
Interfaces:		
Remarks: This WP will likely merge with GWP-S-834		

Gaia DPAC WP: GWP-M-436-00000

Title: Eclipsing Systems

Provider(s): G. Sadowski, C. Siopis, B. Tingley

Manager(s): C. Siopis

Start: Phase B	**End**: Phase D2	**Total Effort**: 90.0 MM

Objective:
The creation of DPAC-compliant software for the extraction of physical parameters from the light and radial velocity curves of the eclipsing binary stars that will be detected by Gaia.

Tasks:

M-436-00200	Light-Curve Analysis
M-436-00400	Timing Binaries
M-436-00500	Eclipsing Binaries with Variable Component
M-436-00600	Variable Eclipsing Binaries in Multiples

Input:
Light curves in multiple passbands, velocity curve, orbital period, [eccentricity, argument of periastron, argument of the ascending node, mass ratio, parallax, spectral type of primary and/or secondary star, surface temperature of primary and/or secondary star, metallicity of primary and/or secondary star].

Output:
Mass ratio (and sometimes individual masses), relative or absolute luminosities, ratio of radii (and sometimes individual radii and semi-major axis), orbital period, inclination, eccentricity, argument of periastron, argument of the ascending node, synchronicity parameters (stellar rotation), surface temperatures, metallicities, limb darkening coefficients.

Deliverables:
Software and all documentation missing

Dependencies:
CU6 (Spectroscopic Processing) for the velocity curve, orbital parameters (eccentricity, argument of periastron), spectral type(s), surface temperatures, and metallicities; CU7 (Variability Processing) for the light curves, orbital period; CU3 (Core Processing) for the parallax.

Interfaces:
Gaia Catalog (no more information on feedback to other WPs at this moment).

Remarks:
Not all input parameters are required for a solution, and not all output parameters are always produced (depending on input). Some output parameters could be eliminated if they are not needed by other pipeline modules *and* will not be included in the Gaia Catalog.
The minimum requisite input parameter is one light curve. It can, in principle, yield orbital inclination, eccentricity, arguments of periastron and of the ascending node, luminosity ratio, and ratio of radii. If spectroscopic analysis (CU6) provides an estimate for the eccentricity and the arguments of periastron and of the ascending node, these parameters can be used as initial guesses in the eclipsing binary parameter fit.
Availability of light curves at multiple passbands could help determine surface temperatures, the presence of a “third light”, and the presence of extrasolar planets. Surface temperatures, as well as spectral type information, can help reduce the size of parameter space.
The period can be obtained from the light curve (CU7) as well as from the velocity curve (CU6), and can be either used as such or as an initial guess (in which case it is also included in the output). A good period determination is necessary, even if it is allowed to float in the fitting stage.
Knowledge of parallax yields absolute luminosities and hence luminosity classes which, in turn, can be useful for reducing the size of parameter space.
Metallicity is another model parameter to be fit; even a rough estimate of its value can be used as an initial guess.
If one velocity curve (SB1) is available additionally to the light curve, it can yield individual masses, individual radii and the semi-major axis of the binary.
If two velocity curves are available (SB2), individual masses can be deduced independent of photometry, which can be used as initial guesses for the parameter fit.

Gaia DPAC WP:		GWP-M-437-00000
Title: Extrasolar Planets		
Provider(s): A. Sozzetti, N. Rambaux, D. Ségransan		
Manager(s): A. Sozzetti		
Start: Phase B	**End**: Phase F	**Total Effort**: 150 MM

Objective:
Characterize the likely nature of a companion, assess the detection of a planet candidate, search for and characterize additional components, investigate the possible existence of strong mutual gravitational interactions in multiple-planet systems and establish the degree of dynamical stability of the detected systems

Tasks:

- M-437-00100: Single-Component Solutions
    - M-437-00110: Companion Characterization
    - M-437-00120: Assessment of Planet Candidate Detection
- M-437-00200: Multiple-Component Keplerian Orbital Solutions
    - M-437-00210: Multiple-Planet Orbital Fits (All Periods Determined)
    - M-437-00220: Multiple-Planet Orbital Fits (with Acceleration Terms)
    - M-437-00230: Multiple-Planet Orbital Fits (Independent Assessment: Genetic Algorithms)
- M-437-00300: Non-Keplerian Orbital Solution & Stability Analysis
    - M-437-00310: MEGNO Stability Analysis
    - M-437-00320: N-Body Integration & MEGNO Stability Analysis

Input:
Astrometric binary solution (12 parameters) + time-series of observations + multiple statistical indicators of the goodness of fit + photometric and radial-velocity variability information

Output: A measure of the likelihood of the first detected companion being a planet, and of the statistical robustness of the single-component solution. Multiple-component solution (5+7*N parameters, eventually acceleration terms), including covariance matrix, χ^2, multiple statistical indicators of the robustness of the solutions, and a measure of their relative agreement. Indicators of the degree of dynamical interaction and of the system long-term stability.
Deliverables: Well-documented software codes performing the above mentioned tasks and subtasks
Dependencies:
Interfaces: Use of the astrometric binary solution from M-432-00200
Remarks:

Gaia DPAC WP: GWP-M-438-00000

Title: Simulated Test data for NSS processing

Provider(s): F. Arenou and the other NSS WP managers

Manager(s): F. Arenou

Start: Phase B	**End**: Phase D2	**Total Effort**: 15.0 MM

Objective:
Simulated data are produced to fulfill the specific needs of the various Non Single Stars tasks, and adapted to the validation of these tasks. The simulations should ensure a realistic distribution of the parameters of the various NSS objects and will thus make use of the tools developed within CU2. However, input for these tools from the other NSS DU managers is needed, e.g. for close binaries. This leads to the development of dedicated simulations which are then included inside the CU2 overall chain. Finally, the testing data is defined, produced, and delivered to the WP managers.

Tasks:

M-438-00100	Contribution to the Gaia Simulator
M-438-00200	Simulated data for the NSS WP managers
M-438-00300	Production of test data for algorithm validation

Input:
Instrument and mission parameters, Galaxy model generation with the stellar associated methods in CU2.

Output:
Simulated data used as inputs by the different Non Single Stars tasks.

Deliverables:
Simulated observations of specific Non Single Stars in the Gaia instruments.

Dependencies:
CU2 simulation methods for stellar objects and instrument.

Interfaces:
CU2 and the other NSS WPs

Remarks:

Gaia DPAC WP:		GWP-M-439-00000
Title: NSS solution combiner		
Provider(s): E. Gosset, D. Pourbaix, C. Siopis		
Manager(s): D. Pourbaix		
Start: Phase B	**End**: Phase F	**Total Effort**: 6.0 MM
Objective: Once several solutions have been derived depending on different data sets, this DU checks their consistency (quality control of the solutions) and, when positive, derives a simultaneous solution in order to improve the quality of the original fits		
Tasks: M-439-00100 Joined astro+spectro orbit M-439-00200 Eclipsing spectroscopic binaries		
Input: Same as DU432, DU434, and DU436 + their solutions		
Output: Same as DU432, DU434, and DU436 but the correlation matrix copes with all the parameters fitted simulatenously		
Deliverables: Java codes with detailed descriptions of the methods and documentation (how to use them).		
Dependencies: This DU is only activited if at least two solutions are availables for the same object		
Interfaces: Same as DU432, DU434, and DU436		
Remarks:		

Gaia DPAC WP: GWP-M-450-0000

Title: Management and implementation of Solar System Object processing

Provider(s): F. Mignard, A. Cellino, K. Muinonen, P. Tanga

Manager(s): P. Tanga

Start: Phase B	**End**: Phase E1	**Total Effort**: 21.0 MM

Objective:
Coordinate the activity concerning Solar System data reduction by supervising the structure of the data flow, task and subtask definition. Manage the interactions with the CU4 manager and the other CUs, maintain an up-to-date documentation of the software produced to accomplish the different tasks, and define the standards allowing the optimization of software interactions (data exchange format, common tools, conventions, etc.).

Tasks:

1. M-450-00100 Task and subtask definition
2. M-450-00200 Interaction between tasks
3. M-450-00300 Schedule Management
4. I-450-00400 Interaction with CU4 manager and other CUs
5. M-450-00500 Revision and integration of the documentation
6. M-450-00600 Definition of standards for software development (in relation with CU1)

Input:

Output:

Deliverables:

Dependencies:

Interfaces: Activity developed in close collaboration with the CU4 manager, the other CU leaders, and the WP leaders of Solar System processing.
Remarks:

Gaia DPAC WP: GWP-M-451-00000

Title: Auxiliary data for SSO processing

Provider(s): J. Berthier, K. Minonen

Manager(s): J. Berthier

Start: Phase B	**End**: Phase D2	**Total Effort**: 7.2 MM

Objective:
Provide the solar system object reference data and their update with new ground based data and data resulting from the data processing. At the time Gaia will fly, the vast majority of Solar System objects will be known and characterized by an orbit sufficiently precise to be useful for identification purposes in the Gaia dataset. It will provide to other WP an up-to-date set of data for the different object categories. Updating will require the access (with a frequency still to be determined) to different data centers, and the development of specific software tools.

Input:
The up to date reference data (orbital, physical) of the small solar system bodies, comets and natural satellites

Output:
Database containing the data as given by the task list

Tasks:

M-451-00100 Orbital elements for minor planets and comets

M-451-00200 Orbital elements for natural satellites

M-451-00300 Physical parameters (absolute magnitudes, size, rotation period, pole, etc.)

M-451-00400 Ephemeris of minor bodies

M-451-00500 Radial velocity for RVS calibration

Deliverables:
Software tools to access the data and extract computer files containing the above cited data sets

Dependencies:
Interfaces: GWP-S-335-17300 Physical data for minor planets
Remarks:

Gaia DPAC WP:		GWP-M-452-00000
Title: Solar System objects cross matching		
Provider(s): J. Berthier		
Manager(s): F. Mignard		
Start: Phase B	**End**: Phase D2	**Total Effort**: 7.2 MM

Objective:
Associate the provisional tag assigned to observations of suspect solar system objects to already known sources, when possible. This task takes as input the coordinates and detection epochs for sources that have failed the ”cross matching” inside IDT, and compares them to the ephemerides of known Solar System objects, taking into account the uncertainty associated both to source coordinates at IDT output and to the available ephemerides.

Input:

1. The intermediate data processed in the IDT of observations at each transit for the sources which have failed the standard cross-matching in the IDT process
2. The most of up to date file of orbital elements of the minor planets, including those identified by Gaia when appropriate
3. orbital elements, or precomputed ephemeris, of satellites of major planets

Output:

1. ID numbers of the solar system bodies when the search ends successfully
2. provisional ID numbers or flag for the unmatched observations

Tasks:

M-452-00100 Update of orbital elements (or any other convenient approach to the ephemeris computation) to the most suitable epoch

M-452-00200 Cross-check with computed positions of known objects

M-452-00300 Assignment of final ID numbers to the cross-matched sources

Deliverables:
ID numbers for all identified sources or flagged observations when no successful cross match is found

Dependencies:

Interfaces:
M-451-0000 : Auxiliary data

Remarks:
The numbering scheme for moving objects must be decided by the DPACE. Observations associated to a known objects are automatically threaded together.

Gaia DPAC WP:		GWP-M-453-00000
Title: CCD processing for SSO observations		
Provider(s): A. Dell'Oro, D. Hestroffer, A. Cellino		
Manager(s): A. Dell'Oro		
Start: Phase B	**End**: Phase D1	**Total Effort**: 35.0 MM

Objective:
Processing at the transit level of the astrometric and photometric data, by fitting the AF, BP and RP signal with a suitable model of the object. The model should take into account size, shape, orientation and movement. Most physical models (especially concerning rotation and shape) will become available toward the end of the mission. This task will thus activate modules of growing complexity. The signal will also be analyzed in order to identify hints of cometary activity or faint, undetected (V>20) companions of the detected source. Depending upon windowing and sampling choices, imaging of extended sources can also be implemented.

Input:
CCD raw counts, CCD calibration files, color-dependent PSFs, photometric calibrations, smoothed out background

Output:
Calibrated centroids and magnitudes for each chip crossing with accurate timing. Preliminary size estimate (sky-projected). Hints of cometary activity. Apparent motion on the focal plane.

Remarks:

Deliverables:
Software routines for the treatment of the signal with output centroids, magnitudes, motion and preliminary sizes.

Dependencies:
GWP-M-452-00000 “Detection of moving objects” calibrated magnitude, PSF from CU5

Interfaces:
M-452-00000, M-454-00000, M-455-00000, M-456-00000, M-458-00000.

Tasks:

M-453-00100	PSF-based centroiding and timing
M-453-00200	Motion measurement
M-453-00300	Preliminary size estimate
M-453-00400	Magnitude estimate in G-band
M-453-00500	Magnitude estimate in Blue and Red bands
M-453-00600	Detection of anomalies in the astrometric signals suggesting hints of cometary activity
M-453-00700	Detection of anomalies in the astrometric signals suggesting the existence of undetected companions surrounding detected asteroids

Gaia DPAC WP: GWP-M-454-00000

Title: Astrometric Reduction for SSO

Provider(s): J.-E. Arlot, T. Pauwels, P. De Cat

Manager(s): J.-E. Arlot

Start: Phase B	**End**: Phase D2	**Total Effort**: 10.8 MM

Objective:
Astrometry of the moving source on the sky based on the centroiding, attitude and reference system. This processing is independently performed over all Solar System detections.

Input:
Centred position in each CCD from task M-453-00000, attitude, system definition, timescale calibration files, physical parameters of the sources (photocentric shift).

Output:
Gaia-centered astrometric position (1D or 2D ?) and covariance matrix.

Remarks:
Iterations: Yes, compelling with updated system and attitude, orbit, colors, size and shapes. This processing is independently performed over all Solar System detections.

Deliverables:

Dependencies:

Interfaces:
M-453-00000, M-458-00000

Tasks:

M-454-00200	Conversion of CCD pixel coordinates to focal plane coordinates
M-454-00400	Conversion of focal plane coordinates to ICRS coordinates
M-454-00600	Computation of corrections to apply for high precision
M-454-00800	Estimate of speed based on a single transit

Gaia DPAC WP: GWP-M-455-00000

Title: Threading of SSO

Provider(s): F. Mignard, A. Morbidelli, J.-M. Petit

Manager(s): J.-M. Petit

Start: Phase B	**End**: Phase D1	**Total Effort**: 10.0 MM

Objective:
Thread together the epoch observations of every source that does not have a previously determined orbit. Criteria that can be used include both dynamical and physical properties.

Input:
Astrometric positions from task M-454-00000 and timing, magnitude and colors

Output:
Individual sources threaded together into a single data vector.

Remarks:
Iterations: Yes, compelling with new observations, improved orbits.

Deliverables:

Dependencies:

Interfaces:
M-454-00000

Tasks:

M-455-00100	Threading by linear extrapolations
M-455-00200	Threading by short arc trajectory
M-455-00300	Threading by preliminary orbits (2-body orbit)
M-455-00400	Threading to Earth-based observations

Gaia DPAC WP:	GWP-M-456-00000

Title: Orbital inversion for SSO		
Provider(s): J.-E. Arlot, J. Berthier, A. Cellino, B. Davidsson, M. Delbó, A. Dell'Oro, M. Fouchard, M. Granvik, D. Hestroffer, S. Klioner, V. Lainey, K. Muinonen, P. Tanga, J. Vaubaillon		
Manager(s): K. Muinonen		
Start: Phase B	**End**: Phase D2	**Total Effort**: 72.0 MM

Objective:
Determine or improve the orbits of solar system objects, in particular, single and non-single asteroids, comets, and planetary satellites. This involves the solution of a dynamical model including all the physics necessary to properly fit the observations by Gaia (relativistic effects, objects shapes and sizes, perturbations, etc.). This module will run at different complexity levels depending upon the timescale: as soon as new data become available as IDT output, orbits are computed with simple dynamical models in order to identify possible candidates for ground-based follow-up. At end-of-mission, all perturbations can be taken into account.

Input:
Astrometric positions from task M-454-00000 with threading from M-455-00000, timing, masses from M-457-00000, sizes from M-453-00000, physical parameters from M-458-00000

Output:
Orbital elements and their uncertainties at a given epoch, nongravitational parameters and their uncertainties. Rapid identification of critical objects in need of immediate ground-based follow-up.

Remarks:
Iterations: with new positions and threading (GWP-M-454-0000 and GWP-M-455-00000), with improved perturber masses (GWP-M-457-00000), with improved photocenter shift (GWP-M-453-00000), with improved spin, shape, and scattering parameters (GWP-M-458-00000)

Deliverables:

Dependencies:

Interfaces:
M-453-00000, M-454-00000, M-455-00000, M-457-00000, M-458-00000)

Tasks:

M-456-00100	Asteroids
M-456-00110	2-body orbits
M-456-00120	N-body orbits
M-456-00130	N+M -body orbits
M-456-00140	N+M -body photocenter-corrected orbits
M-456-00150	N+M -body photocenter-corrected orbits and nongravitational parameters
M-456-00160	Identification of critical asteroids
M-456-00200	Comets
M-456-00210	2-body orbits
M-456-00220	N-body orbits
M-456-00230	N+M -body orbits
M-456-00240	N+M -body orbits and nongravitational parameters
M-456-00250	Identification of critical comets
M-456-00300	Binary and multiple asteroids
M-456-00310	2-body orbits
M-456-00320	N-body orbits
M-456-00400	Planetary satellites
M-456-00410	2-body orbits
M-456-00420	N-body orbits
M-456-00430	Impact on ephemerides of planets
M-456-00500	Simultaneous global inversion for all objects and parameters

Gaia DPAC WP:	GWP-M-457-00000

Title: Global Effects on Solar System Dynamics		
Provider(s): Ph. Bendjoya (PB), J. Berthier (JB), A. Cellino (AC), F. Colas (FC), M. Delbo (MD), A. Fienga (AF), M. Fouchard (MF), D. Hestroffer (DH), S. Klioner (SK), O. Michel (OM), F. Mignard (FM), A. Morbidelli (AM), S. Mouret (SM), F. Vachier (FV), J. Vaubaillon (JV)		
Manager(s): D. Hestroffer		
Start: Phase B	**End**: Phase D2	**Total Effort**: 54 MM
Objective: Determination of global parameters affecting the astrometric observations of all or a subset of the small bodies of the solar system bodies. The dynamical model is taking into account gravitational interactions, relativistic perturbations, and non-gravitational effects.		
Input: Astrometric positions from task M-454-00000, timing and error, 2-body orbit from M-451-00000 (and M-456-00000), identification and threading from M-452-00000 and M-455-00000, masses of binaries from M-456-00000, sizes and other physical parameters from M-451-00000 (M-453-00000 and M-458-00000) for non-gravitational effect and photocenter offset; positions of major planets from M-456-00400; PPN parameter γ for the derivation of β; ephemerides of perturbing bodies (planets and asteroids).		
Output: Initial position and velocity at a reference time (osculating elements) in Cartesian coordinates and/or elliptic elements — masses of perturbing asteroids — parameterization of non-gravitational effects; PPN parameters and solar quadrupole J_2; link of quasi-inertial reference frames — dynamical families — planetary ephemerides.		
Remarks: Iterations: not necessarily; could be performed late during the reduction using the best parameters.		
Deliverables:		
Dependencies:		

Interfaces:
M-451-00000, M-454-00000, M-455-00000, M-456-00000, M-833-00000, M-300-00000 ephemerides of planets (TBC)

Tasks:

M-457-00100	Mass determination for a subset of asteroids from gravitational perturbation during close encounters (SM, FM, DH)
M-457-00110	Searches of relevant close approaches from 2-body orbits
M-457-00120	Setting observation equations from numerical integrations
M-457-00130	Deriving initial conditions of targets and masses of perturbers
M-457-00200	Non gravitational effects on asteroids and comets (MF, MD, JV, FC)
M-457-00210	Setting observation equations from numerical integrations
M-457-00220	Including and deriving the effect on outgassing comets
M-457-00230	Including and deriving the Yarkovsky effect on small asteroids
M-457-00300	Link of quasi-inertial reference frames (DH, SM, FM, SK, AF). Rotation and rotation rate of the dynamical reference frame with respect to the Gaia-ICRF
M-457-00400	Local test of General Relativity (DH, SM, FM, SK)
M-457-00410	Setting of relevant targets for test of GR (among small bodies and dwarf planets)
M-457-00420	Derivation of global parameters (PPN β, η, solar J_2, $\dot{G}/G$, etc.)
M-457-00500	Impact on the ephemerides of major planets (AF)
M-457-00510	Ephemerides of major planets including their positions derived from the planetary satellites (456-00400)
M-457-00520	Ephemerides of major planets including the masses of asteroids (457-00100)
M-457-00600	Proper elements and dynamical families (AM, PB, OM, AC)
M-457-00610	Computing proper elements for a subset of asteroids from numerical integration including all perturbative effects and osculating elements from the global solution
M-457-00620	Identification of families from clustering in the dynamical phase space, including the taxonomy (M-833-00000)

Gaia DPAC WP:		GWP-M-458-00000
Title: SSO physical parameters		
Provider(s): A. Cellino, M. Delbó, A. Dell'Oro, D. Hestroffer, S. Mottola, J. Torppa, P. Tanga		
Manager(s): A. Cellino		
Start: Phase B	**End**: Phase D2	**Total Effort**: 55.0 MM
Objective: Determination of the size, shape, rotational properties, albedo, average density (when possible) of minor planets from the inversion of photometric or astrometric data, or a combination of both.		
Input: Threaded photometric data, threaded CCD counts, preliminary orbits, timing, measured masses		
Output: Sizes, rotational parameters, photocentric shifts, albedoes, average densities		
Remarks: Iterations: yes with improved orbits and additional observations		
Deliverables: Software for disk-integrated photometry inversion; software for improved size determination based on input from photometry inversion; software for the estimate of the photocentric shift; software for average density computation		
Dependencies: Threaded asteroid signals and disk-integrated photometry and timing from CCD Processing; distances from Dynamical Modeling; masses from Mass Determination		
Interfaces: M-451-0000, M-453-00000, M-456-00000, M-457-00000		

Tasks:

M-458-00100	Size determinations from PSF and/or LSF convolution and fitting
M-458-00200	Lightcurves inversion for shape, rotation, scattering properties
M-458-00300	Estimate of the photocentric shift for each observation
M-458-00400	Densities from masses, sizes and shapes
M-458-00500	Albedoes from sizes and magnitudes
M-458-00600	Validation of results against *in situ* space probe data.
M-458-00700	Global astrometric and photometric solution

Gaia DPAC WP:	GWP-M-459-00000

Title: Ground based observations

Provider(s):

Manager(s): W. Thuillot

Start: Phase C	**End**: Phase E2	**Total Effort**: 30.0 MM

Objective:
To supplement Gaia observations in specific areas. In particular, three situations will be relevant for ground-based follow up: (1) New Earth-crossers or ”inner Earth” objects that will have few Gaia observations (this will be common due to their orbit configurations). In some cases, without ground-based observations, it will be impossible to usefully constrain their orbit. (2) Asteroids following peculiar orbits that can be improved by extended the observation arc beyond the Gaia operation, even with lower accuracy astrometry; (3) Asteroids having mutual close encounters at beginning or end of the mission, needing additional astrometry in order to improve mass determinations.

Input:
Preliminary ephemeris

Output:
Astrometric positions of followed targets

Remarks:

Deliverables:

Dependencies:

Interfaces:
M-456-00000

Tasks:

M-459-00100	Solar System alerts: very short timescale (± 1day)
M-459-00200	Solar System follow-up: additional observations (± 1week)

Gaia DPAC WP:		GWP-M-460-00000
Title: Simulated Test Data for SSO processing		
Provider(s): P. Tanga, F. Mignard, N. Rambaud		
Manager(s): F. Mignard		
Start: Phase B	**End**: Phase D2	**Total Effort**: 18.0 MM

Objective:
Dedicated simulations for solar system objects developed internally. This is independent as much as possible from the Gaia simulation task that will provide also simulated data to validate the overall processing chain rather than providing customised data to check the individual algorithms with great flexibility. Specific needs of the different WPs will drive the simulation development

Input:
Instrument and mission overall parameters relevant for these test data, digitised Solar System

Output:
Customised simulated test data used as inputs in the different tasks and subtasks

Remarks:
I. An already advanced version of the simulator does exist and provides satisfactory results. However its implementation (in Fortran) has reached its final evolutionary state and must be completely rewritten to include new functionalities.
II. Main steps in the development.

1. Step 1: Definition of the needs and requirements in term of content, kind of sources, accuracy in the modelling.
2. Step 2: Overall structure of the simulation to obtain the sequences of observations for standard minor planets with detailed photometric modelling
3. Step 3: Extension to particular objects like satellite, binary asteroids, NEOs.

Deliverables:

Dependencies:

Interfaces:

Tasks:

M-460-00100	Contribution to the Gaia simulator
M-460-00200	Internal simulations for the SSDPC
M-460-00300	Production of test data for piecewise algorithm validation

Gaia DPAC WP:		GWP-D-470-000000
Title: Extended Objects		
Provider(s): G. Bourda, P. Charlot, C. Ducourant, P. Gavras, A. Krone-Martins, D. Sinachopoulos, G. Medina-Tanco, E. Slezak, R. Teixeira		
Manager(s): C. Ducourant		
Start: Phase B	**End**: Phase D2	**Total Effort**: 18.0 MM
Objective: Determine morphological parameters and improve astrometric parameters for extended objects.		
Input: 2D reconstructed images and characterization from CU5, classification flag (star, galaxy, …) from CU8, astrometric solution with goodness of fit from CU3.		
Output: Improved astrometric solution and morphological parameters for extended objects. Overlapped reconstructed 2D images of large extended objects.		
Remarks:		
Deliverables:		
Dependencies:		
Interfaces:		

Tasks:

M-470 00000	Management and implementation of Extended Object processing
M-470 00100	Scientific analysis of 2D images for the determination of morphological parameters
M-470 00200	Overlap of 2D reconstructed images for large extended objects
M-470 00300	Very compact unresolved Galaxies
M-470 00400	Host galaxies of Quasars
M-470 00500	Structure of Quasars
M-470 00600	Planetary nebula
M-470 00700	HII regions
M-470 00800	Unresolved Globular clusters in nearby galaxies
M-470 00900	Host galaxies of Red Giant Branch stars

C.6 Top-level Work Packages of CU5

Gaia DPAC WP: GWP-M-501-00000

Title: Planning, management, and coordination of CU5 activities

Provider(s): IOA

Manager(s): F. van Leeuwen

Start: Phase B	**End**: Phase D2	**Total Effort**: 65 MM

Objective:
Planning, management and coordination of CU5 activities

Tasks:

1. Plan and coordinate the activities of the CU5 management team
2. Through the management team, plan, coordinate and monitor the activities of the DUs
3. Coordinate the CU5 activities with those of other CUs, where relevant, through the DPACE
4. Through the management team, manage the resources assigned to the DUs
5. Through the management team, collect, monitor and adjust where necessary the CU5 requirements
6. Collect and monitor input for monthly reporting by DU managers
7. Prepare progress reports for Livelink
8. Prepare reports for national funding agencies
9. Organize reviews and other meetings, including teleconferences, where and when necessary
10. Report on the progress of CU5 to the DPACE
11. Maintain the risk register

Input: Progress reports from DU managers, report on resources from national coordinators
Output: Documents and reports
Deliverables: Documents
Dependencies:
Interfaces: National funding negotiations, other CUs through the DPACE
Remarks:

Gaia DPAC WP:		GWP-T-502-00000
Title: Architecture and technical coordination of CU5		
Provider(s): IoA		
Manager(s): P. S. Bunclark		
Start: Phase B	**End**: Phase D2	**Total Effort**: 24 MM
Objective: See Sect. 7.6.3.2.		
Tasks: See Sect. 7.6.3.2.		
Input:		
Output:		
Deliverables:		
Dependencies:		
Interfaces:		
Remarks:		

Gaia DPAC WP:		GWP-T-503-00000
Title: Quality assurance and configuration management for CU5		
Provider(s): RAL, IoA		
Manager(s): P. J. Richards		
Start: Phase B	**End**: Phase D2	**Total Effort**: 41 MM
Objective: See Sect. 7.6.3.3.		
Tasks: See Sect. 7.6.3.3.		
Input:		
Output:		
Deliverables:		
Dependencies:		
Interfaces:		
Remarks:		

Gaia DPAC WP:		GWP-T-504-00000
Title: Integration, validation and operation of CU5		
Provider(s): IoA, RAL		
Manager(s): F. De Angeli		
Start: Phase B	**End**: Phase D2	**Total Effort**: 117 MM
Objective: See Sect. 7.6.3.4.		
Tasks: See Sect. 7.6.3.4.		
Input:		
Output:		
Deliverables:		
Dependencies:		
Interfaces:		
Remarks:		

Gaia DPAC WP: GWP-T-505-00000

Title: Technical support

Provider(s): IOA, RAL

Manager(s): F. De Angeli

Start: Phase B	**End**: Phase D2	**Total Effort**: 46 MM

Objective:
Development of support tools, for both general use and CU5-specific use, support for software development activities.

Tasks:

1. Support the CU5 team in all software development activities
2. Organise development tools workshops for CU5 members
3. Provide and maintain support tools and facilities
4. Provide an interface with CU1

Input:
Requirements from other DUs

Output:

1. Software documentation and software user guide
2. Programming guidelines

Deliverables:

1. Software modules
2. Programming workshops

Dependencies:

Interfaces: CU1
Remarks:

Gaia DPAC WP: GWP-T-510-00000

Title: PSF and LSF calibration

Provider(s): UOL

Manager(s): M. A. Barstow

Start: Phase B	**End**: Phase D2	**Total Effort**: 190 MM

Objective:
DDT of software for PSF and LSF calibrations.

Tasks:

1. Characterize the instrumental optical profiles (PSF and LSF) for SM, AF, and BP/RP sections across the focal plane for calibration purposes.
2. Determination of the methods to characterize PSF and LSF in SM and AF, these methods may be quite different from the ones needed for BP/RP.
3. Determination of the input data relevant to the task.
4. Interface with laboratory data, particularly dealing with issues like radiation damage, charge injection and introduction of gates during the pre-launch phase.
5. Usage of simulations to fine-tune the software and "volume test" the procedures to produce scalable procedures.
6. Construction of the software and incorporation of it within the First Look in the CU3 pipeline.
7. Design methods which would allow the storage of the calibration results in a "historical context", possibly building/maintaining our own "database".
8. Use the history of the PSF/LSF to monitor the trend of the instrument performance and raise warning flags if needed.

Input:
CCD characteristics, SM, AF5 and BP/RP data streams for calibration of PSF, AF1-AF9 and BP/RP data streams for LSF calibration. Use of full windows extracted for monitoring stars across AF1-9. For each object, information about nearest detected objects (physical proximity or FOV1/FOV2 overlap), colour indices, geometrical details of the transit, etc.

Output:
PSF and LSF curves per CCD, resolved on stellar magnitude bins (TBD), colour index, and transit AC coordinate, application modules for these curves in IDT and IDU

Deliverables:
Software (classes) to handle the calibration and application of the PSF and LSF, documentation of those classes and the methods used, assessment of expected reconstruction accuracy and the relation between this accuracy and the parameters derived from the images.
Methods to notify other components (CU3, IDT?) if the data analysis suggests some unexpected behaviour of the telescope/detectors.

Dependencies:
CU1 for the input data stream. Knowledge/access to the stored parameters (main database?). Simulations generated by CU2.

Interfaces:
CU3 and CU2

Remarks:

Gaia DPAC WP: GWP-M-511-00000

Title: BP/RP flux extraction and initial data treatment

Provider(s): LEI, IOA, INAF-OARm

Manager(s): A. Brown

Start: Phase B	**End**: Phase D2	**Total Effort**: 232 MM

Objective:
DDT of the BP/RP flux extraction software package and BP/RP initial data treatment.

Tasks:

1. Perform spectral extraction under the assumption of a known position and for complex or multiple sources.
2. Explore, define, implement and test methods for background determination for dispersed images over the full range of brightness and for multiple sources and crowded regions.
3. Define broad-band flux and colour parameters from single extracted BP/RP spectra.
4. Explore, define, implement and test the algorithms for the determination of the effective wavelength from BP/RP spectra.
5. Explore the possible use of different parameter sets for saturated images.
6. Prepare software modules for the processing of BP/RP spectra in the initial data treatment (determining colours and RVS fluxes).

Input:
Raw BP/RP data (2D or 1D dispersed images), tools and data necessary to make transit predictions for BP and RP for each source in the two fields of view, G-band fluxes, calibrated LSFs/PSFs

Output:
Extracted BP/RP spectra, sky background measurements, broad-band flux estimates (including RVS band), broad-band colour estimates, effective wavelength estimates

Deliverables:
BP/RP spectral extraction, effective wavelength, and broad-band flux- and colour-parameter estimation packages. A simplified version of these for implementation in the IDT/FL pipeline. Test packages for all software. Documentation.

Dependencies:
CU3 for providing the tools and data to make transit predictions for BP/RP

Interfaces:
CU1, requirements for input data and CU3 for defining the colour information and IDT/FL BP/RP processing requirements and for providing the tools and data to make transit predictions for BP/RP.

Remarks:

Gaia DPAC WP: GWP-T-512-00000

Title: Photometric calibration models for G and BP/RP		
Provider(s): UB, IoA, LEI		
Manager(s): C. Jordi		
Start: Phase B	**End**: Phase D2	**Total Effort**: 560 MM

Objective:
Providing adequate and effective models for the internal calibrations of G and BP/RP photometry.

Tasks:
This DU is in charge of the research into the internal photometric calibration models. It includes exploring provisions for the wide range of large- and small-scale influences on the observed fluxes and wavelength scale, defining methods for comparing and combining different dispersion spectra, and accommodating in the calibration models the effects of CTI and aging of the instrument.

Input:
Descriptions of the instrument and its expected evolution with time. Description of available data for every transit. Description of house-keeping data with impact on photometry calibration.

Output:
Description of the calibration models for SM, AF and BP/RP relating aspects of the instrument and of the observations with a set of parameters for the internal photometric calibration.

Deliverables:
One or more photometric-calibration models research papers.

Dependencies:
DU11, DU15, DU16

Interfaces:
CU1, CU3

Remarks:
The integration of the models in the CU5 pipeline is the responsibility of DU15.

Gaia DPAC WP:		GWP-T-513-00000
Title: Instrument absolute response characterisation: ground-based preparation		
Provider(s): INAF-OABo, INAF-OAPd, UB, RUG		
Manager(s): E. Pancino		
Start: Phase B	**End**: Phase D2	**Total Effort**: 410 MM

Objective:
Obtain the ground-based observations required for the absolute calibration of Gaia photometric data.

Tasks:

1. Assessment of needed ground-based observations
2. Acquisition and reduction of those observations
3. Preparation of the data for the application in the CU5 pipeline

Input:
Requirements from GWP-M-514-00000 on the absolute calibration of the Gaia photometric systems

Output:
List of spectrophotometric standard stars (SPSS) and database of the fully reduced ground-based observations of these stars; a full accuracy assessment of these ground-based data

Deliverables:
Database of ground-based observations and documentation describing their characteristics

Dependencies:
None

Interfaces:
CU7 and CU8 for GB observations coordination; GWP-M-519-00000 for database and archive; GWP-M-514-00000

Remarks:

Gaia DPAC WP: GWP-T-514-00000

Title: Instrument absolute response characterisation: definition and application

Provider(s): INAF-OABo

Manager(s): C. Cacciari

Start: Phase B	**End**: Phase D2	**Total Effort**: 235 MM

Objective:
DDT of the methods and software to solve for and apply the absolute calibration model.

Tasks:

1. Explore, define, develop and test methods for calibration of the absolute response of the instrument
2. Define and derive the reconstruction-accuracy expectations and limitations
3. Provide Java class to apply the absolute calibration procedure to the data.

Input:
Ground-based observations of SPSS from GWP-M-513-00000; G mean fluxes, BP/RP mean spectra from GWP-M-515-00000

Output:
Flux calibration parameters and model; algorithm to apply the absolute calibration to the internally calibrated photometry

Deliverables:
The instrument response characterisation model and Java class to convert measured fluxes into absolute fluxes, with documentation

Dependencies:
GWP-M-513-00000 (SPSS fluxes); GWP-M-515-00000 (G mean fluxes, BP/RP mean spectra); CU2 (simulations)

Interfaces:
GWP-M-513-00000, CU2

Remarks:

Gaia DPAC WP:		GWP-T-515-00000
Title: Internal photometric calibration and its application		
Provider(s): IOA		
Manager(s): D. W. Evans		
Start: Phase B	**End**: Phase D2	**Total Effort**: 535 MM

Objective:
DDT of the methods and software required for the internal photometric calibration and its application

Tasks:
This unit is responsible for:

- The development, testing and implementation of the internal calibration procedures following models proposed by DU12;
- The application of the internal and external calibration models to all sources;
- Providing a simplified photometric calibration application for the FL processing to CU3;
- The design, development and testing of the methods and software for the accumulation of the mean flux information;
- The selection of constant stars from the accumulated flux information;
- Trend monitoring of calibration parameters;
- The preparation of the processed data for release to the central data base;

Input:
Interface control documents, providing data descriptions for:

- photometric parameters for AF and SM;
- dispersion spectra in BP and RP (from CU3 and DU11);
- the internal calibration standards (from DU16);

the external calibration model application functions (from DU14); the internal calibration model (from CU12); input requirements for the central data base (from CU1).

Output:
Internal photometric calibration model parameters; calibrated (internally and externally) photometric observations; not-variable flags.

Deliverables:
The software and documentation for

- the internal calibration procedures;
- the application of the internal calibrations;
- the accumulation of calibrated data;
- the selection of non-variables;
- trend monitoring of the calibration parameters;
- the preparation of the data for release to ESAC.

Dependencies:
DU11 (BP/RP data), DU12 (model), DU14 (external calibration model), DU16 (standards), DU19 (internal data base), CU1 (data base), CU3 (photometric parameters AF, SM, photometry for IDU, application of simplified photometric calibration model), CU7 (Requirements on epoch photometry)

Interfaces:
External: CU1 (central data base); Internal: DU11, DU12, DU14, DU16, DU17.

Remarks:

Gaia DPAC WP:		GWP-M-516-00000
Title: Selection of internal reference sources		
Provider(s): UB, IoA		
Manager(s): C. Jordi		
Start: Phase B	**End**: Phase D2	**Total Effort**: 135 MM

Objective:
Selection of reference sources for internal calibration of fluxes (G and BP/RP), wavelength scale and absolute zero wavelength (BP/RP).

Tasks:
Define, implement and test the software for the selection of reference sources for internal calibration of fluxes (G and BP/RP), wavelength scale and absolute zero wavelength (BP/RP).

1. Define suitable criteria for selecting the reference sources
2. Investigate the relation between density and quality of reference sources and the accuracy of the photometric calibration
3. Define, implement and test the software for the selection of reference sources

Input:
Mean *G* fluxes, mean BP and RP spectra, colour information, statistics of fluxes (from DU15), information about the kind of source and its parameters (from CU8), astrophysical information from CU6, results of variability analysis (from CU7), contamination (from DU18, CU4, ...), flags from other processing chains (CU1, CU3).

Output:
Suitable sources for flux calibration; suitable sources for wavelength scale calibration; suitable sources for absolute zero wavelength calibration

Deliverables:
The package to be applied a few times during the mission for selecting internal reference sources

Dependencies:
DU12, DU15, DU19

Interfaces: CU1, CU2, CU3, CU4, CU6, CU7, CU8
Remarks:

Gaia DPAC WP:		GWP-T-517-00000
Title: Flux and classification-based science alerts		
Provider(s): IOA		
Manager(s): S. Hodgkin		
Start: Phase B	**End**: Phase D2	**Total Effort**: 145 MM

Objective:
DDT for flux and classification-based anomaly detections.

Tasks:

1. Define, develop, test and implement methods to provide rapid detection of large flux anomalies in the first look
2. Define, develop, test and implement methods for classification-based anomaly detections to be integrated in the CU5 pipeline
3. Design methods to filter the anomaly candidates appropriately for suitable science alerts
4. Tests on the statistical success rate of the methods, and thus the reliability of the “alerts”

Input:
First look astrometry and photometry, and calibrated photometry from the CU5 pipeline

Output:
Science alerts together with significance

Deliverables:
Flux and classification-based anomaly detection classes, software packages for the first-look flux anomaly detection

Dependencies:
First-look in CU3

Interfaces:
CU6, CU7 and CU8

Remarks:

Gaia DPAC WP:		GWP-M-518-00000
Title: 2D image restoration		
Provider(s): LEI, Bordeaux		
Manager(s): A. Brown		
Start: Phase B	**End**: Phase D2	**Total Effort**: 280 MM

Objective:
DDT of the software required for reconstructing images from SM1,2 and AF2,5,8 transits.

Tasks:

1. Explore, define, develop and test algorithms to produce 2D maps from stacked SM1,2 and AF2,5,8 data
2. Develop methods and software to automatically analyse the maps and derive image parameters for use in the photometric processing pipeline and by other CUs

The output of this image processing will be a characterisation of faint disturbing sources around the target source and this information will be incorporated in a TBD way in the astrometric and photometric data processing. Also images of extended sources will be produced.

Input:
SM1,2 and AF2,5,8 raw data, local plane coordinates for each of the input windows

Output:
2D images, and parameters/flags describing these images

Deliverables:
Image reconstruction and analysis classes

Dependencies:
CU3 for the transformation to local plane coordinates

Interfaces:
CU3 for the delivery of the local plane coordinate data; CU3, CU4, CU6, CU7, CU8 for interface requirements (which image parameters are needed by these CUs)

Remarks:

Gaia DPAC WP:		GWP-T-519-00000
Title: Photometric data base and archive		
Provider(s): IFA		
Manager(s): N. Hambly		
Start: Phase B	**End:** Phase D2	**Total Effort:** 210 MM
Objective: DDT of methods and software for the photometric database.		
Tasks: 1. Create a data base and archive that can be used efficiently and effectively for holding and distributing the data used and produced by CU5 2. Explore and implement methods for fast extraction of the data from the data base and archive		
Input: Requirements from other DUs		
Output: Data base and archive structure and documentation		
Deliverables: The data base and archive interface package		
Dependencies:		
Interfaces:		
Remarks:		

C.7 Top-level Work Packages of CU6

Gaia DPAC WP:		GWP-M-601-00000
Title: Management and Scientific Coordination of CU6		
Provider(s): Obs. Paris		
Manager(s): D. Katz		
Start: Phase B	**End**: Phase F	**Total Effort**: 140 MM
Objective: Manage the coordination unit CU6 spectroscopic processing.		
Tasks: See section 7.6.3.1		
Input: See section 7.6.3.1		
Output: See section 7.6.3.1		
Deliverables: See section 7.6.3.1		
Dependencies: —		
Interfaces: —		
Remarks: —		

Gaia DPAC WP:		GWP-M-602-00000
Title: Architecture and Technical Coordination of CU6		
Provider(s): CNES		
Manager(s): A. JeanAntoine		
Start: Phase B	**End**: Phase E	**Total Effort**: 54 MM
Objective: Define the architecture of the overall structure of the spectroscopic processing system.		
Tasks: See section 7.6.3.2		
Input: See section 7.6.3.2		
Output: See section 7.6.3.2		
Deliverables: See section 7.6.3.2		
Dependencies: —		
Interfaces: —		
Remarks: —		

Gaia DPAC WP:		GWP-M-603-00000
Title: Quality Assurance and Configuration Management for CU6		
Provider(s): CNES		
Manager(s): T. Levoir		
Start: Phase B	**End**: Phase E	**Total Effort**: 36 MM
Objective: Provide Software Quality Assurance Support and Configuration Control for CU6		
Tasks: See Section7.6.3.3		
Input: See Section7.6.3.3		
Output: See Section7.6.3.3		
Deliverables: See Section7.6.3.3		
Dependencies: —		
Interfaces: —		
Remarks: —		

Gaia DPAC WP:		GWP-M-604-00000
Title: Integration, Validation and Operation of CU6 System		
Provider(s): CNES		
Manager(s): A. JeanAntoine		
Start: Phase B	**End**: Phase F	**Total Effort**: 166 MM
Objective: Integrate and validate the software products delivered by the other CU6 development units and operate the spectroscopic processing system.		
Tasks: See section 7.6.3.4		
Input: See section 7.6.3.4		
Output: See section 7.6.3.4		
Deliverables: See section 7.6.3.4		
Dependencies: —		
Interfaces: —		
Remarks: —		

Gaia DPAC WP: GWP-T-610-00000

Title: CU6 Host Software Framework development		
Provider(s): CNES		
Manager(s): A. JeanAntoine		
Start: Phase B	**End**: Phase F	**Total Effort**: 153 MM

Objective:
Develop and validate the Host Software Framework

Tasks:
The host software framework is required :
- to monitor activations of each scientific software (provided by the DUs), managing their dependencies into workflows ;
- to manage their input/output, including interfaces with the GAIA Main Data Base (at ESAC) - for that, a data access layer will be implemented ;
- to manage resource allocation (CPU, network,..).
The main tasks of this WP are:

1. Gather detailed software requirements (functions, performance, quality, tests,..).
2. Follow up sub contracted development (including detailed design, component development and integration phases) and software reviews.
3. Integrate and validate the framework on a reference platform at CNES (Toulouse).

Input:
Task 1 : GWP-T-x02-10000 output.
Task 2 : ITT document, software requirement specifications for the host software framework.
Task 3 : Framework software product, purchased reference platform, scientific software units in test, simulated data.

Output:
Task 1 : Software Requirement Specifications for the host software framework.
Task 2 : Product documentation (Software design document, Integration plan,..), Host software framework product.
Task 3 : Tests results, non conformance reports, change requests, and so on.

Deliverables: Progress report on the framework development
Dependencies: With GWP-T-x02-10000
Interfaces: TBD
Remarks: - Framework development process could be incremental to define as soon as possible the framework design for its main functions. - Host software framework product should be common for all the sub systems hosted by CNES DPC (spectroscopic processing, object processing, astrophysical parameters).

Gaia DPAC WP:		GWP-S-620-00000
Title: Spectra extraction		
Provider(s): MSSL		
Manager(s): S. Rosen		
Start: Phase B	**End**: Phase F	**Total Effort**: 185 MM

Objective:
To extract (and if necessary deblend) spectra, to subtract the backgrounds (internal and external) and to apply the calibration to convert data to standard units. Also to produce a normalised spectrum for later radial velocity determination.

Tasks:

1. 01000 Provide management, configuration management & interfaces with CU6
2. 02000 Carry out detailed functional analysis of the spectra extraction
3. 03000 Carry out extraction of spectra from raw images
4. 04000 Apply calibration
5. 05000 Model background generated by extended sources
6. 06000 Model background generated by point like sources
7. 07000 Clean spectra
8. 08000 Normalise spectra to the continuum
9. 09000 Validate spectral extraction

Input:
Data from IDT or simulations from GIBIS, GASS, GOG. Auxiliary data and data from photometry/astrometry. Calibration files from GWP-S-630

Output:
Cleaned calibrated spectra, normalised spectra.

Deliverables:

1. requirements and design specifications
2. verification specifications
3. working code to applicable standards
4. test harnesses and files
5. code description/manual

Dependencies:
GWP-S-6x-x

Interfaces:
TBI

Remarks:

Gaia DPAC WP:		GWP-S-630-00000
Title: Calibration of the spectroscopic instrument		
Provider(s): MSSL		
Manager(s): S. Rosen		
Start: Phase B	**End**: Phase F	**Total Effort**: 250 MM

Objective:
To determine the calibration of the RVS (throughput, wavelength scale, CCD characteristics *etc.*.

Tasks:

1. 01000 Provide management, configuration management & interfaces with CU6
2. 02000 Carry out detailed functional analysis of the calibrations
3. 03000 Interface with Quick Look group
4. 05000 Detailed First Look & validation: raw data
5. 06000 Implement SGIS
6. 07000 Calibrate CCD bias, CCD readout and dark noises, CCD blemishes
7. 08000 Calibrate photometric throughput, CCD flat field, linearity, saturation level
8. 09000 Determine calibration of AL & AC LSF
9. 10000 Determine wavelength scale, Distortion map
10. 11000 Derive calibration for scattered light & ghosts
11. 12000 Implement detailed first look & validation: calibration
12. 13000 Implement detailed first look: faint stars

Input:
Calibrated data from GWP-S-620. Auxiliary data and data from photometry/astrometry.

Output:
Updated calibrations.

Deliverables:

1. requirements and design specifications
2. verification specifications
3. working code to applicable standards
4. test harnesses and files
5. code description/manual

Dependencies:
GWP-S-6x-x

Interfaces:
TBI

Remarks:

Gaia DPAC WP: GWP-S-640-00000

Title: Radial Velocity zero point

Provider(s): GRAAL

Manager(s): G. Jasniewicz

Start: Phase B	**End**: Phase F	**Total Effort**: 139 MM

Objective:
To ensure the best accuracy for RV measurements obtained with the RVS, in guaranteeing the best RV zero point.

Tasks:

1. Management, configuration management & interfaces of the TWP-S-640.
2. To define simulation requests for potential reference sources.
3. To observe stars and asteroids, and to update consequently lists of stars and asteroids to be used as reference RV sources.
4. To assess asteroids/stars RV zero point consistency.
5. To estimate all effects altering the RV measurements, and to determine the astrophysical zero point

Input:
Catalogues of stars and asteroids extracted from literature, archives and databases

Output:
Release of lists of stars and asteroids with RVs known with an accuracy better than 0.3km.s^{-1} for the RVS zero point.

Deliverables:
Release of reliable lists of RV reference sources (with RV values & errors).

Dependencies:
This TWP depends on the availability of ground-based spectroscopic observations during the next 5 years in the northern and southern hemispheres.

Interfaces: Interface with WP-S-630 (calibration of the RVS), WP-S-811-20000 (Provide calibration of training data), CU2 (general simulation), GWP-M-604-10000.
Remarks: RV = ”Radial Velocity”.

Gaia DPAC WP:		GWP-S-650-00000
Title: Single transit analysis		
Provider(s): Obs. Paris		
Manager(s): Y. Viala		
Start: Phase B	**End**: Phase F	**Total Effort**: 250 MM
Objective: Analysis of the spectra obtained during a single transit in order to determine, using various algorithms, the radial and rotational velocities of the single or multiple observed object(s). For the faintest objects, for which the radial and rotational velocity cannot be derived from a single transit, provide the cross correlation function to GWP-S-660 in charge of performing a multi-transit analysis. Another objective is to perform a detailed first look so as to validate the analysis and to possibly provide science alerts.		

Tasks:
Listing of all tasks this WP consists of

1. Task 1 : Management, configuration managements and interfaces.
2. Task 2 : Definition of test campaigns and comparison of algorithms performances : Define the set of object spectra to be tested by all algorithms ; test algorithms performances (vrad and vrot determinations) on this set of spectra ; refine the set of object spectra and perform algorithms tests every 6 months.
3. Task 3 : Perform a detailed functional analysis and provide a functional diagram of single transit analysis.
4. Task 4 : Make a detailed critical review of existing techniques for spectra analysis and for radial and rotational velocities determination.
5. Task 5 : Manufacture spectroscopic masks to be used in cross-correlation techniques (data space, Fourier space, TODCOR method, ...).
6. Task 6 : Coarse characterization of sources. The main aim is to provide informations about the sources : nature, multiplicity, variability, magnitude, identification of main lines (absorption/emission).
7. Task 7 : Radial and rotational velocity determination by cross-correlation with a template and/or a mask in data space.
8. Task 8 : Radial velocity determination by cross-correlation with a template and/or a mask in Fourier space.
9. Task 9 : Rotational velocity determination by Fourier transform.
10. Task 10 : Radial and rotational velocity determination by minimum distance method.
11. Task 11 : Rotational velocities determination by Neural network.
12. Task 12 : Radial and rotational velocity determination for multi-line spectra by TODCOR-like method.
13. Task 13 : Radial and rotational velocity determination for multi-line spectra by spectrum subtraction method.
14. Task 14 : Single transit detailed first look (DFL) and validation. DFL monitors the output of a single transit in order to perform data and instrument sanity check.
15. Task 15 : Science alerts : provide a quick scientific diagnosis for specific objects for which it would be scientifically interesting to have a quick follow-up with other astronomical facilities.

Input:
Output:
Deliverables: Reports. Functional diagrams. Algorithms. Single transit correlation function. Radial and rotational velocities. Science alerts.
Dependencies: CU2 provides simulated object spectra ; CU 8 provides synthetic spectra (masks, templates for cross-correlation techniques)
Interfaces: CU2 provides simulated object spectra ; CU 8 provides synthetic spectra (masks, templates for cross-correlation techniques)
Remarks:

Gaia DPAC WP:		GWP-S-660-00000
Title: Multiple transits analysis		
Provider(s): MSSL		
Manager(s): S. Rosen		
Start: Phase B	**End**: Phase F	**Total Effort**: 140 MM

Objective:
To produce the final RVS radial velocities from mission-aggregated data.

Tasks:

1. 01000 Provide management, configuration management & interfaces with CU6
2. 02000 Carry out detailed functional analysis of multiple transits data
3. 03000 Make an overview of existing techniques
4. 04000 Derive radial velocities from multi transit data (skew analysis)
5. 05000 Assess sources spectroscopic stability/variability
6. 06000 Combine spectra optimally
7. 07000 Determine mean radial and rotational velocities
8. 08000 Validate multi-transit data

Input:
Calibrated data from mission up to time of run. Auxiliary data.

Output:
Radial and rotational velocities derived from data available at time of run. Assessment of variability and binarity. Aggregated optimally-combined spectra.

Deliverables: 1. requirements and design specifications 2. verification specifications 3. working code to applicable standards 4. test harnesses and files 5. code description/manual
Dependencies: GWP-S-6x-x
Interfaces: TBI
Remarks:

C.8 Top-level Work Packages of CU7

Gaia DPAC WP:		GWP-M-701-00000
Title: Management and Scientific Coordination of CU7		
Provider(s): ObsGE, IoA		
Manager(s): L. Eyer, D.W. Evans, P. Dubath		
Start: Phase B	**End**: Phase F	**Total Effort**: 203 MM
Objective: 7.6.3.1		
Tasks: See section 7.6.3.1		
Input: See section 7.6.3.1		
Output: See section 7.6.3.1		
Deliverables: See section 7.6.3.1		
Dependencies: —		
Interfaces: —		
Remarks: —		

Gaia DPAC WP:		GWP-M-702-00000
Title: Architecture and Technical Coordination of CU7		
Provider(s): ObsGE		
Manager(s): M. Beck		
Start: Phase B	**End**: Phase F	**Total Effort**: 7 MM
Objective: See section 7.6.3.2		
Tasks: See section 7.6.3.2		
Input: See section 7.6.3.2		
Output: See section 7.6.3.2		
Deliverables: See section 7.6.3.2		
Dependencies: —		
Interfaces: —		
Remarks: —		

Gaia DPAC WP:		GWP-M-703-00000
Title: Quality Assurance and Configuration Control for CU7		
Provider(s): ObsGE		
Manager(s): I. Lecoeur		
Start: Phase B	**End**: Phase F	**Total Effort**: 69 MM
Objective: Provide Software Quality Assurance Support and Configuration Control for CU7		
Tasks: See Section7.6.3.3		
Input: See Section7.6.3.3		
Output: See Section7.6.3.3		
Deliverables: See Section7.6.3.3		
Dependencies: —		
Interfaces: —		
Remarks: —		

Gaia DPAC WP:		GWP-M-704-00000
Title: Integration, Validation and Operation of the CU7 System		
Provider(s): ObsGE		
Manager(s): M. Beck		
Start: Phase B	**End**: Phase F	**Total Effort**: 80 MM
Objective: See section 7.6.3.4		
Tasks: See section 7.6.3.4		
Input: See section 7.6.3.4		
Output: See section 7.6.3.4		
Deliverables: See section 7.6.3.4		
Dependencies: —		
Interfaces: —		
Remarks: —		

Gaia DPAC WP: GWP-M-705-00000

Title: Host Framework for CU7

Provider(s): ObsGE

Manager(s): M. Beck

Start: Phase B	**End**: Phase F	**Total Effort**: 75 MM

Objective:
Define and implement the framework to run the variability processing

Tasks:

- gather requirements for the variability framework
- implement functionality to import/export real and simulated data to/from the CU7 system
- implement functionality for data I/O
- implement common functionality used by CU7 software
- implement control and monitoring functionality for variability processing
- test provided functionality

Input:

- Software Development Plan
- Software Requirements Specification
- Software Verification Plan
- Coding Guidelines
- MDB Interface Control Document
- CU7 Interface Control Document

Output:
Deliverables: • Software Requirements Specification • Software Design Document • Software User Manual • tested software modules
Dependencies: This work package depends on all non-common CU7 workpackes
Interfaces: This work has an interface with the Gaia MDB
Remarks: —

Gaia DPAC WP:		GWP-M-710-00000
Title: Special Variability Detection & Analysis		
Provider(s): INAF-OACt, ObsGE, ULB, IAP, LAL, ARI		
Manager(s): A. Lanzafame		
Start: Phase B	**End**: Phase F	**Total Effort**: 211 MM

Objective:
Study, develop and test algorithms that use a priory information to detect small signature of variability that may not be efficiently detected by the usual, standard variability tests.

Tasks:
Specific variability detection methods will be developed for detecting the following types of objects, we provide an estimation of the percentage of work-package effort per subject:

1. planetary transits (e.g. BLS method), 15% of work-package effort
2. extremely short periods (photometry per CCD, SdB stars, ZZ Ceti stars, binary WDs), 30% of work-package effort
3. small amplitude periodic variables, 5% of work-package effort
4. stars scintillations, 30% of work package effort
5. solar-like variability detection 20% of work-package effort

Input:
Photometric and Spectroscopic time-series. Astrophysical Parameters from CU8.

Output:
List of stars candidates for specific variability class.

Deliverables:

- Before the operations:
  - Feasibility studies reports
  - Software documentation and software packages
    - Requirement and Design documents
    - Software deliveries in compliance with the cycle development approach
    - Test Reports
- During the operations:
  - Software and documentation updates

Dependencies:
Variability Characterisation

Interfaces:
CU8, GWP-M-701, GWP-T-702, GWP-T-703, GWP-T-704, GWP-T-705

Remarks:
WP description to be finalised after first feasibility studies.

Gaia DPAC WP:		GWP-M-711-00000
Title: Variability Characterisation		
Provider(s): ROB, KUL, EPFL, ObsGE, INAF-OACt		
Manager(s): J. Cuypers		
Start: Phase B	**End**: Phase F	**Total Effort**: 191 MM

Objective:
Characterise with the simplest possible description the observed variations

Tasks:
The variability will be characterised in different ways, through:

1. Statistical Parameter Determination (GWP-S-711-01000)
2. Period Search (GWP-S-711-02000)
3. Model Fitting (GWP-S-711-03000)

In each case, the results are for each (variable) object a number of parameters providing a description of the variability characteristics.

Input:

Photometric and Spectroscopic time-series,
variability flag(s) from CU5,
list of variable star candidates from Special Variability Detection & Analysis

Output:
Variability parameters.

Deliverables:

- Before and during the operations:
  - Requirement and Design documents
  - Software deliveries in compliance with the cycle development approach
  - Test reports
- During the operations:
  - Follow-up of the delivery of parameters that characterize the variability for the objects with already a sufficient number of observations
  - Validation and evaluation of analysis results
- At the end and after the mission:
  - assistance in the delivery of the final product (variability catalogue, final catalogue), validation and evaluation.

Dependencies:

- Special Variability Detection & Analysis
- CU5
- Variability Classification; the results will go to the variability catalogue and the MDB.

Interfaces:
CU5

Remarks:

Gaia DPAC WP: GWP-M-712-00000

Title: Classification		
Provider(s): KUL, SVO, EPFL, ROB, OAL, ObsGE, INAF-OACt		
Manager(s): C. Aerts		
Start: Phase B	**End**: Phase F	**Total Effort**: 421 MM

Objective:
Perform a systematic classification of all variable sources.

Tasks:
The classification tasks can be divided into three method categories, which will form three work-packages of equal effort. They will be developed and run in parallel:

1. a set of extractors,
2. supervised global classification,
3. un-supervised global classification.

These methods are further described/covered in sub-workpackages.
This work-package contains the following steps:

- Exploration of the different classification methods and their efficiency
- Development of new methods adapted to Gaia
- Design and coding of the methods
- Optimization of the methods
- Merging of different classifications obtained from the three general subwork-packages
- Validation of the classification

Input:
The necessary information from the MDB to classify the variable objets.

Output:
Each object with n associated class of variability (known, unknown, unsolved). This should preferably be with a probability of membership. The objects can be known previously or not.

Deliverables:

- Before the operations:
    - Study report on classification methods
    - Software documentation and software packages
        - Requirement and Design documents
        - Software deliveries in compliance with the cycle development approach
        - Test Reports
- During the operations:
    - Software and documentation updates

Dependencies:

- Variability Characterisation
- Supplementary observations
- Specific Object Studies

Interfaces:
—

Remarks:
—

Gaia DPAC WP: GWP-M-720-00000

Title: Specific Object Studies		
Provider(s): ObsGE, KUL, ARI, INAF-OABo, OCA, UNIVIE, INAF-OACt, INAF-OACn, ONDREJ, KONK, UIUC, AU, INAF-OARm		
Manager(s): N. Mowlavi		
Start: Phase B	**End**: Phase F	**Total Effort**: 295 MM

Objective:
Perform specific analysis for given classes of objects knowing their variability properties, to

- validate the classification
- provide parameters specific to the given variability types for the Gaia catalogue

Tasks:
Coordinate the sub-workpackages that deal with the extraction of variability parameters specific to known classes of variable objects, and monitor the activities related to those sub-workpackages.
For each sub-workpackage determine if the analysis of the various object groups belongs to the Data Processing tasks or if the analysis should be left to the community for the scientific exploitation of the Gaia catalogue. A selection of the topics will therefore be made in agreement with the GST. Thus, the estimation of the work effort is very uncertain.
The classes of objects that are identified to date are (the list remains modifiable as necessary):
1) Opacity driven oscillators in main sequence stars (delta Scuti, gamma Dor, SPB, beta Cep stars, ...). 2) Rapidly Oscillating Ap stars (roAp). 3) RR Lyrae stars and Cepheids. 4) Long Period Variables (LPVs). 5) Solar-like oscillators. 6) Compact oscillators. 7) Pre-main sequence oscillators. 8) Solar-like (magnetic-related) and rotation-induced variable stars. 9) Flare stars. 10)Eruptive stars. 11) Cataclysmic variables. 12) Eclipsing binaries. 13) Rapid phases of stellar evolution. 14) Optical counterparts of high energy sources. 15) Active Galactic Nuclei. 16) Microlensing events. 17) Solar System Objects 18) Other types.
The purpose of the sub-workpackages analysis is to compute a number of class-specific parameters that characterize the variability of these specific objects and that will populate the Gaia catalogue. The definition of those parameters, their astrophysical relevance, and the algorithm to compute them from the Gaia lightcurves will be described in dedicated sub-workpackages, one per object class.
The following phases are foreseen for the sub-workpackages:

- Before the operations:
  - *Evaluation phase*: Evaluate for each class of specific objects the scientific goals to be addressed with the analysis of the Gaia data.
  - *Algorithm definition phase*: Define the variability parameters relevant for each class of specific objects and the numerical methods to compute them from the Gaia database.
  - *S/W development phase*: Develop the software to be integrated in the CU7 data processing pipeline. This phase is divided into more sub-phases with Requirements and Design documents, the deliveries and tests in compliance with the cycle development approach.
- During the operations: Document and software updates. Run the analysis on the Gaia data and populate the Gaia database with the specific object variability parameters.

Input:
Photometric and Spectroscopic time-series from the Database, as well as the results of the variability characterisation and classification performed earlier in the automatic CU7 data processing pipeline.

Output:
Variability parameters for the Gaia catalogue and summary reports for each class of specific objects.

Deliverables:
The following sub-workpackage deliveries are foreseen for each class of specific objects:

- Evaluation phase document: A document describing the goals aimed for each specific object.
- Requirements and Design documents
- Full software package.
- Test reports.
- *Operations phase*: Software updates and statistics reports on the results for each class of specific objects.

Dependencies:
—

Interfaces:
CU4, CU8

Remarks:
The specific object sub-workpackages that are monitored in this workpackage include contributions from many institutes throughout the CU7 consortium, each institute having the responsibility of one or more sub-workpackage.

Gaia DPAC WP:		GWP-M-721-00000
Title: Global Variability Studies		
Provider(s): SVO, KUL, ROB, ObsGE, INAF-OABo, OAL		
Manager(s): L. Sarro		
Start: Phase B	**End**: Phase F	**Total Effort**: 158 MM

Objective:

The statistical analysis and characterization of the Gaia variability catalogue. This includes the forecast of population prevalences based on past and ongoing surveys, the evaluation of mission biases and their impact on the resulting database and the definition of useful statistics for the description and comparison with other variability catalogues, along with the detection of non gaussianity in the space of parameters used to describe variability. The characterization of the Gaia variability survey in terms of bias, completeness, etc, is an essential service which will greatly enhance the post-mission science exploitation.

Tasks:

Global Variability Studies consists of three main sub-tasks:

1. Variability catalogue visualisation
2. Variability catalogue quality Assessment
3. Survey comparison

Two main stages can be defined: the early- and late- mission studies. In the early mission phase, different surveys will be analysed and their statistical properties and quality measures compared. Based on these, extrapolations to Gaia's characteristics will be put forward. During the mission, intermediate studies of the variability catalogue will be carried out and corresponding reports produced, measuring the discrepancies between the projections and the real statistical properties of the intermediate databases. Assessments of the identified biases and quality checks will be performed. This will allow an independent verification of the calibrations.

Input:

Past and ongoing surveys databases. Gaia variability database (intermediate and end of mission).

Output:
Reports on statistical properties and quality assessments.

Deliverables:

1. Definitions of sufficient statistics for the description of the variability catalogue.
2. Requirement and Design documents and their updates throughout the mission in agreement with the cycle development approach.
3. Software for the automated statistical analysis of both Gaia and similar surveys and for quality assessment.
4. Reports on the three aspects that define the WP: the statistical properties of the database in itself and as compared to other databases and on its quality.

Dependencies:
The WP depends mainly on the database definition WP and, through it, on the Special Variability Detection & Analysis WP (GWP-M-710-00000) and the Variability Characterisation WP (GWP-M-711-00000). These will define the parameters used to describe the various kinds of variable objects in the database. Furthermore, the Classification WP both in its supervised and unsupervised subWPs will provide the class definitions to be considered in the population studies and initial priors for the parameters in each class. The dependence is in both directions and close interaction will be needed in order to iterate the solutions. There will possibly be another dependence with the CU2 Data Simulation.

Interfaces:

Remarks:
—

Gaia DPAC WP:		GWP-M-731-00000
Title: Analysis of Impact on Astrometry		
Provider(s): ULB		
Manager(s): A. Jorissen		
Start: Phase C	**End**: Phase F	**Total Effort**: 48 MM
Objective: Assess the astrometric problems linked to variability		

Tasks:

1. Evaluate to which extent Hipparcos data and parallaxes of long-period variable stars are affected by the pulsations or surface brightness asymmetries (spots)
2. Parametrize these effects in a simple way (in terms of intrinsic photocenter motion, characterized by frequency and amplitude) and use the Gaia simulator to evaluate the possible impact on the accuracy of Gaia parallaxes and proper motions of pulsating/spotty stars.
3. If possible, implement algorithms correcting the parallaxes or, at least, raise a flag indicating that the parallax may be affected by variability effects.

Input:
Astrometric and photometric data

Output:
List of possible variability effects affecting astrometry, list of objects where the astrometry can be affected, new astrometry, new parallaxes. More precisely:

- Before the launch: Identification of the regions of the Gaia colour-colour and colour-magnitude diagrams where stars with intrinsic photocentre motion reside: active M dwarfs, RS CVn binaries, Mira and semi-regular variables, red supergiants with irregular photometric variations, …
- During data processing:
  - Identify such stars, and raise a flag
  - Correct the astrometry (if at all possible).

Deliverables:

- Requirements and Design documents in compliance with the cycle development approach.
- Software deliveries
- Report, identifying the threats of intrinsic photocentric motion, and the regions of the colour-colour diagrams which are likely to be mostly affected.

Dependencies:
Test data for this activity are astrometric abscissae (as in the Hipparcos Intermediate Astrometric Data), photometric colour indices and light curves

Interfaces:
CU3, CU4

Remarks:
—

Gaia DPAC WP: GWP-M-732-00000

Title: Supplementary Observations

Provider(s): INAF-OABo, KUL, INAF-OACt, ObsGE, OHP, ONDREJ

Manager(s): G. Clementini

Start: Phase B	**End**: Phase F	**Total Effort**: 65 MM

Objective:
For certain groups of variable objects, observations either from ground based observatories, or from other satellites are needed in order to prepare or verify the variability results in the Gaia catalogue. Observations may also be needed for quality control to check the studies developed in the CU7.

Tasks:

- To gather data from literature and complete them with observations if necessary, to help the classification of objects.
- To perform observations for quality control on specific variability classes.

Input:
Astrometric and photometric data

Output:
List of sources to be observed. Observing Proposals. Meetings to organise/coordinate multi-site observational campains.

Deliverables:

- Requirement/specification document
- First round of observations
- Second round of observations
- Third round of observations

Dependencies:
All other data processing CU7 tasks may put requirements on supplementary observations

Interfaces: As CU7 is not the only Coordination Unit to need other observations, there are connections with the other CUs to organise global observations campaigns. However as CU7 needs time series, the requirements are different from other CUs since we need a specific object to be observed repeatedly and/or with appropriate scheduling in time.
Remarks: —

C.9 Top-level Work Packages of CU8

Gaia DPAC WP: GWP-M-801-00000

Title: Management and scientific coordination

Provider(s): MPIA Heidelberg, CNES

Manager(s): C.A.L. Bailer-Jones

Start: phase B	**End**: phase F	**Total Effort**: 45 MM

Objective:
Manage the Coordination Unit, define work packages, set schedules, follow up on tasks, interface to other coordinating units and to the the DPACE.

Tasks:

1. Define the requirements of the classification aspect of Gaia and set priorities
2. Define the work breakdown structure, i.e. the work packages
3. Allocate groups from the community to the work packages
4. Set milestones and schedules
5. Monitor progress
6. Oversee test runs and define analysis/interpretation protocols
7. Interface to other CUs
8. Report back to the DPACE

Input:
Reports from CU8 DUs and DPACE

Output:
see tasks

Deliverables:
progress reports on CU8

Dependencies:

Interfaces: to all other CUs
Remarks:

Gaia DPAC WP: GWP-T-802-00000

Title: Architecture and technical coordination CU8

Provider(s): CNES

Manager(s): A.M. Janotto

Start: phase B	**End**: phase D2	**Total Effort**: 43 MM

Objective:
This WP has 2 main objectives:

- Define System architecture :
  The Define System architecture activities includes a definition phase and a design phase. It is dedicated both to the architecture of the software to be developed to fulfill all the requirements and to the data model to be managed at DPC level. The CU8 system software is structured in components, with identification of the dependencies between them, the common software,...
  This design phase also identifies the components developed by laboratories to be delivered to (and operated by) the DPC and the components to be developed by the DPC (components of the host framework).
- System administration : Install and maintain hardware (for development, integration, and operation phases) and development tools.

Tasks:
The main tasks of this WP are:

1. system definition.
2. System design.
3. System administration.

Input:
CU1 activities results (Overall system architecture, operation schedule, assurance quality, system ICD,...).
Estimations of computer and data storage resources (provided by each development unit in charge of a component delivered to DPC).

Output: - Functional analysis document - Software System Specifications - Interface Control Documents (or inputs to CU1 ICD describing interfaces between MDB and CU8 system) - Software Design Document - Dedicated technical notes (studies results, choices justifications,. . .).
Deliverables: See above - output
Dependencies: Overall system definition
Interfaces: With CU1 and CU3 activities and choices.
Remarks:

Gaia DPAC WP:		GWP-T-803-00000
Title: Quality assurance and config management for CU8		
Provider(s): CNES		
Manager(s): T. Levoir		
Start: phase B	**End**: phase D2	**Total Effort**: 35 MM

Objective:
- Define and organize the software quality assurance
- Define and organize the configuration management

Tasks:
The main tasks of this WP are:

1. Derive a product assurance plan for the CU8 from the general product assurance plan provided by CU1
2. Organize and manage reviews during the development cycle (phases B, C,D) on documentation and software
3. Support labs to apply quality assurance rules and recommendations during scientific software development and control their application.
4. Define the configuration management plan in conformance with the rules defined in the product assurance plan
5. Manage configuration and releases in the framework (to be defined with CU1)
6. Manage issue, anomaly, change with the support of a tool (to be defined with CU1)
7. Manage and organize Configuration Control Board activities and contribute for technical sides.

Input:
product assurance plan from CU1

Output:
CU8 product assurance plan, review reports

Deliverables:
Task output - see above

Dependencies: With all DUs in CU8
Interfaces: With CU1
Remarks:

Gaia DPAC WP: GWP-T-804-00000

Title: Integration, Validation and Operation of CU8 system

Provider(s): CNES

Manager(s): A.M. Janotto

Start: phase B	**End:** phase F	**Total Effort:** 163 MM

Objective:
- Simulation and test data management
- CU8 System Integration and Validation at DPC
- Optimize CU8 system
- CU8 system operations and monitoring
- Maintain the CU8 system

Tasks:
The main tasks of this WP are:

1. Define simulation data requirements (in conjunction with CU2) and test, verify received data
2. Define test plans and carry out subsystem integration and validation at the DPC according to test plans.
3. Optimize the sub system during the software development, integration and validation phases.
4. Operate and monitor the subsystem at CNES DPC.
5. Maintain all software components.

Input:

Output:

Deliverables:
Task output - see above

Dependencies:

Interfaces:
With DU managers and developers for software corrections /changes.
With CU1 for end-to-end test at the GAIA system level.

Remarks:
Lab efforts in test definition and testing will be identified in dedicated DUs WPs. Note that the *scientific* validation of results during the operations phase is carried out by the individual algorithmic top-level work packages, specifically GWP-S-821 to GWP-S-836 inclusive.

Gaia DPAC WP: GWP-T-805-00000

Title: CU8 Host Software Framework development		
Provider(s): CNES		
Manager(s): A.M. Janotto		
Start: phase B	**End**: phase F	**Total Effort**: 194 MM

Objective:
Develop and validate the Host Software Framework

Tasks:
The host software framework is required :
- to monitor activations of each scientific software (provided by the DUs), managing their dependencies into workflows ;
- to manage their input/output, including interfaces with the Gaia Main Database (at ESAC). . . for that, a data access layer will be implemented ;
- to manage resource allocation (CPU, network,. . .).
The main tasks of this WP are:

1. Gather detailed software requirements (functions, performance, quality, tests,. . .).
2. Follow up sub contracted development (including detailed design, component development and integration phases) and software reviews.
3. Integrate and validate the framework on a reference platform at CNES (Toulouse).

Input:
for Task 1 above: GWP-T-802-10000 output.
for Task 2 above: ITT document, software requirement specifications for the host software framework.
for Task 3 above: Framework software product, purchased reference platform, scientific software units in test, simulated data.

Output: from Task 1 above: Software Requirement Specifications for the host software framework. from Task 2 above: Product documentation (Software design document, Integration plan,. . .), Host software framework product. from Task 3 above: Tests results, non conformance reports, change requests,. . .
Deliverables: Progress reports on the framework development
Dependencies: With GWP-T-802-10000
Interfaces: With other CUs supported by CNES
Remarks: - Framework development process could be incremental to define as soon as possible the framework design for its main functions. - Host software framework product should be common for all the systems hosted by CNES DPC (spectroscopic processing, object processing, astrophysical parameters).

Gaia DPAC WP: GWP-T-806-00000

Title: Data model and utility library

Provider(s): MPIA Heidelberg

Manager(s): C. Tiede

Start: phase B	**End**: phase F	**Total Effort**: 19 MM

Objective:
Produce and maintain the software data model and utility library for the CU. The data model is the set of common Java classes which store data and provide basic manipulations specific to CU8 yet common to many CU8 algorithms. The utility library is a set of classes for common data manipulation and pre- and post-processing.

Tasks:

1. define the data model
2. implement and maintain the data model
3. identify useful software utilities
4. implement and maintain utilities
5. document software

Input:
requirements from the CU, coordinated by the CU8 Configuration Control Board

Output:
data model and utility library Java classes

Deliverables:
documentation on the software

Dependencies:
none

Interfaces:
to the rest of the CU (for requirements) and to the Gaia Toolbox committee to identify what classes can be implemented in the Toolbox instead.

Remarks:

Gaia DPAC WP: GWP-S-811-00000

Title: Provide training and testing data

Provider(s): Nice, INAF-OAPd, Bourdeaux, Uppsala, Montpellier, others

Manager(s): F. Thévenin

Start: phase B	**End**: phase F	**Total Effort**: 349 MM

Objective:
This is a relatively large WP and comprises three well-defined subpackages:

- GWP-S-811-10000: Provide synthetic stellar spectra. Provide synthetic spectra for single stars across the full AP space expected. These should be at sufficient wavelength resolution and coverage to ensure accurate simulations in RP/BP and RVS (although separate simulations could be provided for each of RP/BP and RVS). The synthetic spectra must show variance in (at least) $T_{\rm eff}$, $\log g$, [Fe/H], [α/Fe].

- GWP-S-811-20000: Provide calibrations of training data. Provide the data and parameters required to calibrate the classification and AP estimation methods. The primary method will be to identify a grid of stars on the sky with a wide range of APs which can be used to calibrate the AP estimation models. Determine accurate APs for these using existing data and/or new data. Take overall responsibility for determining what auxiliary data (i.e. real data) are required for all classification algorithms. This may be existing data from catalogues or libraries or may involve new observations, either spectroscopy or photometry.

- GWP-S-811-30000: Assemble training data sets. Coordinate the data requirements of the various classification algorithms within this CU and construct the required training data sets. This refers to both stellar and non-stellar objects.

The relationship between these three subpackages is described in Sect. 8.9.1.3, in particular Fig. 61. The three subpackage above correspond to the three yellow ellipses on the right of the diagram. The “source specific WPs” on the left of that figure and the simulated data inputs from GWP-S-831, -832, -833, -834 and -835.

Tasks:

1. Use stellar atmospheric models to synthesize the required spectra.
2. Maintain and improve atomic and molecular line lists required by the models.
3. Make improvements to stellar structural, evolutionary and atmospheric models to more realistically include relevant phenomena (e.g. 3D, diffusion, NLTE).
4. Use models and spectra to improve the AP determinations of real (auxiliary) data used in the training data or for calibration
5. Establish what real data are required to correct the synthetic data to improve performance of the AP estimation algorithms. Obtain such data and perform the corrections.
6. Establish a grid of "Gaia calibrator stars" and estimate the APs of these using existing or new data. These are relatively bright stars which Gaia will observe and which have (or will have) well-determined APs covering a range of APs as wide as possible.
7. Obtain any catalogue/library auxiliary data as required.
8. Coordinate, plan, execute and reduce any new observations as required for calibration purposes.
9. Work with the rest of the CU to determine what training data (synthetic or real) are required for which classification tasks.
10. perform synthetic interstellar reddening of all source spectra.
11. Simulate spectra of binary systems using single star spectra.
12. Act as an interface between the CU and CU2 Simulations, specifically for specifying and soliciting the necessary simulations, passing the necessary data to CU2 and receiving simulated data from CU2.
13. Apply any necessary data transformations of the classification model inputs and outputs.
14. Construct and maintain libraries of training data as required by the CU.

Input:
Gaia simulations from CU2. Ground-based data.

Output:
See deliverables.

Deliverables:

1. A complete grid of synthetic stellar spectra showing variance in $T_{\rm eff}$, $\log g$, [Fe/H], $[\alpha/{\rm Fe}]$ at sufficient wavelength resolution and coverage to ensure accurate simulations in RP/BP and RVS
2. Complete calibration grid (i.e. all stars identified along with their approximate APs) by launch.
3. Accurate APs for half of the stars in the calibration grid by launch and for the rest by end-of-mission.
4. Software for converting simulated spectra provided by the simulations CU2 into training files for the various classification algorithms, including any necessary input/output transformations.
5. Software for artificially reddening an arbitrary spectrum.
6. Software for “correcting” synthetic spectra using real spectra.
7. Libraries of training data on various types of objects as required by the CU.
8. Regular progress reports as specified by the CU manager.

Dependencies:

1. Telescope time for new observations as appropriate
2. Specifications and requirements from the rest of CU8
3. Synthetic spectra of other sources from the SSUs in CU8
4. Reddening specifications in coordination with GWP-S-812-00000
5. Simulated data received from CU2 and then passed to rest of CU8

Interfaces:

1. Provide CU2 with source spectra
2. Receive from CU2 Gaia simulated spectra and astrometry
3. Interface with all CUs for the co-ordination of ground-based observations
4. Interface with all algorithmic WPs in CU8

Remarks:

1. This WP provides all data used by the CU8 algorithms for development (training and testing) purposes. Source simulations for non-stellar objects and source simulations of various complex phenomena required by GWP-S-835-00000 (Extended Stellar Parametrizer) are the responsibility of the top-level WP which requires them. But in all cases these source spectra will be passed via this WP for performing instrument simulations (by CU2) and for constructing data sets.

2. The performance of the various classification algorithms depends heavily on the quality of the training data. In general this will be a combination of synthetic and real data, with real data may use to make continuum corrections to synthetic spectra (for example). Auxiliary data are used to determine the APs of the Gaia calibrator stars and thus calibrate the Gaia classification algorithms. (The Gaia data form the training input data, the accurate APs from the auxiliary data the training output – or target – data.) Alternatively (or in addition) auxiliary data are used to calibrate or adjust directly the synthetic data prior to training. This procedure is described in more detail in ICAP-CBJ-005. See ICAP-AB-001, ICAP-AB-002, GAIA-CBJ-001 and ICAP-CS-001 for early plans concerning obtaining new data. More FTE will be required for this task is new observations have to be obtained and reduced.

3. This WP operates during both the pre-launch and mission phases. The latter is necessary so that (a) the classifications and parametrizations are be based on the latest source models and (b) the training data for the classifiers reflect any changes in the instrumentation which differ from the nominal simulations. The latter are not trivial due to radiation damage plus the evolution of the instruments during operations.

Gaia DPAC WP: GWP-S-812-00000

Title: Interstellar extinction

Provider(s): V. Straizys, K. Zdanavicius, A. Kazlauskas, R. Lazauskaite, J. Knude

Manager(s): R. Drimmel

Start: phase B	**End**: phase F	**Total Effort**: 72 MM

Objective:
Establish an optimal method for determining the line-of-sight interstellar extinction to individual stars, based on RP/BP data (in GSP-phot) and RVS data (in GSP-spec). Advise the algorithm writers on the best way to implement this and oversee the data simulations to ensure that extinction is handled properly.

Tasks:

1. Advise on how best to parameterize extinction (e.g, with A_V or G-band extinction; with more than one parameter (e.g. including R_V)?)

2. In cases where extinction parameters are estimated, provide algorithms for calculating the G-band extinction for each source.

3. Take responsibility for ensuring that extinction simulations are done as required, e.g. according to an appropriate extinction law or laws, at a suitable grid spacing of A_V (or equivalent) values, etc.

4. Advise each set of classification algorithm writers on how to deal with extinction.

5. Assess the global degeneracies between extinction and the other APs, especially effective temperature.

6. Monitor the completeness of the extinction estimates. Is external data necessary to aid the extinction estimation? (This would be in addition to, and not instead of, Gaia-only estimates.) If necessary, assess how to implement this.

7. Analyze and compare the different extinction estimates during the mission.

8. Provide regular progress reports as specified by the CU manager.

Input:

1. Stellar input and simulated RP/BP spectra from GWP-S-811-00000
2. Source APs from GSP-phot (GWP-S-822-00000)

Output:
G-band extinction for each source in the training data

Deliverables:

1. Document with recommendations on how to parameterize extinction.
2. Document describing simulation requirements for extinction, including its frequency dependent effects.
3. Algorithm to calculate the G-band extinction from the estimated source APs
4. Feasibility study of the possibility of determining the extinction law (i.e. R_V) from the Gaia spectro-photometry. Assessment of uncertainties introduced in the case that a nominal extinction law is assumed.
5. Report assessing the global degeneracies between extinction and other APs.
6. Study of various extinction estimates made in CU8 processing, including recommendations regarding how to use or combine these estimates.
7. Conversion tables for deriving extinctions in standard photometric bands from the adopted Gaia extinction measures.
8. A requirement study regarding the need for external data and/or models to aid extinction determination.
9. Depending on the outcome of the above study, a software module which will estimate the extinction (semi)independently of the other algorithms.

Dependencies:
None.

Interfaces:

1. Advise GWP-S-821 (DSC), GWP-S-822 (GSP-phot) and GWP-S-823 (GSP-spec) on determination of extinction)
2. For input source APs: GWP-S-822
3. For simulation recommendations and simulated spectra: GWP-S-811
4. For G-band output: GWP-S-825

Remarks:
Interstellar extinction is considered as an intrinsic AP from the point of view of AP estimation, and is estimated star-by-star along with the other APs. However, as it is not simulated by model stellar atmospheres, it needs special attention.

Gaia DPAC WP:		GWP-S-821-00000
Title: Discrete Source Classifier		
Provider(s): MPIA Heidelberg		
Manager(s): C.A.L. Bailer-Jones		
Start: phase B	**End**: phase F	**Total Effort**: 170 MM
Objective: Study, develop and test the algorithms for the Discrete Source Classifier (both the "early mission", e-DSC, and "late mission", l-DSC, versions; see section 5.5.3 and Fig. 39). DSC takes photometry on unresolved "slow" sources (see "Remarks") and determines the broad class of the object, i.e. whether it is a single star, binary star, QSO etc.		

Tasks:

1. Study the problem of doing discrete source classification according to the requirements and guidelines laid out in the ICAP documentation.
2. Study the statistics literature and assess the suitability and performance of existing models. Report on this to the rest of the CU.
3. Develop, implement and test new classification methods as appropriate.
4. Assess the performance of classification algorithms and report on these.
5. Study how best to introduce parallaxes, proper motions and variability information into this classification work.
6. Study and recommend the appropriate input data and output classes to be used by e-DSC and l-DSC.
7. Optimize the codes (from the point of view of the method).
8. Define, implement and test preliminary and final e-DSC and l-DSC algorithms which will be used during the mission.
9. Verify and validate both test results (development phase) and mission results (operations phase) from the algorithms.
10. Modify and optimize the algorithms during the operations (mission) phase based on what is learned about the real (rather than simulated) Gaia data.
11. Participate in construction of the final Gaia catalogue using the results from these algorithms.

Input:
e-DSC: calibrated RP/BP data (all robust time averages) and their covariances (from photometric noise model) on unresolved, “slow” sources.
l-DSC: as e-DSC plus parallax and proper motions and their covariances (from astrometric noise model) and suitable variability indices.

Output:
e-DSC: probability of class membership (see GAIA-C8-SP-MPIA-CBJ-019).
l-DSC: as e-DSC.

Deliverables:

1. A project plan outlining how the problem will be addressed for the period 2006–2011.
2. Reports on the suitability of various approaches to the problem including performance assessments and recommendations for the classes.
3. Report on the study of how best to introduce parallaxes, proper motions and variability indices into l-DSC.
4. Fully functional version of e-DSC for full-scale pre-launch tests.
5. Fully functional version of l-DSC for full-scale pre-launch tests.
6. Documentation on the algorithms.
7. Performance assessment and predictions for e-DSC and l-DSC by launch.
8. Improved version of e-DSC during the mission as required (i.e. to adapt to lessons learned on real Gaia data).
9. Improved version of l-DSC during the mission as required (i.e. to adapt to lessons learned on real Gaia data).
10. Regular progress reports as specified by the CU manager.

Dependencies:

1. During operations: Calibrated, combined RP/BP spectra from CU5
2. During operations: Calibrated astrometry from CU3

Interfaces:

1. CU3 for identification of QSOs for the astrometric reference frame.
2. GWP-S-812 for parametrization of extinction.
3. GWP-S-824 and GWP-S-825 for identification of outliers and natural classes.

Remarks:
A “slow” source is one which was not detected to have any proper motion by the onboard detection. Non-slow sources are fast moving solar system objects and thus do not need to pass through the DSC. l-DSC may be based on a parallel classifier. In this case, dedicated modules for identifying (i.e. assigning a probability to) a source as being each of a QSO, unresolved galaxy or minor planet, may be provided by the relevant SSU (TBD). For prior work on DSC see ICAP-CH-001, ICAP-AB-003 and ICAP-AB-004.
Method: A supervised method based on a combination of synthetic and real data. This must accommodate missing and truncated data and make optimal use of the covariances (errors) on the inputs. This provides multiple class probabilities so will presumably be based on modelling the probability density function over the classes as a function of the inputs. A simple partitioning of the data space to give “yes/no” classifications is insufficient. It must be possible to specify explicitly the prior classifications for individual sources.

Gaia DPAC WP:		GWP-S-822-00000
Title: Generalized Stellar Parametrizer – Photometry (GSP-phot)		
Provider(s): MPIA Heidelberg		
Manager(s): C.A.L. Bailer-Jones		
Start: phase B	**End**: phase F	**Total Effort**: 176 MM
Objective: Study, develop and test the algorithms for GSP-phot (both the "early mission", e-GSP-phot, and "late mission", l-GSP-phot, versions; see section 5.5.3 and Fig. 39). GSP-phot estimates APs for the full range of type of single stars based on photometry (e-GSP-phot) or photometry plus parallaxes (l-GSP-phot).		

Tasks:

1. Study the problem of doing continuous stellar astrophysical parameter estimation according to the requirements and guidelines laid out in the ICAP documentation.
2. Study the statistics literature and assess the suitability and performance of existing models. Report on this to the rest of the CU.
3. Develop, implement and test new parametrization methods as appropriate.
4. Assess the performance of parametrization algorithms and report on these. In particular address the issue of whether to use a regression approach or an estimation approach.
5. Study how best to introduce parallaxes into stellar parametrization work.
6. Study the issue of degeneracies (a given set of input data corresponding to more than one set of output APs).
7. Study and recommend the appropriate input data and output APs which are achievable for each of e-GSP-phot and l-GSP-phot.
8. Optimize the codes (from the point of view of the method) as necessary.
9. Define, implement and test preliminary and final e-GSP-phot and l-GSP-phot algorithms which will be used during the mission.
10. Verify and validate both test results (development phase) and mission results (operations phase) from the algorithms.
11. Modify and optimize the algorithms during the operations (mission) phase based on what is learned about the real (rather than simulated) Gaia data.
12. Participate in construction of the final Gaia catalogue using the results from these algorithms.

Input:
e-GSP-phot: calibrated RP/BP data (all robust time averages) and their covariances (from photometric noise model) on single stars (as defined by the DSC).
l-GSP-phot: as e-GSP-phot plus parallax and its (asymmetric) variance or PDF (from astrometric noise model).

Output:
e-GSP phot: $\mathrm{T}_{\mathrm{eff}}$, $\log g$, [Fe/H], $[\alpha/\mathrm{Fe}]$, A_{G} plus their uncertainties. More than one set of APs must be reported in the case of degeneracies, along with a probability of each.
l-GSP-phot: as e-GSP-phot

Deliverables:

1. A project plan outlining how the problem will be addressed for the period 2006–2011.
2. Reports on the suitability of various approaches to the problem including performance assessments.
3. Report on the study of how best to introduce parallaxes in l-GSP-phot.
4. Report on the study of degeneracies and how best to handle these.
5. Fully functional version of e-GSP-phot for full-scale pre-launch tests.
6. Fully functional version of l-GSP-phot for full-scale pre-launch tests.
7. Documentation on the algorithms.
8. Performance assessment and predictions for e-GSP-phot and l-GSP-phot by launch.
9. Improved version of e-GSP-phot during the mission as required (i.e. to adapt to lessons learned on real Gaia data) within 1 year from start of observations.
10. Improved version of l-GSP-phot during the mission as required (i.e. to adapt to lessons learned on real Gaia data) by end-of-mission.
11. Regular progress reports as specified by the CU manager.

Dependencies:

1. During operations: Classification from GWP-S-821 (DSC)
2. During operations: Calibrated, combined RP/BP spectra from CU5
3. During operations: Calibrated astrometry from CU3

Interfaces:

1. GWP-S-821
2. CU6 (to provide stellar astrophysical parameter estimates required for template selection for the cross correlation in the wavelength calibration)

Remarks:
Various ICAP working group reports have already studied various aspects of the GSP-phot. See ICAP-CBJ-002, ICAP-VM-001, ICAP-VM-002, ICAP-AB-003, ICAP-AB-004, ICAP-PW-001, ICAP-PW-004, ICAP-PW-005, ICAP-PW-006, ICAP-PW-008.
Method: A supervised method based on a combination of synthetic and real data. This must accommodate missing and truncated data and make optimal use of the covariances (errors) on the inputs. It must be possible to specify explicitly the prior classifications for individual sources (TBD). A significant issue to address is whether to take a regression approach (pre-trained mapping of the input space to the output space) or a case-by-case estimation. Another significant issue is how to identify and record degeneracies.

Gaia DPAC WP: GWP-S-823-00000

Title: Generalized Stellar Parametrizer – Spectroscopy (GSP-spec)		
Provider(s): Nice		
Manager(s): A. Recio-Blanco		
Start: phase B	**End**: phase F	**Total Effort**: 172 MM

Objective:
Study, develop and test the algorithm for GSP-spec. Using both the RVS spectra and AP estimates from GSP-phot, this re-estimates APs for the full range of type of single stars. It will only be applied to stars which have an RVS spectrum of sufficient quality (i.e. sufficient SNR and for which background subtraction was successful). It may also estimate A_{V} based on the Diffuse Interstellar Band (DIB).

Tasks:

1. Study the problem of doing continuous stellar astrophysical parameter estimation.
2. Study the statistics literature and assess the suitability and performance of existing models. Report on this to the rest of the CU.
3. Develop, implement and test new parametrization methods as appropriate.
4. Assess the performance of parametrization algorithms and report on these. In particular address the issue of whether to use a regression approach or an estimation approach.
5. Study the issue of degeneracies (a given set of input data corresponding to more than one set of output APs).
6. Study and recommend the appropriate input data and output APs which are achievable.
7. Optimize the codes (from the point of view of the method) as necessary.
8. Define, implement and test preliminary and final algorithm which will be used during the mission.
9. Study methods for determining $A_{\rm V}$ from the DIB, taking into account the APs provided by GSP-phot.
10. Verify and validate both test results (development phase) and mission results (operations phase) from the algorithms.
11. Modify and optimize the algorithms during the operations (mission) phase based on what is learned about the real (rather than simulated) Gaia data.
12. Participate in construction of the final Gaia catalogue using the results from these algorithms.

Input:
Calibrated RVS spectra (robust time average) plus flux covariances (from spectroscopic noise model) on single stars (the “SB0” stars, as determined by CU6).

Output:
T_{eff}, $\log g$, [Fe/H], [α/Fe] and A_{G}, plus their uncertainties. More than one set of APs must be reported in the case of degeneracies, along with a probability of each. Other APs may be determined (e.g. emission line strengths, specific element abundances), although recall that the SSUs are intended to make more detailed analyses on restricted sets of stars.

Deliverables:

1. A project plan outlining how the problem will be addressed for the period 2006–2011.
2. Reports on the suitability of various approaches to the problem including performance assessments.
3. Report on the study of degeneracies and how best to handle these.
4. Fully functional version of GSP-spec for full-scale pre-launch tests.
5. Documentation on the algorithms.
6. Performance assessment and predictions for GSP-spec by launch.
7. Improved version of GSP-spec during the mission as required (i.e. to adapt to lessons learned on real Gaia data) within 1 year from start of observations.
8. Regular progress reports as specified by the CU manager.

Dependencies:

1. During operations: Classification from GWP-S-821 (DSC)
2. During operations: Stellar astrophysical parameters from GWP-S-822 (GSP-phot)
3. During operations: Calibrated, combined RVS spectra from CU6

Interfaces:
CU6 provides additional identification of spectroscopically single stars.

Remarks:
GSP-spec could take AP estimates as inputs either from e-GSP-phot or l-GSP-phot. It could only be applied to fainter stars later in the mission, i.e. for which sufficient spectra have been obtained to achieve sufficient SNR. Note that there are not separate early and late mission versions of GSP-spec (because astrometry and variability are not explicit inputs). For studies of this problem see, for example, ICAP-TK-001 or Bailer-Jones (2003, ASP Conf. Ser. 298, 199–208).
Method: A supervised method based on a combination of synthetic and real data. This must accommodate missing and truncated data and make optimal use of the covariances (errors) on the inputs. It must be possible to specify explicitly the prior APs for individual sources (i.e. those from the photometry). A significant issue to address is whether to take a regression approach (pre-trained mapping of the input space to the output space) or a case-by-case estimation. Another significant issue is how to identify and record degeneracies.

Gaia DPAC WP: GWP-S-824-00000

Title: Object Clustering Analysis

Provider(s): Spanish Virtual Observatory

Manager(s): L.M. Sarro

Start: Phase B	**End**: Phase F	**Total Effort**: 115 MM

Objective:
Develop an algorithm which can identify the 'natural' or 'intrinsic' classes in the Gaia data, and thus lead to the identification of new or unusual types of objects. Produce both early and late mission versions of appropriate algorithms. This WP provides important feedback to GWP-S-821 (DSC) to help define the class distributions and boundaries.

Tasks:

1. Overall coordination of the Work Package
2. Data gathering and preprocessing
3. Software development
4. Algorithm testing
5. Verify and validate both test results (development phase) and mission results (operations phase) from the algorithms.
6. Modify and optimize the algorithms during the operations (mission) phase based on what is learned about the real (rather than simulated) Gaia data.
7. Participate in construction of the final Gaia catalogue using the results from these algorithms.

Input:

1. BP & RP spectrophotometry
2. RVS spectra or processed features such as line fluxes or equivalent widths.
3. Astrometry

Output:

1. Probabilistic estimation of the number of clusters in the data
2. Statistical description of the members of each cluster. Identification when and where possible of the main contributors to each cluster in terms of known astrophysical labels.
3. Clustering quality measures
4. Identification of low (but nonzero) probability regions of input hyperspace
5. Within cluster correlations of any kind.
6. For individual objects: probabilistic cluster assignments.

Deliverables:

1. A project plan outlining how the problem will be addressed.
2. Produce a requirements document for the software delivered by the WP.
3. Report on existing software for clustering and an identification of libraries of reusable software both for the statistical analysis of input data and output clusters , the pre- and post-processing stages and the clustering itself. Find visualization tools for the multidimensional.
4. First version of the document describing the protocol for the algorithm selection, specifications required and a priori criteria for the assessment of the different clustering approaches.
5. Document with assessments of the validity of the different clustering approaches as applied to existing astrophysical databases, in particular, the Sloan Digital Sky Survey. Evaluate hierarchical *versus* partitional, divisive *versus* agglomerative, deterministic *versus* stochastic, monothetic *versus* polythetic, hard *versus* fuzzy and incremental *versus* non-incremental approaches using available software. Check interpretability and analyze statistical properties of resulting clusters. Learn from the results obtained for the SDSS broad band photometry and degraded spectra and elaborate projections for Gaia.
6. Final report on the best paradigm(s) to be used, results of the testing and comparison phase, justification of the choice.
7. Produce a final report on the purpose and specific deliverable of the OCA algorithms, plus a detailed implementation plan.
8. Fully functional version of e-OCA for full-scale pre-launch tests.
9. Fully functional version of l-OCA for full-scale pre-launch tests.
10. Documentation on the algorithms.
11. Improved version of OCA algorithms during the mission.

Dependencies:

1. During operations: Calibrated, combined RP/BP spectra from CU5
2. During operations: Calibrated astrometry from CU3

Interfaces:

1. GWP-S-821 (DSC)

Remarks:
Physical parameter estimation methods necessarily rely on physical models in order to infer astrophysical parameters, because these parameters are not directly observable. Although this is the ultimate goal of the Gaia classification work, these methods can only correctly parametrize known types of objects. They will fail on unknown objects (or rather, any type of object not included in the training template set). OCA is intended to complement the other classification algorithms by making a non-physical analysis of the collective properties of objects to see what are the natural clusterings in the data. By identifying where known objects lie, we can identify previously unknown types of objects and subject them to more detailed examination. OCA therefore forms the basis for an internal classification system of the Gaia data –either through the definition of internal parameters or discrete classes– which may then be calibrated onto physical parameters according to a variety of physical models.

Gaia DPAC WP:		GWP-S-825-00000
Title: Final Luminosity, Age and Mass Estimator (FLAME)		
Provider(s): Paris, Nice		
Manager(s): Y. Lebreton		
Start: phase B	**End**: phase F	**Total Effort**: 128 MM

Objective:
Estimate the age and mass of individual stars based on their APs, parallax and apparent G-band magnitude. A crucial intermediate step is to derive the absolute G-band magnitude and intrinsic luminosity for each star using the provided extinction, parallax and other APs. Thus this WP can be considered as transforming the *atmospheric* stellar parameters to *global* stellar parameters.

Tasks:

1. Develop and test an algorithm which provides an optimal estimate of the intrinsic stellar G-band magnitude and total luminosity from the APs ($T_{\rm eff}$, $\log g$, [Fe/H], $A_{\rm G}$), average apparent G-band magnitude and the parallax, optimally using the uncertainties on (or even probability distribution over) these.
2. Develop and test an algorithm which estimates age and mass for individual stars from the available data (the intrinsic luminosity or G-band absolute magnitude plus $T_{\rm eff}$ and [Fe/H]) and provides confidence intervals on these.
3. Investigate whether/how to extend this to binary systems
4. Verify and validate both test results (development phase) and mission results (operations phase) from the algorithms.
5. Modify and optimize the algorithms during the operations (mission) phase based on what is learned about the real (rather than simulated) Gaia data.
6. Participate in construction of the final Gaia catalogue using the results from these algorithms.

Input:
APs as determined by l-GSP-phot; G-band magnitude; parallax, for single stars.

Output:
Absolute G-band magnitude, intrinsic luminosity, mass and age along with confidence intervals on these (generally asymmetric).

Deliverables:

1. Reports on the suitability of various approaches to the problem including performance assessments.
2. Fully functional version of FLAME for full-scale pre-launch tests.
3. Documentation on the algorithms.
4. Improved version of FLAME during the mission as required (i.e. to adapt to lessons learned on real Gaia data) by end-of-mission.
5. Regular progress reports as specified by the CU manager.

Dependencies:

1. Classification from GWP-S-821 (DSC)
2. Stellar astrophysical parameters from GWP-S-822 (GSP-phot)

Interfaces:
GWP-S-821, GWP-S-822, CU4

Remarks:
FLAME is best done once the parallaxes are available, so there is no early mission version of the algorithm.

Gaia DPAC WP:		GWP-S-831-00000
Title: Quasar classifier		
Provider(s): Liège		
Manager(s): J.-F. Claeskens		
Start: phase B	**End**: phase F	**Total Effort**: 128 MM

Objective:
Study, develop and test algorithms which provide optimal astrophysical parameters estimates for quasars, based on the assumption that the object is restricted to this class. Provide all synthetic data required to identify and classify these objects with Gaia, as required by the DSC algorithm.

Tasks:

1. Develop and provide software to determine QSO parameters.
2. Advise the DSC group on how best to identify QSOs, i.e. separate them from all other objects Gaia will observe, using photometry, proper motions, parallaxes and variability information.
3. Obtain and provide any necessary synthetic data required to operate and test the classification algorithms.
4. Interact with the WP “Provide calibrations and auxiliary data” to obtain any real data required.
5. Verify and validate both test results (development phase) and mission results (operations phase) from the algorithms.
6. Modify and optimize the algorithms during the operations (mission) phase based on what is learned about the real (rather than simulated) Gaia data.
7. Participate in construction of the final Gaia catalogue using the results from these algorithms.

Input:
Calibrated RP/BP spectra. Variability indices (from CU7).

Output:
Prediction of at least three parameters for each QSO (redshift, continuum slope, emission line strength).

Deliverables:

1. Libraries of QSO spectra as required for the DSC development.
2. Reports on the suitability of various approaches to the problem, objectives and performance assessments. Final report required by mid 2009 (i.e. prior to selecting and developing the final algorithm).
3. Regular progress reports as specified by the CU manager.

Dependencies:

1. During operations: Classifications from GWP-S-821 (DSC)
2. During operations: Calibrated, combined RP/BP data from CU5
3. During operations: Variability information from CU7

Interfaces:
GWP-S-821 (DSC) for exchange of information/ideas on optimal methods to identify QSOs.

Remarks:
Prior work is presented in ICAP-JCF-001 and Claeskens et al. 2006, MNRAS

Gaia DPAC WP:		GWP-S-832-00000
Title: Unresolved galaxy classifier		
Provider(s): UOA Athens		
Manager(s): M. Kontizas		
Start: phase B	**End**: phase F	**Total Effort**: 128 MM
Objective: Study, develop and test algorithms which provide optimal astrophysical parameter estimates for unresolved galaxies, based on the assumption that the object is restricted to this class. Provide all synthetic data required to identify and classify these objects with Gaia, as required by the DSC algorithm.		

Tasks:

1. Libraries of galaxy spectra
   (a) Build a library of Synthetic galaxy spectra
      i. Use of the existing codes (PEGASE) in order to prepare a comprehensive and realistic set of typical galaxy spectra.
      ii. Search the key Astrophysical Parameters for the construction of the galaxy spectra
   (b) Build one or two libraries of real galaxy spectra
      i. Selection of a typical set (probably more than one) of real spectra covering the wavelength range of GAIA observations.
      ii. Fitting with the synthetic spectra
2. Parametrization - Classification
   (a) Search optimum GAIA ”Colour Indices” and/or broad band spectral features
   (b) Extended tests of various classification algorithms, statistical tests of performance, estimates of execution speed etc
   (c) Indentify key parameters for galaxies to determine
   (d) Classify (or sub-classify) extended extragalactic objects with respect to the spatial behaviour of their spectral properties (Nice, Strasbourg)
3. Radial Velocities High Resolution Synthetic Spectra
   (a) PSF-like but extragalactic. (Slezak, Bijaoui and Recio-Blanco)
   (b) Set of Simulated Galaxy Spectra with high resolution
4. Verify and validate both test results (development phase) and mission results (operations phase) from the algorithms.
5. Modify and optimize the algorithms during the operations (mission) phase based on what is learned about the real (rather than simulated) Gaia data.
6. Participate in construction of the final Gaia catalogue using the results from these algorithms.

Input: Calibrated RP/BP spectra.
Output: Prediction of up to 5 astrophysical parameters for each target plus discrete (e.g. Hubble type) classes.
Deliverables: 1. Libraries of galaxy spectra as required for the DSC development. 2. Production of Simulated Galaxy Spectra with resolution R=11500 for GAIA spectroscopy at wavelengths from 848-874A (external auxiliary data). 3. Research for important spectral features in this wavelength range taking into account realistic redshifts. 4. Algorithm for doing galaxy classification (discrete types) and parameter estimation.
Dependencies: 1. During operations: Classifications from GWP-S-821 (DSC) 2. During operations: Calibrated, combined RP/BP data from CU5
Interfaces: CU2, CU6, GWP-S-821 (DSC)
Remarks: Initial work reported in Tsalmantza et al. (2007), A&A, submitted

Gaia DPAC WP:		GWP-S-833-00000
Title: Solar system object classifier		
Provider(s): UAO Uppsala		
Manager(s): C.-I. Lagerkvist		
Start: phase B	**End**: phase F	**Total Effort**: 103 MM
Objective: Study, develop and test algorithms which provide taxonomic classes and/or parameter estimates for solar system objects, based on the assumption that the object is restricted to this class. Provide all synthetic data required to identify and classify these objects with Gaia, as required by the DSC algorithm.		

Tasks:

1. GWP-M-833-10000: Develop, implement and test classification algorithms for solar system objects (asteroids, Trojans, Centaurs, transneptunian objects, comets)
    (a) GWP-M-833-11000: Define optimal taxonomic classes for BP/RP system
        i. GWP-M-833-11100 (Lagerkvist & Warell, Uppsala): Analysis of the suitability of the Tholen and Bus and Binzel taxonomic classification systems in the Gaia BP/RP system, and definition of recognizable taxonomic classes. Based on supervised classification. The algorithm (in matlab) is working and will now be tested using simulated asteroid spectra.
        ii. GWP-M-833-11200 (Bendjoya, Nice): Analysis of data from the 52-colour system to see how it fits with the classification of Bus and Binzel. It is important to see how the infrared part is reflected in the different taxonomic types of Bus and Binzel.
        iii. GWP-M-833-11300 (Barrucci, Paris; Dotto, Roma): Analysis of how the classification is affected if we also have albedo information. They use IRAS data and G-mode classification to test the robustness of the Bus and Binzel classification.
    (b) GWP-M-833-12000 (Lagerkvist & Warell, Uppsala): Evaluate anticipated result of algorithms based on simulation of expected full observational sample of SEDs of minor bodies
2. GWP-M-833-20000: Construct reliable taxonomic template spectra for each class. Perform spectroscopic observations of selected representative objects in each taxonomic class in order to provide new reliable template spectra for full BP/RP wavelength range.
    (a) GWP-M-833-21000 (Lagerkvist & Warell, Uppsala): VIS-NIR low-resolution spectroscopy with the Nordic Optical Telescope, La Palma.
    (b) GWP-M-833-22000 (other contributors, e.g. Bendjoya)
3. Verify and validate both test results (development phase) and mission results (operations phase) from the algorithms.
4. Modify and optimize the algorithms during the operations (mission) phase based on what is learned about the real (rather than simulated) Gaia data.
5. Participate in construction of the final Gaia catalogue using the results from these algorithms.

Input:

1. Solar SED (obtained in collaboration with Korn (GWP-S-835))
2. Minor planet taxonomic template spectra (SEDs), final number of which will be determined in simulations. Initial templates are derived from existing ground-based spectroscopic surveys, and final templates will be derived from dedicated ground-based spectroscopic observations carried out by us. Approximately 30 in number, for asteroids, comets, transeptunian objects and other object types.

Output:

1. Probability-based taxonomic classifications of observed solar system objects based on observed SEDs. For the final Gaia catalogue, listing of the three most probable taxonomic classifications for each object, with probability estimate for each taxonomic class.

Deliverables:

1. Libraries of asteroid spectra as required for the DSC development.
2. Algorithm for doing taxonomic classification (discrete types) and parameter estimation.

Dependencies:

1. During operations: Classifications from GWP-S-821 (DSC)
2. During operations: Calibrated, combined RP/BP data from CU5 (including of solar analogue)

Interfaces:
GWP-S-811 for obtaining ground-based calibration data

Remarks:

Gaia DPAC WP: GWP-S-834-00000

Title: Multiple Star Classifier (MSC)

Provider(s): MPIA Heidelberg

Manager(s): C.A.L. Bailer-Jones

Start: phase B	**End**: phase F	**Total Effort**: 110 MM

Objective:
Study, develop and test algorithms which provide optimal AP estimates for both components of an unresolved binary system. This may be either a physical or an optical binary.

Tasks:

1. Develop, implement and test classification algorithms as required.
2. Obtain and provide any necessary synthetic data required to operate and test the algorithms.
3. Interact with the WP "Provide calibrations and auxiliary data" to obtain any real data required.
4. Investigate how to optimally include astrometric and RV information – which indicates the presence of a physical binary – for determining the APs of the individual components.
5. Analyse both test and mission results from the algorithms.
6. Participate in construction of the final Gaia catalogue using the results from these algorithms.
7. Verify and validate both test results (development phase) and mission results (operations phase) from the algorithms.
8. Modify and optimize the algorithms during the operations (mission) phase based on what is learned about the real (rather than simulated) Gaia data.
9. Participate in construction of the final Gaia catalogue using the results from these algorithms.

Input:
RP/BP; RVS (for brighter objects); SB1 and SB2 indications from CU6; astrometric information from Astro on physical binaries (e.g. common proper motion, orbital solutions); AP estimates from GSP-phot (and GSP-spec where available).

Output:
APs: $T_{\rm eff}$, $\log g$, [Fe/H], $[\alpha/{\rm Fe}]$, $A_{\rm G}$, plus their confidence limits for both components. Additional APs or flags. More than one set of APs must be reported in the case of degeneracies, along with a probability of each.

Deliverables:

1. A project plan outlining how the problem will be addressed for the period 2006–2011.
2. Reports on the suitability of various approaches to the problem, objectives and performance assessments.
3. Fully functional version of all classification algorithms for full-scale prelaunch tests.
4. Documentation on the algorithms.
5. Improved version all classification algorithms (by end-of-mission at the latest) as required to adapt to lessons learned on real Gaia data.
6. Regular progress reports as specified by the CU manager.

Dependencies:

1. During operations: Classification from GWP-S-821 (DSC)
2. During operations: Calibrated, combined RP/BP spectra from CU5
3. During operations: Calibrated, combined RVS spectra from CU6
4. During operations: Information on the presence of (astrometric) binaries from CU4

Interfaces:
GWP-S-821 (DSC)

Remarks:
This WP has the potential to become very elaborate, but this should be avoided. A preliminary study is presented in ICAP-PW-002 and ICAP-PW-003.

Gaia DPAC WP:		GWP-S-835-00000
Title: Extended Stellar Parametrizer (ESP)		
Provider(s): various		
Manager(s): Y. Frémat		
Start: phase B	**End**: phase F	**Total Effort**: 304 MM

Objective:
Study, develop and test algorithms which provide parameter estimates for "extreme" types of stars. These are defined as those which require additional parameters to those provided by GSP-phot and GSP-spec and/or which require special treatment.
This is a relatively large WP and currently comprises five well-defined and independent subpackages, each corresponding to a different type of "extreme" object:

- GWP-S-835-10000: Hot stars
- GWP-S-835-20000: Cool stars
- GWP-S-835-30000: Ultra cool dwarfs (brown dwarfs)
- GWP-S-835-40000: Abundance anomalous stars
- GWP-S-835-50000: Emission line stars

Tasks:

1. Investigate and recommend what subpackages are required and what they should do (e.g. which APs it should determine and over what range) to provide additional information to that provided by GSP-phot and GSP-spec.
2. Develop, implement and test classification algorithms as required.
3. Obtain and provide any necessary synthetic data required to operate and test the algorithms.
4. Coordinate with GWP-S-811-00000 to obtain any real data required for calibration purposes.
5. Verify and validate both test results (development phase) and mission results (operations phase) from the algorithms.
6. Modify and optimize the algorithms during the operations (mission) phase based on what is learned about the real (rather than simulated) Gaia data.
7. Participate in construction of the final Gaia catalogue using the results from these algorithms.

Input:

RP/BP; RVS (for brighter objects); parallaxes (for late mission phase, CU3); AP estimates from GSP-phot; AP estimates from GSP-spec (if available); Variability (for late mission phase, CU7); Radial velocity and projected rotation velocity (if available and for late mission phase, CU6); Binarity flag and/or orbital parameters (if available and for late mission phase, CU4). Energy distributions of stars of various peculiarity types, synthetic spectra of model atmospheres, interstellar reddening law and its variations.

Output:

Algorithms for identification and parameterizing of stars of various peculiarity types. This includes additional parameters appropriate for the special type of star and/or improved estimates of $T_{\rm eff}$, $\log g$, [Fe/H], [α/Fe], [O/C], $A_{\rm G}$ (plus their confidence limits) if possible and necessary. More than one set of APs must be reported in the case of degeneracies, along with a probability of each. All of the subpackages also provide source spectra which are processed by GWP-S-811 and delivered to CU2 for simulating sources and thus generating the CU8 training data (used in particular by GWP-S-821, -822 and -823); see Sect. 8.9.1.3 and Fig. 61.

Deliverables:

1. A project plan outlining how the problem will be addressed for the period 2006–2011.
2. Libraries of galaxy spectra as required for the DSC, GSP-phot and GSP-spec development.
3. Reports on the suitability of various approaches to the problem, objectives and performance assessments. Report on the performance relative to GSP-phot and GSP-spec.
4. Fully functional version of all classification algorithms by for full-scale pre-launch tests.
5. Documentation on the algorithms.
6. Improved version all classification algorithms (by end-of-mission at the latest) as required to adapt to lessons learned on real Gaia data.
7. Regular progress reports as specified by the CU manager.

Dependencies:
DSC (GSP-S-821), GSP-phot (GWP-S-822), GSP-spec (GWP-S-823)

Interfaces:
Extinction methods (GWP-S-811), DSC (GSP-S-821), GSP-phot (GWP-S-822), GSP-spec (GWP-S-823)

Remarks:
A note on the relationship of this WP to GSP-phot (GWP-S-822) and GSP-spec (GWP-S-823) is in order. ESP re-examines a restricted subset of “extreme” stars in order to: (1) test alternative assumptions (stellar models); (2) provide additional parameters suitable to the specific type of object (e.g. Carbon abundance); (3) explore using alternative combinations of the Gaia data. For the set of stars in common between ESP and GSP-phot (small in number, but scientifically interesting) the APs may be discrepant. APs from GSP-phot are always recorded in the final catalogue. Additional APs (e.g. from ESP, but also from GSP-spec) may also be recorded too. (Because all APs are model-dependent, it is fine to report multiple APs provided the models/assumptions used are described).

Gaia DPAC WP:		GWP-S-836-00000
Title: Outlier Analysis		
Provider(s): Galician Group for Gaia (GGG)		
Manager(s): M. Manteiga		
Start: phase B	**End**: phase F	**Total Effort**: 126 MM

Objective:
Outliers are objects which cannot be identified by DSC (GWP-S-8310) or OCA (GWP-S-8350) as a known class. The objective of this WP is to analyse the data on outliers, in particular, to see whether they are misclassified "known" objects, and to flag such objects to the DSC group in order to help them to improve the DSC. The remaining outliers ("UFOs") will be studied by matching astrometry with surveys and considering the possibility of merged or non-single objects. Finally, the nature of the remaining real unidentified objects will be studied in more detail, possibly via follow-up observations. This WP is a relevant part of the main Gaia data processing because it provides direct feedback to DSC. (DSC is a supervised algorithm and can only classify what it knows about. The present WP improves the class definitions based on *observed* Gaia data.)

Tasks:

1. Perform outliers analysis simulations using available datasets from sky surveys.
2. On an early phase of the mission, analyse outliers on RP/BP data, considering both outliers detection by DSC and OCA. On DSC outliers, perform supervised classification taking into account OCA natural classes. Classify the remaining outliers and try to identify them in terms of physical parameters.
3. Perform an outliers analysis on RVS data for those objects that GSP-spec (GWP-S-8340) can not identify as normal classes.
4. On a late phase of the mission, outliers on RP/BP data will be analysed, taking into account all available information (spectroscopy for brighter objects, parallaxes, proper motions and variability indices).
5. Schedule and develop ground-based observations as needed.
6. Verify and validate both test results (development phase) and mission results (operations phase) from the algorithms.
7. Modify and optimize the algorithms during the operations (mission) phase based on what is learned about the real (rather than simulated) Gaia data.
8. Participate in construction of the final Gaia catalogue using the results from these algorithms.

Input:
RP/BP and RVS data, DSC classes, OCA classes; parallaxes, proper motions and variability indices (for late mission phase).

Output:
Identification of known objects which have been incorrectly classified as UFOs by DSC. Heuristic classes for the remaining unidentified objects.

Deliverables:

1. A project plan outlining how the problem will be addressed for the period 2006–2011.
2. Reports on the suitability of various approaches to the problem including performance assessments.
3. A final report on the purpose and specific deliverables of the selected supervised and unsupervised algorithms.
4. Fully functional version of algorithms for full-scale pre-launch tests.
5. Fully functional version of late mission algorithms for full-scale prelaunch tests.
6. Documentation on the algorithms.
7. Improved version of early algorithms during the mission as required (i.e. to adapt to lessons learned on real Gaia data) within one year of the start of observations.
8. Improved version of late mission algorithms during the mission as required (i.e. to adapt to lessons learned on real Gaia data) by end-of-mission.
9. Regular progress reports as specified by the CU manager.

Dependencies:

1. DSC (GWP-S-821): probability of memberships and initial outlier identification
2. OCA (GWP-S-824): natural classes and complementary outlier identification
3. GSP-phot (GWP-S-822) and GSP-spec (GWP-S-823): for objects with rare or unusual spectra

Interfaces:

1. GWP-S-821 (DSC)
2. GWP-S-824 (OCA)
3. GWP-S-834 (MSC), because many outliers could be unusual (or poorly modelled) binaries.

Remarks:

C.10 Top-level Work Packages of DPC-B

Gaia DPAC WP:		GWP–M-B01-00000
Title: Management of Gaia at the BPC		
Provider(s): BPC		
Manager(s): S.Girona		
Start: Phase B	**End**: Phase F	**Total Effort**: 65 MM
Objective:		
Tasks: Management of implementation and execution of tasks listed in GWP-O-B01		
Input:		
Output:		
Deliverables: BPC Status Reports		
Dependencies:		
Interfaces:		
Remarks: The BPC dedicates part of his time to Gaia according to the proposals presented		

Gaia DPAC WP:	GWP-O-B10-00000

Title: Software deployement and operation at BPC		
Provider(s): BPC		
Manager(s): S. Girona		
Start: Phase B	**End**: Phase F	**Total Effort**: 370 MM

Objective:
This WP covers work performed by the Operations Team of the BPC DPC

Tasks:
This WP covers operation of:

1. The GASS simulator(see section Sect. 6.4)
2. The GOG simulator (see section Sect. 6.6)
3. The IDT test (see section Sect. 4.1)
4. The IDU, PSF and 2-D imaging processing (see section Sect. 5.1.6)

Input:
SW delivered from GWP-345

Input:
SW delivered from GWP-350

Input:
raw data from local DB, source data, attitude, calibration and auxiliary data from MDB

Output:
GASS Telemetry stream

Output:
GOG simulated data

Output:
Intermediate data updated after AGIS run

Deliverables:
Simulated telemetry data, GOG simulated data, Intermediate data updated

Deliverables:
DPACC standard documentation

Dependencies: CU2, CU3 software
Interfaces: Mainly with CU2 and CU3. Interface with CU1 for transfer of MDB data and architectural aspects
Remarks:

C.11 Top-level Work Packages of DPC-C

Gaia DPAC WP: GWP–M-C01-00000

Title: Management of the CNES DPC		
Provider(s): CNES		
Manager(s): X.Passot		
Start: Phase B	**End**: Phase F	**Total Effort**: 89 MM
Objective: To organize the development, acceptance and operations of the CNES DPC		
Tasks: Tasks definitions, subcontracting host framework development, operations organization.		
Input:		
Output:		
Deliverables: Management reports, development plan, management plan		
Dependencies:		
Interfaces:		
Remarks: the CNES DPC is entirely funded by CNES.		

Gaia DPAC WP:		GWP-O-C10-00000
Title: Operation of DPC systems		
Provider(s): CNES		
Manager(s): X. Passot		
Start: Phase B	**End**: Phase F	**Total Effort**: 309 MM

Objective:
This WP covers work performed by the Operations Teams of the Gaia DPC at CNES

Tasks:
The tasks of this work package are:
This WP covers operation of:

1. CU4 objects processing
2. CU6 Spectroscopic processing
3. CU8 Astrophysical parameters processing
4. Gaia Transfer System

Input:

Output:

Deliverables:
Operations reports.

Dependencies:

Interfaces:
Mainly with the MDB from ESAC (as DPC for IDT/FL/AGIS).

Remarks:
CNES DPC operations are entirely funded by CNES.

Gaia DPAC WP:		GWP-T-C60-00000
Title: CNES DPC Host Framework development		
Provider(s): CNES		
Manager(s): X. Passot		
Start: Phase B	**End**: Phase F	**Total Effort**: 725 MM
Objective: The CNES DPC will runs CU4, 6, 8 processing. See Sec. 9.3 for details.		
Tasks: Define Overall System hardware and software Architecture for Processing. Specify and manage the design and the development of the computer system until acceptance.		
Input:		
Output:		
Deliverables: Development documents as required by the product assurance plan		
Dependencies:		
Interfaces: CNES DPC interfaces mainly with the ESAC DPC .		
Remarks: the CNES DPC develoment is entirely funded by CNES.		

C.12 Top-level Work Packages of DPC-E

Gaia DPAC WP:		GWP–M-E01-00000
Title: Management of the ESAC DPC		
Provider(s): W. O’Mullane, U. Lammers		
Manager(s): W. O’Mullane		
Start: Phase B	**End**: Phase F	**Total Effort**: 56 MM
Objective: Sect. 7.6.3.1		
Tasks: Tasks and sub tasks of GWP-M-E01-00000 As in Sect. 7.6.3.1.		
Input:		
Output:		
Deliverables:		
Dependencies:		
Interfaces:		
Remarks: ESAC is entirely funded by ESA and the DPC is an ESA service to DPAC.		

Gaia DPAC WP: GWP-O-E10-00000

Title: Operation of ESAC (DPC-E) systems		
Provider(s): J. Hoar, A. Volpicelli, BPC		
Manager(s): W. O'Mullane		
Start: Phase B	**End:** Phase F	**Total Effort:** 126 MM

Objective:
This WP covers work performed by the Operations Teams of the Gaia SOC at ESAC

Tasks:
The tasks of this work package are:
This WP covers operation of:

1. Initial Data treatment (IDT)
2. First Look (FL)
3. Astrometric Global Iterative Solution (AGIS).
4. Gaia Main Database
5. Gaia Transfer System

Input:
Operations manual, operations plans

Output:

Deliverables:
Operations reports.

Dependencies:

Interfaces:
Mainly between CU3 (ARI, Lund, Barcelona) and ESAC as DPC for IDT/FL/AGIS. Interface to all DPCs for transfer of MDB data.

Remarks:
ESAC is entirely funded by ESA and the DPC is an ESA service to DPAC.

Gaia DPAC WP:		GWP-T-E02-00000
Title: Architecture and technical coordination DPC-E		
Provider(s): W. O'Mullane, J. Hoar		
Manager(s): W. O'Mullane		
Start: Phase B	**End**: Phase F	**Total Effort**: 221 MM
Objective: Sect. 7.6.3.2 ESAC is the hub of the DPAC system all data flows through the DPC and must be integrated in the DPC.		
Tasks: Tasks and sub tasks of GWP-T-102-00000 Sect. 7.6.3.2 Define Overall System Architecture for Processing (GSDD).		
Input:		
Output:		
Deliverables:		
Dependencies:		
Interfaces: All other DPCs for data transfer.		
Remarks: ESAC is entirely funded by ESA and the DPC is an ESA service to DPAC.		

C.13 Top-level Work Packages of DPC-G

Gaia DPAC WP:		GWP-M-G01-00000
Title: Management of DPG		
Provider(s): ObsGE		
Manager(s): L. Eyer, M. Beck		
Start: Phase B	**End:** Phase F	**Total Effort:** 40 MM
Objective: 7.6.3.1		
Tasks: See section 7.6.3.1		
Input: See section 7.6.3.1		
Output: See section 7.6.3.1		
Deliverables: See section 7.6.3.1		
Dependencies: —		
Interfaces: —		
Remarks: —		

Gaia DPAC WP: GWP-O-G10-00000

Title: Operation of the DPC-G systems

Provider(s): ObsGE

Manager(s): M. Beck

Start: Phase D2	**End:** Phase F	**Total Effort:** 81 MM

Objective:
This WP covers work performed by the Operations Team of the DPG

Tasks:
This WP covers operation of:

1. CU7 variability processing
2. Gaia file transfer system

Input:

Output:

Deliverables:
Operations reports

Dependencies:
—

Interfaces:
Interface to the DPE for transfer of MDB data and interface to CU7

Remarks:
—

Gaia DPAC WP: GWP-T-G02-00000

Title: Architecture and Technical Coordination of DPG		
Provider(s): ObsGE		
Manager(s): M. Beck		
Start: Phase B	**End**: Phase F	**Total Effort**: 16.5 MM
Objective: See section 7.6.3.2		
Tasks: See section 7.6.3.2		
Input: See section 7.6.3.2		
Output: See section 7.6.3.2		
Deliverables: See section 7.6.3.2		
Dependencies: —		
Interfaces: —		
Remarks: —		

Gaia DPAC WP: GWP-T-G05-00000

Title: Host Framework for CU7

Provider(s): ObsGE

Manager(s): M. Beck

Start: Phase B	**End:** Phase F	**Total Effort:** 75 MM

Objective:
Define and implement the framework to run the variability processing

Tasks:

- gather requirements for the variability framework
- implement functionality to import/export real and simulated data to/from the CU7 system
- implement functionality for data I/O
- implement common functionality used by CU7 software
- implement control and monitoring functionality for variability processing
- test provided functionality

Input:

- Software Development Plan
- Software Requirements Specification
- Software Verification Plan
- Coding Guidelines
- MDB Interface Control Document
- CU7 Interface Control Document

Output:
Deliverables: • Software Requirements Specification • Software Design Document • Software User Manual • tested software modules
Dependencies: This work package depends on all non-common CU7 workpackes
Interfaces: This work has an interface with the Gaia MDB
Remarks: —

C.14 Top-level Work Packages of DPC-I

Gaia DPAC WP:		GWP–M-I01-00000
Title: Management of the IoA DPC		
Provider(s): IoADPC		
Manager(s): F. van Leeuwen		
Start: Phase E	**End**: Phase F	**Total Effort**: 89 MM
Objective: To organize the development, acceptance and operations of the IoADPC		
Tasks: Tasks definitions, subcontracting host framework development, operations organization.		
Input:		
Output:		
Deliverables: Management reports, development plan, management plan		
Dependencies:		
Interfaces:		
Remarks:		

Gaia DPAC WP:		GWP-O-I10-00000
Title: Operation of DPE systems		
Provider(s): F. van Leeuwen, F. De Angeli		
Manager(s): F. van Leeuwen		
Start: Phase E	**End**: Phase F	**Total Effort**: 126 MM
Objective: This WP covers work performed by the Operations Teams of the Gaia SOC at the IoADPC		
Tasks: The tasks of this work package are: This WP covers operation of: 1. Receive and archive on a daily basis photometric parameter data from ESAC 2. Internal photometric calibrations 3. Application of calibrations, accumulation statistsics per object, distribution of data. 4. Selection of suitable internal standards 5. External calibration and application 6. Receive and incorporate catalogue updates from ESAC		
Input:		
Output:		
Deliverables: Operations reports.		
Dependencies:		
Interfaces: Mainly between CU3 (ARI, Lund, Barcelona) and ESAC as DPC for IDT/FL/AGIS. Interface to all DPCs for transfer of MDB data.		

Remarks:
ESAC is entirely funded by ESA and the DPC is an ESA service to DPAC.

C.15 Top-level Work Packages of DPC-T

Gaia DPAC WP:		GWP–M-T01-00000
Title: Management of the INAF-OATo DPC		
Provider(s): INAF		
Manager(s): A. Volpicelli		
Start: Phase B	**End**: Phase F	**Total Effort**: 360 MM
Objective: Sect. 7.6.3.1		
Tasks: Tasks and sub tasks as in Sect. 7.6.3.1.		
Input:		
Output:		
Deliverables: DPC Status Reports		
Dependencies:		
Interfaces:		
Remarks: The INAF-OATo DPC is a dedicated DPC of CU3. See section Sect. 9.6.		

Gaia DPAC WP:		GWP-O-T10-00000
Title: Operation of INAF-OATo systems		
Provider(s): INAF		
Manager(s): R.Morbidelli		
Start: Phase B	**End**: Phase F	**Total Effort**: 360 MM
Objective: This WP covers work performed by the Operations Team of the INAF-OATo DPC		
Tasks: This WP covers operation of: 1. the Astrometric Verification Unit (AVU) processing SW. 2. BAM data processing SW (see section Sect. 5.1.8.6) 3. Astrometric Instrument Model SW (described in section Sect. 5.1.8.5) 4. WFS data processing SW (if operated after commissioning)		
Input: SW delivered from GWP-340.		
Output:		
Deliverables: Operations reports.		
Dependencies:		
Interfaces: Mainly with CU3 as DPC for AVU. Interface with ESAC for transfer of MDB data.		
Remarks: The INAF-OATo DPC is a dedicated DPC of CU3. The estimated Required Effort does not take into account the system management for the DPC (operating the infrastructure).		

Gaia DPAC WP:		GWP-T-T02-00000
Title: Architecture and technical coordination INAF-OATo DPC		
Provider(s): INAF		
Manager(s): A. Volpicelli		
Start: Phase B	**End**: Phase F	**Total Effort**: 240 MM
Objective: Sect. 7.6.3.2		
Tasks: Sect. 7.6.3.2 Define system architecture for AVU and BAM processing.		
Input:		
Output:		
Deliverables:		
Dependencies:		
Interfaces: ESAC DPC for MDB data.		
Remarks: The INAF-OATo DPC is a dedicated DPC of CU3.		

D ESA deliverables

This appendix describes the deliverables that the DPAC needs from ESA to perform the data processing.

Aside from the deliverables specified below it would be proper to mention resources that the Gaia Project Team supporting the PS has already made available to the Gaia community. These include:

- Gaia Parameter Database (also mentioned below)
- Gaia webpage, including My Portal access
- Gaia People Database
- Gaia Wiki
- Livelink
- Subversion source code control system
- Mantis bug/issue tracker

We hope that these resources will continue to be available in the future, with appropriate access permissions to members of the DPAC.

D.1 ESAC

As stated in the Gaia SMP, the Gaia DPAC is, ”a collaboration between the ESA Gaia Science Operations Centre (SOC) [located at ESAC] and a substantial and broad scientific community.” As part of the DPAC, and for the sake of completeness, the ESAC contribution has been included in this proposal, though appropriately highlighted, as it is a service provided to the scientific community at large and is at the disposal of any consortium that may respond to the AO.

ESAC has a major role to play in DPAC, it forms the ’hub’ for data processing and system engineering as outlined in Sect. 8.2 and will perform a major part of the core processing Sect. 9.2 as well as being involved in the development of the core processing software. ESAC team members are well integrated in the different levels of the DPAC managment from DPACE down to DU level.

This section details precisely the resources that make up the ESAC contribution to the DPAC. These resource are already accounted for in the Gaia Cost At Completion.

D.1.1 Human Resources

The table Tab. 175 provides a summary of the effort provided by ESA to DPAC Tab. 176 breaks this down per work package, these are a small subset of the work-packages outlined in Appendix. A.

Table 175: Summary of ESA effort in DPAC (in months)

	ESA Effort	**DPAC Effort**	**Total Effort**
CU1 Summary	1051	138	1189
CU3 Summary	462	4651	5113
DPAC Total Summary	**1513**	**19087**	**20600**

Table 176: Details of ESA effort at WP level (in months)

WP Number	**Name**	**ESA Effort**	**DPAC Effort**	**Total Effort**
`GWP-M-101`	Management and scientific coordination of CU1	151	3	154
`GWP-T-102`	Architecture and technical coordination CU1	99	14	113
`GWP-T-103`	Quality assurance and config management for CU1	156	26	181
`GWP-M-104`	Integration,Validation of CU1 systems	31	0	31
`GWP-T-110`	Coordination common software resources	110	6	116
`GWP-T-140`	Technology Trend Monitoring	21	66	87
`GWP-T-150`	End-to-end system testing	69	11	79
`GWP-T-160`	Host Framework	17	0	17
`GWP-T-170`	Interaction with ESOC	56	0	56
`GWP-T-180`	Main Database Design/ Implement	307	2	309
`GWP-T-190`	Gaia Transfer System Design/Implement	36	10	46
`GWP-M-320`	Manage and Implement AGIS (Core Algorithm Framework)	285	256	541
`GWP-M-350`	Manage and Implement IDT (Initial Data Treatment)	164	962	1125
`GWP-D-360`	Manage and Implement First look	14	729	742
`GWP-M-E01`	Management of ESAC DPC	56	0	56
`GWP-T-E02`	Architecture and technical coordination ESAC DPC	221	0	221
`GWP-O-E10`	Operation ESAC Systems(AGIS,IDT,FL,MDB,GTS)	126	0	126
Totals		**1513**	**2084**	**3597**

D.1.2 Computing resources

As described in Sect. 9.2 ESA will provide substatial hardware for the execution of IDT, FL, AGIS and the MAINDB. The actual hardware numbers are not presented in this proposal but have been included in the Cost at Completion for Gaia.

D.2 Instrument specifications

The instrument specifications are primarily needed for a proper simulation of the Gaia data, which must take into account any relevant instrumental effects needed to support developement of the data processing system, to support the monitoring of the instrument as it evolves during operations, and to support the validation of the resulting science products. Given the aggressive precision and accuracy goals of the Gaia mission, a detailed knowledge of the instrument will be necessary.

Note that the following specifications are not exhaustive nor detailed; for a detailed list see [Bab07]. More specifications may arise when the relevant documentation become available. Note also that the DPAC models that will be based on the different specifications requested below will have to be validated against the various documentation and data that can be made available. These validation data are equivalent to the pre-launch calibration data (Sect. D.5). Ideally competent technical interfaces with ESA will have to be guaranteed during all the phases of mission implementation (development, construction, ground testing and calibration, in-flight commissioning), including support during science operations.

D.2.1 General

- all relevant technical documentation on the spacecraft and instrument design
- maintenance of the content of the Gaia Parameter Database
- all relevant technical documentation on the spacecraft and instrument operations
- all science relevant documentation on the spacecraft and instrument procurement

D.2.2 Optical specifications

A precise description of the optical configuration is needed by the data analysis community in order to model the projection from the FoVRS to the FPRS, and the expected range of variation of image parameters over the focal plane and over time. Given the instrumental precisions to be achieved, this description must include finite element and thermal models of the optics and support structure. A realistic range of configurations is needed to allow the construction and the validation of the data reduction algorithms.

D.2.3 Detector specifications

A precise description of the detectors response with any possible defect or ageing are needed by the DPAC in order to construct and validate the data reduction algorithms. In particular, sufficient information must be made available to construct the detector models, including inputs describing the device and readout chain physical and operating parameters at beginning of life and their expected variation with time.

D.2.4 Orbit and attitude specifications

In order to study the effects of the satellite attitude on the mission and on the data to reduce, detailed models are needed of the satellite dynamics, the on-board attitude control loop, the rigidity of the satellite, its expected reaction to external hits, and torques caused by the satellite itself (cf. [vT06]).

D.2.5 Auxiliary equipment (BAM, WFS) specifications

The detailed description of the auxiliary equipment is required for implementation of a model able to link their measurements with the actual status of the science payload, also taking into account the science data where appropriate. In particular, the WFS appears to provide an assessment of the optical response in two positions wide apart in the along scan direction, and the science data can be used to extrapolate the optical response to the whole science field (across scan direction, SM, AF, BP/RP, RVS). For the BAM, it is crucial that the science community is able to derive a model linking the astrometric measurements from the two arms of the instrument; at first order, the BAM should provide the direct measurement of variations to the basic angle, assumed as common mode over the field and for all sources. A detailed description of the optical model is therefore requested for the auxiliary instruments, as well as information on the operations of the auxiliary equipment.

D.3 On-board processing specifications

The on-board data processing for Gaia consists of the detection and confirmation of sources, the allocation of windows to these sources, their tracking across the focal plane, reading out of the window samples, and then sending the data to the payload data handling unit. From there the data will be transmitted to Earth immediately or stored for later transmission according to a pre-defined priority scheme. All of this on-board processing (together with telemetry losses) will determine the selection function for the survey carried out by Gaia. A detailed understanding of the selection function is critical to both the data processing and eventually the scientific exploitation of the final catalogue data.

Therefore the DPAC needs enough information on both the video processing unit (VPU) and payload data handling unit (PDHU) on-board processing algorithms such that all the stages from detection to data transmission can be simulated in detail. For a detailed list of the information needed, see [Bab07].

The information should be in the form of the (pseudo) source-code mimicking the exact behaviour of the on-board processing. Before those algorithms are actually available, detailed descriptions of the algorithms under development are needed.

D.4 Telemetry specifications

The DPAC processing will of course operate on the data collected by the instruments on the Gaia spacecraft and ESA is expected to deliver all the necessary telemetry to DPAC through ESOC-MOC and/or by direct transfer to ESAC. Data transferred directly to ESAC or through ESOC-MOC should include not only the telemetry from the scientific instruments, but satellite 'house-keeping' telemetry, and any data and reports generated by the on-ground processing performed by ESOC.

The data processing starts with the unpacking and ingestion of the telemetry received by ESAC-SOC and a detailed knowledge of both its contents and format is of course mandatory. The information on the telemetry contents and format should be contained in a MOC-SOC interface control document. Recently a proposal for the telemetry was put forward by members of the DPAC (see [PFea07]).

D.5 On-ground and commissioning calibrations and tests

A critical input to the DPAC calibration procedures for all Gaia instruments will be their detailed characterization before launch, and then during commissioning, by specifically designed calibration measurements. Note that a number of the calibration measurements can only be obtained through (laboratory) experiments on ground (for example, the characterization of the response of individual pixels on the CCDs).

The following is neither a complete nor fully detailed list of all the required calibration data. For a detailed list see [Bro07]. The latter document also describes the relation between the calibration requests and the calibration plan under development by EADS-Astrium and ESA.

D.5.1 Pre-launch (laboratory) calibration data

This section refers to data which should be provided by EADS-Astrium before the launch of Gaia in order to support the data processing preparations. The term 'calibration data' should be understood in a broad sense and covers: laboratory measurements, information obtained from manufacturers/sub-contractors, optical prescriptions, estimates of certain parameters based on experience or known physics, analysis results, etc. The requested calibration data will in some cases overlap with the requested instrument specifications.

Here a summary is provided of the calibration data that is requested. For a complete and detailed list see [Bro07].

- A detailed understanding of the CCDs requires full calibration data, derived from measurements made on each CCD, including CCD QE, MTF and CTI properties. Where relevant the calibration data should be provided for the CCD and proximity electronics module combination.

- The PSF and LSF should be measured in both imaging and TDI mode as a function of focal plane position for all instruments.
- Stray light and ghosts should be mapped per CCD and as a function of AC position.
- The calibration of the photometric throughput of the instruments requires knowledge of the transmission curves of the optical elements of the instruments and their expected evolution.
- In order to calibrate the wavelength scale for the BP/RP and RVS spectra, optical properties of the dispersive optical elements must have calibration measurements made, including transmission curves and wavelength calibrations to be measured in TDI mode for the actual focal plane positions of the detectors.
- The instruments have to be characterized in terms of polarization.
- Finally, the basic angle monitoring device and wavefront sensors should also be calibrated on ground and the results made available to DPAC.

D.5.2 Post-launch commissioning diagnostics, procedures, and data

This section refers to diagnostics and procedures that should be performed by EADS-Astrium and ESA during the commissioning phase in order to assure the correct functioning of the satellite and the instruments.

EADS-Astrium and ESA should deliver a detailed test plan that specifies, which procedures are performed during the commissioning phase. A list of the items to be calibrated and/or verified during the commissioning phase can be found in [Bro07]. For the case that some of the verifications fail, a detailed plan should specify which parameters can be uploaded and what changes/adjustments to the instruments are possible.

Descriptions of all special diagnostic algorithms that are planned for the commissioning phase should be made available in detail so that the First Look Monitor and Evaluators can perform these diagnostics during the commissioning phase and during the operation phase of the mission (on a regular basis or when needed in case that the data indicate non-optimal functioning of instruments).

All HK and science data acquired in the commissioning phase should be made available.

E Qualifications and experience of DPACE members

Below are descriptions of the qualifications and experience of the DPACE members, which includes all the CU managers.

E.1 Francois Mignard, *DPACE Chair*

Trained in astrometry and celestial mechanics, F. Mignard has been involved in the development in several areas of the FAST data processing for Hipparcos (general simulation up to the launch, astrometry, photometry, multiple stars algorithms in the operational treatment) and the production of the final Hipparcos catalogue and documentation. Member of the Hipparcos Science Team (1991-1996) he chaired the Double and Mutliple star working group in charge of producing the combined solution from FAST and NDAC. Managing experience as director of CERGA (3 terms from 1992 to 2003) and many positions in various research councils and advisory groups in France or abroad (eg NASA SIM Science WG), including a 2-term elected position in the French National Committee for Scientific Research and the chairmanship of the High Scientific Committee of the Observatory of Paris.

Member of the SAG from its inception and then later the GST, where he is currently a member. F. Mignard led the Solar system and the Relativity and Reference Frame Working Groups until they were dissolved in December 2005 and was chair of the DACC, until it was succeeded by the DPACE in June 2006. In the DPAC he manages WPs related to Solar System objects in CU4 and the production of Solar System ephemeris for Gaia, in CU3. He is also the national PI for the French participation to the DPAC.

E.2 Ronald Drimmel, *DPACE Deputy Chair*

R. Drimmel received his PhD from the University of Florida in 1995, then transfered to the Astronomical Observatory of Torino in 1996, where he currently holds a tenured position as Research Astronomer. His research interests are in the areas of Galactic structure and dynamics, with emphasis on nonaxisymetric structures, and the galactic distribution of interstellar dust. His interests in space astrometry began with the analysis of Hipparcos data.

R. Drimmel began contributing to Gaia in 1997 during the preparation of the Gaia Study Report. He was an active member of the Photometric and Simulations Working Groups until their disbanding in 2005. Currently he is a WP manager in both CU2 (Simulations), providing the Galactic extinction model, and CU8 (Astrophysical Parameters), where he manages the top-level WP advising on the estimation of interstellar extinction. He has contributed to the Gaia Source Count model, provided to

ESA for making telemetry flow estimations. He is also a member of CU3 (Core Processing). Before being nominated to the DPACE, R. Drimmel also served on the DACC in 2005-2006. He is also the Project Manager for the Italian national funding proposal, and the Scientist in Charge for INAF in the Marie Curie RTN ELSA (European Leadership in Space Astrometry).

E.3 William O'Mullane, *CU1 manager*

W. O'Mullane has an academic background in computer science and over a decade of experience in space missions and space science. He has worked at ESOC on the SCOSII system and contributed to the production of the CDROMs for the Hipparcos catalogue. He has created acclaimed visualisation software for the Hipparcos catalogues and contributed to the quality assurance tools for the Guide Star Catalogue. More recently he has worked on image processing and catalogue interrogation software for the Sloan Digital Sky Survey at the Johns Hopkins University. O'Mullane was also a group leader in the US National Virtual Observatory (NVO) contributing to the overall architecture and technology decisions made by NVO. Outside space science he also has industrial experience in high availability dynamic websites with high user loads.

W. O'Mullane became involved in Gaia as Hipparcos wrapped up in 1997. He has been a member of the DACC since its formation and is now a DPACE member. He has been the ESAC Gaia manager since the team was created in August 2005.

As CU1 leader O'Mullane takes responsibility for the overall architecture of the processing system as well as a few core software items such as the MDB. He has been setting down ideas for the architecture of the system and the GIS since 1997. As ESAC DPC leader contributions will continue in the area of the Astrometric GIS which he has also worked on since 1998.

E.4 Xavier Luri, *CU2 manager*

After obtaining a degree on physics, X. Luri did his Ph.D on astrophysics, developing new statistical methods for the exploitation of the Hipparcos catalogue. This was his first contact with space astrometry, a field he has been working ever since. He also has a wide experience in software development and computer systems (acquired from academic formation and practical experience in software projects, both in private companies and public research centres) and space systems. Dr. Luri presently holds a tenure position at the Department of Astronomy and Meteorology of the University of Barcelona, where he teaches Astronomy, Mathematics and Statistics. He is member of the Spanish Astronomical Society, where he chairs the committee in charge of outreach, and vice-president of the Spanish Institute of Navigation.

Member of the Gaia Science Advisory Group since its early stages and then later the Gaia Science Team, where he is currently a member. He led the Simulation Working Group, providing simulation support to the mission design activities, until it was replaced by the Coordination Unit 2 of the DPAC, which he presently manages. As such, he is also a native member of the DPACE.

E.5 Ulrich Bastian, *CU3 manager*

Trained in theoretical astrophysics and fundamental astrometry, U. Bastian has been involved in the production of the FK5 and PPM star catalogues. He has contributed to the FAST data processing and Input Catalogue production for Hipparcos in several areas and to the production and documentation of the final Hipparcos Catalogue. He had a leading role in the data processing and catalogue production for the Tycho experiment of Hipparcos. He was a member of the Hipparcos Science Team 1990–1997. From 1997 to 2002 he was instrument scientist and astrometry coordinator in the development of the German Hipparcos follow-up project DIVA (which was eventually canceled for financial reasons).

U. Bastian has been involved in the development of Gaia since 1993. He has been a member of the GST since 2002, and he was a member of the DACC from its formation until it was succeeded by the DPACE in June 2006. He has been the scientific manager of CU3 since its constitution in February 2003. He is also the national coordinator for the German participation to the DPAC.

Within CU3 U. Bastian contributes to the primary astrometric reduction and to the definition of the IDT, of the First Look, of the astrometric calibration model and of the astrometric interfaces to the other CUs. From 1993 to mid 1996 he led the development of the First Look task. He directs the efforts towards ground-based optical tracking of Gaia for orbit improvement and chairs the Gaia working group at ARI, Heidelberg.

E.6 Dimitri Pourbaix, *CU4 manager*

D. Pourbaix received his PhD from the Université de Liège in 1998, then transfered to the Université Libre de Bruxelles under PRODEX funding and, finally, the Belgian National Funds for Scientific Research (FNRS) in 2000 where he holds a tenured position since 2002. His research interests focus on orbit modelling of binaries, essentially astrometric, visual, and spectroscopic ones. His interests in space astrometry began with the post-processing of the Hipparcos data. Since 2003, D. Pourbaix has been the leader of International Astronomical Union working group in charge of the preparation and release of a catalogue of spectroscopic binary orbits.

D. Pourbaix began contributing to Gaia in 2001. He ended up as co-leader of the

Double and Multiple Stars and Extrasolar Planets Working Groups until their disbanding in 2005. He took part in the blind test aimed at quantifying the capabilities of Gaia in the context of extrasolar planets. He was a member of the DACC from its formation until it was succeeded by the DPACE in June 2006. Besides the scientific management of CU4, he contributes to several WP of CU4 dealing with astrometric orbit of both resolved and unresolved non-single stars. He is also the Belgian node in the Marie Curie RTN ELSA (European Leadership in Space Astrometry)

E.7 Floor van Leeuwen, *CU5 manager*

Floor van Leeuwen was trained in Astrometry and Photometry at Leiden University, where he was a student of S. Vasilevskis and A. Blaauw. His PhD thesis at Leiden University 1983 was on an astrometric and photometric study of the Pleiades cluster. He was involved in the Hipparcos mission at the Royal Greenwich Observatory, first under Andrew Murray, and from 1985 onwards as team leader, till the publication of the Hipparcos data in 1997. He published several key papers on the Hipparcos mission (a major review in Space Science Reviews in 1997, papers on specific use of the astrometric data, and papers on the derivation of astrometric parameters for open clusters). He is currently rounding of a complete re-reduction of the Hipparcos astrometric data, which, through taking into account dynamical properties of the satellite, has reduced the noise levels on the astrometric data for the brightest stars by a factor 3 to 5. These data will be published early 2007. He has also been data processing manager for the Planck satellite commitments in Cambridge from 1998 till 2006. He has been a member of the Hipparcos Science Team from 1984 till 1997, and of the Gaia Science team since March 2003.

The CU structure for the data processing was first proposed by him at a meeting in Cambridge in February 2005, and then presented in a document which was further developed by Michael Perryman and later by Francois Mignard and Coryn Bailer-Jones. As a member of the DACC he helped structuring the DPAC. In the UK he has been leading the (successful) grant application for the Gaia data processing. Based on his experience with the Hipparcos re-reduction he alerted the astrometric processing to the problems associated with density variations on the sky in obtaining a reliable astrometric solution, and industry on potential problems associated with the rigidity of the satellite.

E.8 David Katz, *CU6 manager*

David Katz his “astronome adjoint” at the Paris Observatory. He was trained in astrophysics, obtained his PhD in 2000 at the Paris Observatory and was Research Fellow at ESA-ESTEC for one year. His main fields of research are: (i) the automated analysis of the stellar spectra and (ii) the study of the structure and formation of the Galaxy and in particular of the Thick Disk. David Katz is teaching observational spectroscopy at the Master of Astronomy of the Paris Observatory.

David Katz started to work on Gaia in 1999 on the assessment of the performance of the Radial Velocity Spectrometer (RVS) and on the optimisation of its design. In 2001, he was appointed coordinator of the RVS working group as well as member of the Gaia Science Team (GST). In 2006, when the working groups were replaced by the coordination units, he became coordinator of the Coordination Unit CU6 spectroscopic processing. He has been member of the DACC until its replacement by the DPACE.

E.9 Laurent Eyer, *CU7 manager*

Laurent Eyer obtained his PhD in 1998 at the Observatoire de Genève, Switzerland, working with Prof. Michel Grenon on the analysis of the variability data from Hipparcos mission, resulting in the Hipparcos Variability Annex and Light Curves (Volume 11 and 12 of the Hipparcos Catalogue). During his first postdoc, in the Catholic University of Leuven, he worked on exploiting Hipparcos findings on certain variable stars. A subsequent post-doctoral position at Princeton University was focused on the analysis of the ASAS, OGLE data, providing experience in the handling of massive photometric data sets of direct relevance to Gaia and produced diverse scientific outputs, e.g. detection, classification and studies of variable stars and asteroids, search for quasars, study of the Galactic structure, proper motion studies, statistical methods on time series analysis, etc. Laurent Eyer obtained a tenure position at Geneva Observatory in 2006 and is member of the Organising Committee of IAU commission 27 (Variable Stars) and 45 (Stellar Classification).

Laurent Eyer has contributed to Gaia since 1998 and became the task-leader of the Variable Star Working Group in 2002. He wrote most of the code used for the Gaia-Grid test. He is the manager of CU7 and leads the Swiss team in Geneva. He is the Swiss node for the Marie Curie Research Training Network ESLA.

E.10 Coryn Bailer–Jones, *CU8 manager*

Coryn Bailer-Jones obtained his PhD from the University of Cambridge in 1997 having worked on the automated classification of stellar spectra. After a postdoc working on the modelling of materials processing at the Cavendish Laboratory in Cambridge, he moved to the Max-Planck-Institute for Astronomy (MPIA) in Heidelberg, Germany in 1998. Other than a one year research visit to Carnegie Mellon University in Pittsburgh (USA) in 2003, he was been at MPIA up until the present day. His research has focused on the development and application of classification and parametrization methods, not just for Gaia but also for SDSS and the (now canceled) German mission DIVA. His experience puts him at the interface between astronomy, instrumentation and statistical data analysis. His other research areas are brown dwarfs and star formation. Coryn presently leads a DFG-funded research group and has a DLR-funded group comprising three postdocs for the CU8 work.

Coryn was involved in the study phase of Gaia between 1998 and 2000, in particular on the photometric/spectroscopic instrument. Starting with the phase A study in 2001, he was appointed to lead the newly formed “Classification Working Group”. In 2001 he was also appointed to the newly constituted Gaia Science Team (GST), a position he still holds. Here he contributed to the overall design and assessment of the Gaia mission, in particular to the design and assessment of the photometric system. In 2005 he was asked by the GST to co-chair the “Data Analysis Coordination Committee” (DACC), the body which oversaw the transition from the informal working group structure to the present Data Processing and Analysis Consortium (DPAC).

E.11 Claude Huc, *CNES Gaia Data processing Center manager*

Claude is a physicist engineer by training. He joined the CNES (French Space Agency) in 1973 to take charge of the design and development of systems for processing and archiving data from European scientific space missions. During the eighties, he was technical manager for the Hipparcos data processing system developed in the framework of the FAST consortium then integrated, validated and operated at the CNES computing centre. This cooperation between CNES and European laboratories within the FAST consortium was most productive Since the early nineties, his activity has been focused on the management, long term preservation, recovery and exploitation of space science data. In this context, various Research & Technology studies have been conducted on the generic nature of systems; he contributed to the creation of specialised CNES services in this field. He also participates in standardisation activities in the field of data archiving. He has been in charge of a department created by CNES to exploit, preserve and add value to space data. Since 1998, he was also the technical manager of CDPP (Plasma Physics data centre).

Claude Huc has been the manager of the CNES Gaia Data processing center until march 2007. His first activity for Gaia since the end of 2004 was assigned to prepare a proposal outlining CNES’s involvement in Gaia data processing in conjunction with the scientific community, the European partners and CNES Programmes Directorate and to obtain for this involvement, the appropriate human and financial resources.

E.12 Xavier Passot, *CNES Gaia Data processing Center manager*

Xavier is a computer scientist engineer by training (1976). He joined the CNES (French Space Agency) in 1983 to take charge of the development of a part of the control center of the SPOT1 earth observation satellite. Later, he has been in charge of the SPOT4 on board software, then of the Locstar ground segment. He has been participating in the Hermes spaceplane adventure as an information manager of the system data. He left CNES for two years to be director of a division in a small company focused on local communities accounting software. He came back in CNES in 1993 to take in charge the development of the VEGETATION image processing center, subcontracted to the belgian and swedish software industry ; he remained in the VEGETATION2 earth observation project until 2002 as a system engineer. Later, he has been manager of the monitoring and control department in the Myriade microsatellites project. He has also been participating in standardisation activities in the field of monitoring and control (ECSS, CNES). He joined the GAIA CNES team in February 2007.

Xavier Passot is the manager of the CNES Gaia Data processing center since march 2007. His first activity for Gaia was the participation of the CNES internal review of the end of phase A of the Gaia DPC.

F Qualifications and experience of CU deputies

Below are descriptions of the qualifications and experience of the deputies of every CU.

F.1 Uwe Lammers, *CU1 deputy manager*

U. Lammers has an academic education in physics and computer science and he holds a PhD in physics. Since graduation he has been working for ESA on several space mission projects in the areas of science operations centre development and design and implementation of large-scale space science data processing systems. Before working on Gaia he was involved in the development of the Science Analysis System software for ESA's XMM-Newton X-ray observatory.

U. Lammers contributes to the Gaia mission since around mid-2003 first as a member of the Project Scientist Support time in ESTEC, The Netherlands. In this position he was working on the modelling and prediction of telemetry data rates and volumes. In September 2005 he joined the newly established Gaia Science Operations Centre development team at ESA's European Space Astronomy Centre (ESAC) in Spain. His main area of responsibility include the planning, coordination, and implementation of the Astrometric Global Iterative Solution (AGIS) system (Sect. 5.1) and general DPAC support in the context of CU1. The role as CU1 deputy manager he shares with T. Levoir.

F.2 Thierry Levoir, *CU1 deputy manager*

T. Levoir has an academic education in computer science and mathematics. Since graduation in 1993, he has been working for CNES on several projects in the areas of archives and data processing, either within a Research and Development framework or for operational projects.

T. Levoir has been member of the CNES Gaia Data processing center team since the beginning of 2006. As system leader of the team he covers the architecture (both hardware and software) of the CNES-DPC system for CU4, CU6 and CU8 and he represents the CNES in the CU1 where he shares the role of deputy manager with U. Lammers.

F.3 Carine Babusiaux, *CU2 deputy manager*

C. Babusiaux completed her PhD in 2003 on ”Photometric Studies of the Milky Way” at the Institute of Astronomy, Cambridge. She made a first postdoc in the Université

Libre de Bruxelles (PRODEX funding) and a second in Paris Observatory (CNES postdoc). She now has a permanent position in Paris Observatory. Her research area is on the structure, formation and evolution of the Milky Way and of the Local Group galaxies through the study of their resolved stellar populations, with emphasis on the central regions of our Galaxy.

C. Babusiaux begun her involvement in Gaia with the development of an on-board algorithm prototype in 1999. She then created, designed and developed the Gaia GIBIS simulator. Since 2001 she co-coordinated the Simulation Working Group and the On-Board Detection Working Group. She is a member of the Gaia Science Team since January 2006, co-manager of the CU2 and manager of the GIBIS simulator.

F.4 Mario G. Lattanzi, *CU3 deputy manager*

M. G. Lattanzi is Senior Associate Astronomer at INAF-Osservatorio Astronomico di Torino (INAF-OATo), lecturer in fundamental astronomy at the University of Turin and leader of the *AstroGalTech* (Astrometry, Galactic-studies and Technology) group at INAF-OATo. His relevant experience includes:
– Development and implementation of the Hipparcos Sphere Reconstruction code in FAST (with B. Bucciarelli and M. Froeschle) and development of the alternative approach named GLOBUS (with F. Sanso, B. Betti, and B. Bucciarelli), 1983-1991.
– Membership of the FAST Committee, for the early evaluation of the results of the Hipparcos mission for the FAST Consortium, 1989-1996.
– Instrument Scientist (1989-1995) and STScI consultant (1995-1997) for the photometric and astrometric calibration and scientific exploitation of the astrometric Fine Guidance Sensor interferometer on board HST.
– Co-PI of the GSC II project, the realization of the largest and deepest astrometric-photometric survey of the whole sky in the visible spectral range to date, 1997-2002.

M. Lattanzi was member of the Gaia Science Advisory Group, and later the Gaia Science Team, from 1997-2005. He is also the PI of the Italian participation to the Gaia mission. His specific contributions to Gaia include:
– Development of the Gaia science case: extra-solar planets, tests of General Relativity, and the study of the galactic disk; development of the astrometric payload: requirements, optics design, basic-angle monitoring (concept, materialization, characterization), chromaticity, calibrations, astrometric instrument simulator; development of the astrometric model (General Relativity, general and instrumental parameters) for the sphere reconstruction.
– Responsible for the development, implementation and operation of AVU and the INAF-OATo DPC.
– Contribution to the CU3 workpackage REMAT.
– Contribution to CU4 (astrometric detection and characterization of extrasolar planets).

– Implementation and use of the GSC II data base as Gaia auxiliary catalogue.

F.5 **Jordi Torra,** *CU3 deputy manager*

After his Ph.D. on galactic kinematics, J. Torra was involved in the Input Catalogue tasks for the Hipparcos mission as well as in its scientific exploitation. He contributed to the preparation of the catalogue of the Optical Monitoring Camera onboard Integral, based on Hipparcos and Tycho catalogues. He has been involved in the Gaia mission since its very beginning by contributing to the *Red Book*, [gai00] and to the first Gaia simulations. His main contribution has been in the development of the GDAAS system producing the first tests on the main Gaia processes. He has managing experience as PI of the group at the UB and as co-director of the IEEC, and he has collaborated in several research councils (Spanish National Commission of Astronomy) and advisory groups (now at the Science Advisory Group of the GTC 10m telescope).

As member and deputy manager of CU3 he is in charge of the development and implementation of both the Initial Data Treatment and the Intermediate-Data Updating. He was member of the Data Processing Working Group and member of the DACC until the creation of the DPACE. He is also the national PI for the Spanish participation to the DPAC.

F.6 **Paolo Tanga,** *CU4 deputy manager*

The research activity of P. Tanga has mainly concerned, on one side, the early stages of planetary formation and, on the other, the collisional evolution of the asteroid belt. On this last topic, he has participated to the modelization of asteroid shapes by interferometric observations using the Hubble Space Telescope FGS. P. Tanga has also developed softwares for the analysis of different kind of data, including interferometry and imaging.

P. Tanga was involved in the Solar System Working Group activity since the beginning. He has contributed to the simulations of Solar System objects, and to study different aspects of the impact of Gaia on Solar System science: photometric inversion, allowing to recover asteroid shapes from light variations; orbit improvement and consequences on occultation studies; measurement of Yarkovsky effect. He is in charge of a management WP for Solar System processing, and is involved in the activity of other CU4 work packages, devoted in particular to simulations and photometric inversion. He will also provide Solar System models for the CU2 activities in the frame of a specific WP.

F.7 Anthony Brown, *CU5 management team*

Anthony Brown was trained in astronomy at Leiden University. His PhD thesis, directed by P.T. de Zeeuw, at the same university in 1996 was on a study of the nearby OB associations using ground based data which was collected in preparation for the Hipparcos mission results. Subsequently he worked with M.A.C. Perryman on a detailed study of the Hipparcos data of the Hyades cluster and was closely involved with the Hipparcos census of the nearby OB associations carried out by J.H.J. De Bruijne and R. Hoogerwerf. During his time as a postdoc at UNAM (Mexico, 1997–1999) he worked on the detection of the remnants of disrupted satellite galaxies in the halo of our galaxy from the Gaia catalogue. This was done by creating a large and detailed Monte Carlo simulation of the Gaia catalogue with a realistic number of entries (3.5×10^8 stars). From 2001–2005 he was involved in the development of the SINFONI instrument for ESO's VLT. He worked on detailed simulations of the adaptive optics module for this instrument.

Brown has been involved with Gaia since 1997 and started out by contributing to the science case for the Concept and Technology Study Report. In 1999 he wrote one of the first detailed simulations of Gaia measurements at the detector level. Elements of these simulations were incorporated in the development of GIBIS. He was a member of the photometric and classification working groups until their disbanding in 2005. In the photometric working group he was very closely involved with the design and optimization of the photometric system resulting in the C1B+C1M filters which were adopted as the baseline at the end of 2004. In 2003 he put together a first planning of the photometric data processing for Gaia, providing detailed work packages. This planning has subsequently been modified and incorporated into the work breakdown structure of CU5. At the introduction of the new photometric instrument (BP/RP) he produced detailed simulations of the prism spectra which have now been incorporated into the CU2 simulation infrastructure. Brown was a member of the Gaia data access and analysis system steering committee, where he was responsible for overseeing the photometric aspects of the Gaia data processing prototype. In 2005-2006 he was a member of the DACC and at the end of 2005 he joined the Gaia science team.

F.8 Carla Cacciari, *CU5 management team*

Carla Cacciari has management experience related to both space and technical projects, including the following:
– June 1980 to September 1983: ESA Research Fellow at IUE-VILSPA (Madrid, Spain), with duties of real time observing preparation and execution, and in charge of program scheduling.
– August 1984 to August 1988 & October 1996 to April 1998: ESA Resident Astronomer at the STScI (Baltimore, USA), with duties of general observers' support

(definition of policies and procedures, program preparation, observing time allocation process).
– June 1991 to September 1996: Chair of the Working Group for the organization of the Observing Program Support for Spectrum UV.
– May 1999 to December 2002: PI of the Italian Consortium for FLAMES.

In the DPAC C. Cacciari is the Manager of the DU14 (Instrument absolute response characterisation: definition and application).

F.9 Carme Jordi, *CU5 management team*

C. Jordi obtained the degree on Physics at the University of Barcelona. Her PhD in 1985 was on astrometry, deriving the fluctuations of the Earth's rotation from lunar occultations. Soon, she started to work on the photometric and astrometric analysis of open clusters and on photometric tasks related with the preparation of the Input Catalogue for the Hipparcos mission. She has also a wide experience in photometric and spectroscopic observations (including Gemini-N and Hubble Space telescopes) and in the determination of stellar astrophysical parameters. The use of M31 eclipsing binary data has recently yielded the first direct determination of the distance to our neighbour galaxy.

C. Jordi presently holds a tenure position at the Department of Astronomy and Meteorology of the University of Barcelona. Managing experience includes 4 years of secretary position of the Department, coordination of software and hardware of the Department for 15 years, PI of research projects in galactic structure and stellar parameters for 4 years. Besides Gaia, her space research experience includes involvements in the Input Catalogues of Hipparcos and of the Optical Camera of INTEGRAL, small contributions to COROT and a proposal for the Spanish Minisat-02.

Co-leader of the Photometric Working Group since its creation in 1999 and until its dissolution and member of the Gaia Science Team since 2002, her main involvement has been in the design of the photometric instrument. She has also contributed to the development of GDAAS prototype. In the DPAC, she manages WPs related to the photometric instrument calibration and she is one of the scientific managers of CU5.

F.10 Dafydd Wyn Evans, *CU7 deputy manager*

D.W.Evans obtained his PhD in Astronomy from the University of Cambridge in 1988 where he was trained in astrometry and photometry. From 1988 to 1997 he was involved in the data analysis for Hipparcos and was leader of the Photometric Working Group. He was responsible for the merging of the FAST and NDAC photometric data and helped produce the Hipparcos Catalogue Epoch Photometry Annex. Since then he has worked on the Carlsberg Meridian Telescope, La Palma, where he was on the

Management Committee, writing the reduction software responsible for calibrating the astrometry and photometry. The latest catalogue (CMC14) was released last year. He has also worked on WFCAM, a wide-field camera for UKIRT, where he developed the PSF fitting software. He is currently the Vice President of IAU Commission 8 (Astrometry) and is on the Organizing Committee of IAU Division I (Fundamental Astronomy).

During the early part of the project, Evans was co-leader of the Variable Star Working Group and a member of the Photometry Working Group. He also had strong interests in the Simulation Working Group and produced a large set of simulated photometry of variable stars that was used to test GaiaGrid. Within DPAC, in addition to his role as CU7 deputy manager, Evans is manager of DU15 (Internal photometric calibration and its application) for CU5.

F.11 Pierre Dubath, *CU7 deputy manager*

Pierre Dubath obtained his PhD with Prof. Michel Mayor in 1992 at the Geneva Observatory. He also worked for ESO, La Silla in Chile during his PhD, and later carried out two postdoctoral stays, first at ESO headquarter in Garching and then at the Lick Observatory, University of California Santa-Cruz. He then moved back to Geneva to work at the INTEGRAL Science Data Centre (ISDC). During the first part of his career, before being heavily involved by INTEGRAL related works, Pierre Dubath developed radial velocity and metallicity determination techniques, and apply them to spectroscopic data coming from large telescopes, such as the ESO NTT/3.6m and the Keck. He used these techniques together with CCD photometry to achieve studies of the dynamics and of the stellar populations of globular clusters and of dwarf galaxies.

Pierre Dubath's experience with spectroscopic and photometric CCD data analysis and with INTEGRAL related software development and data analysis is a very valuable asset in the context of the Gaia data analysis. He has been involved in CU7 since the beginning of 2006, helping Laurent Eyer to setup and organise the CU7 team. He is currently working out the basis CU7 data model and software architecture, trying to extract the best from the ISDC experience and from the concepts and tools proposed by the CU1 team.

F.12 Fredéric Thévenin, *CU8 deputy manager*

Trained in spectroscopic techniques and in the theory of the atmospheres of late-type stars at the observatory of Paris-Meudon, F. Thévenin has been involved in the development of the technique of spectrographs with multi-slits aperture. He participated to studies of chemical element evolution in the Galaxy, and in particular in diagnostic of thermodynamical equilibrium of stellar atmospheres. He did some investigations in stellar interior physics with some applications to asteroseismology. More recently he was involved in stellar diameter measurements with interferometry. He had an experience of managing as deputy of a laboratory of the CNRS and also of the Société Française d'Astronomie et d'Astrophysique. He has been member of the Conseil National des Universits (section astronomy).

Member of the RVS working group (2001-2005), he actively participated to the definition of the instrument RVS in particular in the choice of the resolution. He also led the first Functional Analysis of the data reduction of Gaia with some managers and the CNES. He proposed to create a group for pushing the theory of stellar atmospheres in more realistic ways and to apply the results to the preparation of the interpretation of the spectra of Gaia. Apart from his position of deputy he is also managing the data training in CU8 and is member of the FLAME group (CU8) and of the Functional analysis group of CU6.

G AO Compliance Matrix

G.1 Functional Requirements

Reqirement	Text	Sections
FUNC-3.1-1	The DPAC shall develop all algorithms and processing systems required for the scientific processing of Gaia data and the production of all Gaia products.	Part II.
FUNC-3.1-2	The DPAC shall provide all infrastructure required for their processing systems, also integrating those elements provided by ESAC.	Sect. 9 Part III.
FUNC-3.1-3	The DPAC shall operate the processing systems until the final Gaia products are produced and validated.	Sect. 7.5.1 Appendix. A . Effort and phases are up to 2020.
FUNC-3.1-4	The DPAC shall validate and document the intermediate and final Gaia products.	Sect. 2.1.4 Sect. 10.4.1.
FUNC-3.1-5	The DPAC shall define the milestones and schedule for the data reduction activities, in accordance with the project scientist and the Gaia Science Team.	Sect. 10.4.1.
FUNC-3.1-6	The DPAC shall produce intermediate and final Gaia products, according to the schedule and content defined in accordance with the project scientist and the Gaia Science Team.	Sect. 2.1.4 Sect. 10.4.1.
FUNC-3.1-7	The final Gaia products shall consists of at least astrometric, photometric and spectroscopic data with accompanying interrogation tools. Access tools and mechanisms will be defined at a later date (within CU9; see section 2).	Sect. 2.1.
FUNC-3.1-8	The DPAC shall assign to one centre, the hub, responsibilities for the single point of contact with the ESA Project Team during the development phase (up to and including commissioning), and with the MOC and the mission manager during the operational phase.	Sect. 7.3 Sect. 9.2.
FUNC-3.1-9	The hub shall receive and process telemetry from the MOC.	Sect. 9.1.
FUNC-3.1-10	The hub shall provide feedback to the MOC of required changes to the timeline resulting from first look analysis of the science telemetry.	Sect. 7.2.4 Sect. 7.2.5.
FUNC-3.1-11	The hub shall provide updated calibration information to the MOC.	Sect. 4.2 Sect. 7.2.4 Sect. 7.2.5.
FUNC-3.1-12	The hub shall define the Interface Control Document governing periodic data transfer between the hub and the DPCs, and vice versa.	Sect. 7.1 Sect. 7.3 Sect. 7.4.
FUNC-3.1-13	The hub shall periodically send data to the DPCs according to the governing Interface Control Document and to the schedule referred to in FUNC-3.1-5 and FUNC-3.1-6.	Sect. 7.3.
FUNC-3.1-14	The hub shall integrate reduced science data from the DPCs according to the defined and agreed integration rules.	Sect. 7.4 Sect. 7.4.2.
FUNC-3.1-15	The hub shall be the primary point of distribution of all intermediate and final Gaia products.	Sect. 7.3.
FUNC-3.1-16	The hub shall provide archive and interrogation systems for intermediate and final Gaia products.	Sect. 8.10.
FUNC-3.1-17	The CUs shall develop, validate, and document the processing systems according to their defined role.	Sect. 8.1.
FUNC-3.1-18	The CUs shall deliver the processing systems to the relevant DPC. FUNC-3.1-19 The CUs shall maintain their processing systems until the final Gaia products are produced and validated.	Sect. 8.1.
FUNC-3.1-20	Data Processing Centres shall be responsible for the integration and operation of the processing systems under their remit.	Sect. 8.1.
FUNC-3.1-21	The Data Processing Centres shall receive and process data from the hub according to their defined function.	Sect. 9.
FUNC-3.1-22	The Data Processing Centres shall send the newly-reduced science data to the hub for integration, according to the schedule defined by the DPACE.	Sect. 9.

G.2 Performance Requirements

Reqirement	Text	Sections
PERF-3.2-1	The Data Processing Ground Segment facilities shall be sufficient to produce all intermediate and final Gaia products according to the agreed schedule, in particular the final Gaia data products shall be available three years after the end of the operational phase of the mission.	Sect. 7.5 Sect. 2.1.3 Sect. 2.1.4.
PERF-3.2-2	The Data Processing Ground Segment facilities shall be dimensioned in such a way that they can support without re-design an extension of at least one year of the in-orbit operations.	Sect. 7.5.
PERF-3.2-3	The overall availability figure for the Processing Systems, and connectivity to these systems, as located at the Gaia hub, shall be 95% minimum.	Sect. 9.1.1 Sect. 9.1.2.

G.3 Product and Quality Assurance Requirements

Reqirement	Text	Sections
PAQA-3.3-1	During all phases of the Gaia mission implementation (i.e., design, development, integration and test of the total ground segment both hardware and software) each contributor shall carry out a Product Assurance/Quality Assurance (PA/QA) activity.	Sect. 7.7.1.
PAQA-3.3-2	The PA/QA activity shall also be exercised throughout the operations phase of the mission to ensure that all changes to Processing Systems are carried out in accordance with a formal change control procedure.	Sect. 7.7.1.
PAQA-3.3-3	The PA/QA aspects shall be addressed at each review of the various components (i.e., hub and DPCs) of the Ground Segment as well as during the reviews of the entire Ground Segment and the Mission Level reviews.	Sect. 7.5.4.
PAQA-3.3-4	The DPAC shall (as part of the PA/QA function) carry out a risk assessment of their overall activities. Practical risk mitigation measures shall be identified and implemented. The status shall be reported as part of regular reporting. This aspect shall be addressed on the occasion of the relevant reviews.	Sect. 7.5.4 Sect. 10.4.1.1.
PAQA-3.3-5	Requirement specifications, design specifications, test specifications, interface control documents and user manuals shall be produced as required.	Sect. 7.7.1.
PAQA-3.3-6	Implementation plans and procedures, test plans and procedures, and operations plans and procedures shall be produced for all Processing Systems produced by the DPAC.	Sect. 7.7.1.
PAQA-3.3-7	Software documentation shall conform to the ECSS standards. The ECSS standards must be tailored to each individual project's need. It is foreseen in the CU1 activities that the hub will carry out this tailoring process as part of its contribution to the DPAC. It is assumed that the DPAC as a whole will conform to the SOC (and Gaia project) tailored standards.	Sect. 7.7.1.
PAQA-3.3-8	The Gaia Ground Segment documentation, which must be accessible to all participants in the programme, shall conform to the electronic standards defined for the Gaia project. These standards will be defined and agreed jointly by the DPAC, MOC, project scientist and the Gaia Project.	Sect. 7.7.1.
PAQA-3.3-9	All operational science functions of the Gaia Data Processing Ground Segment shall be tested and validated before launch.	Sect. 7.5.4.
PAQA-3.3-10	Subsystem and system tests shall be conducted according to approved test plans and test reports shall be issued.	Sect. 7.5.4.
PAQA-3.3-11	The operational elements of the Data Processing Ground Segment (hub and DPCs) shall be included, as required, in the Satellite Verification Tests in order to verify their interfaces with the satellite and the other elements of the ground segment.	Sect. 7.5.4.
PAQA-3.3-12	The operational elements of the Data Processing Ground Segment shall, where relevant, be included in the end-to-end tests which validate proper operations of the entire space-ground segment system.	Sect. 7.5.4.
PAQA-3.3-13	The hardware configurations (computers, work-stations, peripherals, LANs, communication equipment, etc.) of the operational elements of the Data Processing Ground Segment shall be maintained under configuration control according to the usually applicable ESA standards.	Sect. 7.5.3.
PAQA-3.3-14	All DPAC processing systems, documentation and data items shall be delivered for integration and archiving in accordance to the DPAC configuration control system.	Sect. 7.5.3.
PAQA-3.3-15	The DPAC implementation shall be carried out in accordance to a (common) Software Project Management Plan to be produced by the Gaia SOC in agreement with the DPAC.	Sect. 7.7.
PAQA-3.3-16	All DPAC processing systems elements shall be produced in accordance to a (common) Software Quality Assurance Plan to be produced by the Gaia SOC in agreement with the DPAC. The Software Quality Assurance Plan shall specify software coding standards applicable to the DPAC.	Sect. 7.7.

G.4 Management Requirements

Reqirement	Text	Sections
MNGT-3.4-1	The SOC development manager, who has the delegated responsibility for all SOC related matters, shall ensure timely delivery of all the SOC deliverables and timely execution of all SOC tasks specified in this document. In practice actual activities will be coordinated with the project scientist.	Not explicitly addressed in the Proposal; DPAC will comply.
MNGT-3.4-2	The DPACE shall be responsible for the monitoring of the DPAC activities in order to verify that the tasks specified in this document are carried out according to specification and schedule.	Sect. 10.4.
MNGT-3.4-3	The DPACE will liaise with the Project Team through the SOC team to verify that the tasks specified in this document are compatible with the overall Ground Segment development schedule.	Sect. 10.4.
MNGT-3.4-4	In response to the requirements specified in this document the DPAC shall issue a proposal (hereafter referred to as the proposal) to ESA. Contingent upon the acceptance of this proposal, this proposal will become the data processing part of the Science Implementation Plan (SIP; see section 3.). The SIP shall serve for monitoring progress of the tasks identified therein.	Sect. 10.4.2.
MNGT-3.4-5	Any change in the contents of the SIP (i.e. the proposal), after its acceptance, might imply changes in cost, schedule and/or performance of the corresponding science function; therefore any modification to the SIP, or tasks, or baseline identified therein shall be formally reviewed and approved by ESA.	Sect. 10.4.1.
MNGT-3.4-6	The proposal shall the define the schedule and plan for completion of Gaia data processing by the DPAC, including as a minimum (a) the overall schedule for activites of the DPAC; (b) the definition of the work packages; (c) the schedule for the complete set of work packages supported by the corresponding schedule planning; (d) the identification of the cost-driving parameters and the corresponding estimates of resources spread over time (manpower, computers and other investment and running expenditure);	Sect. 7.6 Appendix. B.
MNGT-3.4-7	Each work package defined in the proposal shall include the definition of (a) the objective of the work package; (b) the corresponding inputs and output; (c) deliverable items; (d) tasks specifically excluded; (e) progress measurement points; (f) start and completion criteria; (g) cover for the work package.	Appendix. B.
MNGT-3.4-8	The proposal shall define in detail the reporting mechanisms. These reports (hereafter referred to as management reports) shall be produced on a regular (typically quarterly) basis in a format and frequency to be agreed with ESA.	Sect. 10.4.1 Sect. 8.
MNGT-3.4-9	Management reports shall be prepared by each CU and DPC.	Sect. 8 Sect. 9.
MNGT-3.4-10	Management reports shall include the following information: (a) brief summary of the progress achieved since the previous reporting period; (b) concise description of the main problem areas, their criticality and anticipated impacts (e.g., delays in the schedule or non conformance with the requirements); (c) status of the technical design, of proposed solutions to the problem areas and of engineering, product assurance and testing activities; (d) per work package, the manpower usage showing actual versus planned and estimation at completion, with overall manpower usage chart; (e) update of the overall schedule with latest prediction of the completion dates of the identified milestones; (f) a list of relevant action items and their status.	Sect. 8.1.1.
MNGT-3.4-11	The managers of the relevant sub-units (CUs) of the selected consortium have standing invitations to the Gaia Science Team meeting with the status of observer. It is expected that these activities be foreseen in the proposed work breakdown structure and provide the budget to support this.	Sect. 8.1.1.
MGMT-3.4-12	The activities and schedule of the CUs and DPCs are coordinated by the DPACE. The DPACE should have the authority to represent the DPAC and take executive decisions on its behalf. The DPACE reports to the project scientist (who in this respect represents an ESA internal DPAC oversight group) and through him to the Gaia Science Team.	Sect. 10.4.1 Sect. 10.4.2.

H Contributors

Contributors to this document

Name	Name
F. De Angeli	M. Kontizas
C. Babusiaux	C. Lagerkvist
C.A.L. Bailer-Jones	U. Lammers
U. Bastian	M.G. Lattanzi
M. Beck	F. van Leeuwen
M. Biermann	T. Levoir
A.G.A. Brown	L. Lindegren
J. de Bruijne	X. Luri
C. Cacciari	E. Masana
A. Cellino	F. Mignard
J.F. Claeskens	W. O'Mullane
M. Cropper	X. Passot
R. Drimmel	D. Pourbaix
P. Dubath	A. Recio-Blanco
D.W. Evans	A. Robin
L. Eyer	S. Rosen
C. Fabricius	C. Soubiran
M. Gai	P. Tanga
C. Huc	F. Thévenin
S. Jordan	J. Torra
C. Jordi	A. Vallenari
D. Katz	A. Volpicelli
S. Klioner	J. Warell

I Notations

Definitions and notations used in this document

Name	Meaning
α	Right ascension (in the ICRS frame)
δ	Declination (in the ICRS frame)
π	Stellar parallax (when not the mathematical constant $\pi = 3.14159\cdots$)
ϖ	Alternate notation for stellar parallax (when not an orbital element)
μ_α	Proper motion component in right ascension
μ_δ	Proper motion component in declination
v_r	Radial velocity
ξ	Solar aspect angle (fixed at 45 deg.)
γ	Basic angle (BA) (nominally 106.5 deg.)
γ	PPN parameter for the spacetime curvature
η	Along-scan field angle in the SRS (=Scanning Reference System)
ζ	Across-scan field angle in the SRS (Sun > 0)
w	Along-scan field coordinate in the SRS (=Scanning Reference System)
z	Across-scan field coordinate in the SRS (Sun > 0)
f	Index of field of view. 1: preceding, 2: following
k	Pixel index along-scan in data pixel space (integer)
m	Pixel index across-scan (integer)
n	CCD index to number each CCD from 1 to n_{max}
κ	Pixel coordinate along-scan (real)
μ	Pixel coordinate across-scan (real)
$[\mathbf{xyz}]$	Triad directions of the SRS
$[\mathbf{XYZ}]$	Triad directions of the BCRS
$\mathbf{q}$	Attitude unit quaternion vector of components $q_1,\ldots,q_4$
$P(u,v)$	Point Spread Function (PSF). Function of continuous pixel coordinates
$L(u)$	Line Spread Function (LSF). Function of the continuous AL pixel coordinate
G	Magnitude in the G-band
σ_x	RMS, standard error or standard deviation of x
$\mathbf{s}$	Observed proper direction of a source from Gaia
$\mathbf{n}$	BCRS coordinate direction of the light ray from Gaia
σ	BCRS coordinate direction of the light ray at infinity
$\mathbf{k}$	Unit coordinate vector from the source to the observer
$\mathbf{l}$	Coordinate direction of the source in the BCRS

J List of acronyms

The following table has been generated from the on-line Gaia acronym list:

Acronym	Description
2MASS	Two-Micron All Sky Survey
AC	ACross scan (direction)
ACS	Attitude Control Sub-system
ADC	Analogue to Digital Converter (CCD)
AF	Astrometric Field (in Astro)
AGB	Asymptotic Giant Branch (stars)
AGIS	Astrometric Global Iterative Solution
AGN	Active Galactic Nucleus
AIX	Advanced Interactive eXecutive (proprietary operating system developed by IBM based on UNIX System V)
AL	ALong scan (direction)
ANN	Artificial Neural Network
ANSI	American National Standards Institute
AO	Announcement of Opportunity
AOCS	Attitude and Orbit Control Sub-system
AP	Astrophysical Parameters
ARI	Astronomisches Rechen-Institut (Heidelberg; part of ZAH)
ASAP	As Soon As Possible
ASAS	All-Sky Automated Survey
ASCII	American Standard Code for Information Interchange
ASM	Astrometric Sky Mapper (obsolete)
ASP	Alcatel SPace
AU	Astronomical Unit
AVU	Astrometric Verification Unit
BA	Basic Angle
BAM	Basic Angle Monitoring (Device)
BC	Best Case
BCRS	Barycentric Celestial Reference System
BGM	Besançon Galaxy Model
BLOB	Binary Large OBject
BP	Blue Photometer
BSC	Barcelona Supercomputing Centre
C1B	First common baseline (C1) for BBP photometric system, based on F4B photometric system (UB-PWG-028)

C1M	First common baseline (C1) for MBP photometric system, based on S5M photometric system (UB-PWG-029)
CC	Change Control
CCB	Configuration Control Board
CCD	Charge-Coupled Device
CD	Compact Disc
CDROM	Compact Disc Read-Only Memory (also known as CD-ROM)
CESCA	CEntre de Supercomputacío de CAtalunya (involved in GDAAS)
CET	Central European Time
CM	Configuration Management
CN	Change Notice
CNES	Centre National d'Etudes Spatiales (France)
CPU	Central Processing Unit
CTE	Charge Transfer Efficiency
CTI	Charge Transfer Inefficiency
CU	Coordination Unit (in DPAC)
CoMRS	Centre of Mass Reference System
DACC	Data Analysis Coordination Committee
DAL	Data Access Layer
DB	Decibel
DBMS	DataBase Management System
DDS	Data Distribution System
DFL	Detailed First Look
DIB	Diffuse Interstellar Band
DIVA	Deutsches Interferometer für Vielkanalphotometrie und Astrometrie (cancelled)
DLR	Deutsches Zentrum für Luft und Raumfahrt
DP	Data Processing
DPAC	Data Processing and Analysis Consortium
DPACE	Data Processing and Analysis Consortium Executive
DPC	Data Processing Centre
DSC	Discrete Source Classifier
DU	Development Unit (in DPAC)
DVD	Digital Video (Versatile) Disc
EC	Economic Conditions
ECSS	European Cooperation for Space Standardisation
EMC	Electro-Magnetic Compatibility
EO	Extended Object
ESA	European Space Agency
ESAC	European Space Astronomy Centre (formerly known as VilSpa)
ESO	European Southern Observatory

ESOC	European Space Operations Centre (ESA)
ESP	Extra-Solar Planet
ESTEC	European Space research and TEchnology Centre (ESA)
FAME	Fizeau (Fast) Astrometric Mapping Explorer (cancelled)
FAST	Fundamental Astronomy by Space Techniques (Hipparcos)
FCT	Flight Control Team
FEM	Finite-Element Model
FFT	Fast Fourier Transform
FIR	Finite Impulse Response (Filter)
FITS	Flexible Image Transport System (data format in Parameter Database)
FK5	Fifth Fundamental Catalogue
FL	First Look
FLAME	Final Luminosity And Mass Estimator
FLOP	FLoating-point OPeration
FM	Flat Mirror
FNRS	Fonds National de la Recherche Scientifique (Belgium)
FOV	Field of View (also denoted FOV)
FPA	Focal Plane Assembly (Focal Plane Array)
FPRS	Focal Plane Reference System
FTE	Full-Time Equivalent
FoV	Field of View (also denoted FOV)
FoVRS	Field-of-View Reference System
GAIA	Global Astrometric Interferometer for Astrophysics (obsolete; now spelled as Gaia)
GALEX	GALaxy Evolution eXplorer
GASS	GAia System-level Simulator
GAia	Global Astrometric Interferometer for Astrophysics (obsolete; now spelled as Gaia)
GB	GigaByte
GCRF	Gaia Celestial Reference Frame
GCRS	Geocentric Celestial Reference System
GDAAS	Gaia Data Access and Analysis Study
GIBIS	Gaia Instrument and Basic Image Simulator
GIS	(Astrometric) Global Iterative Solution
GOG	Gaia Object Generator
GR	General Relativity
GREM	Gaia Relativity Model
GSC	Guide Star Catalogue
GSC-II	Guide Star Catalogue version 2 (also denoted as GSC-2)
GST	Gaia Science Team

GTS	Gaia Transfer System
GUI	Graphical User Interface
GWP	Gaia Work Package
HK	Housekeeping (also denoted H/K)
HPC	High-Performance Computing
HR	High Resolution
HST	Hubble Space Telescope
HTM	Hierarchical Triangular Mesh
HW	Hardware (also denoted H/W)
I/O	Input/Output
IAU	International Astronomical Union
ICAP	Identification, Classification, and Astrophysical Parametrisation
ICD	Interface Control Document
ICRS	International Celestial Reference System
ID	Identifier (Identification)
IDL	Interactive Data Language
IDT	Initial Data Treatment
IDU	Intermediate Data Update
IEEC	Institut d'Estudis Espacials de Catalunya (involved in GDAAS)
IMCCE	Institut de Mécanique Céleste et de Calcul des Ephémérides
IMF	Initial Mass Function
INAF	Instituto Nazionale di Astrofisica (Italy)
INPOP	Intégrateur Numérique Planétaire de l'Observatoire de Paris
INTEGRAL	INTErnational Gamma-Ray Astrophysics Laboratory (ESA)
IRAF	Image Reduction and Analysis Facility (NOAO)
IRD	Interface Requirements Document
ISDC	INTEGRAL Science Data Centre
IT	Information Technology
ITT	Invitation To Tender
IVOA	International Virtual Observatory Alliance
IoA	Institute of Astronomy (Cambridge)
JPL	Jet Propulsion Laboratory (DE405 ephemeris)
LAN	Local Area Network
LEI	Leiden Observatory
LOS	Loss of Signal
LSF	Line-Spread Function
LSST	Large-aperture Synoptic Survey Telescope
M-L	Mass Luminosity relationship.
MATISSE	Next-generation data management products that offer native object storage and support for SQL
MB	MegaByte

MBA	Main-Belt Asteroid
MCS	Mission Control System
MDB	Main Database
MF	Master Function
MM	Mission Manager
MOC	Mission Operations Centre
MRD	Mission Requirements Document
MS	MicroSoft (software company)
MSL	Modified Scanning Law (Gaia-LL-058)
MSSL	Mullard Space Science Laboratories (involved in RVS)
MTF	Modulation Transfer Function (CCD)
NASA	National Aeronautics and Space Administration (USA)
NDAC	Northern Data Analysis Consortium (Hipparcos)
NEO	Near-Earth Object
NSL	Nominal Scanning Law
NSS	Non-Single Star
NVO	Natioanl Virtual Observatory (United States)
OATO	Torino Observatory
OB	On-Board
ODAS	One-Day Astrometric Solution
ODC	One-Day Calibration
OF	Object Feature
OGLE	Optical Gravitational Lensing Experiment
OM	Object Matching
OPM	Observatoire de Paris-Meudon
PB	Play-Back
PC	Personal Computer
PCA	Principal Component Analysis
PDF	Probability Density Function
PDHU	Payload Data Handling Unit
PEGASE	Projet d’Etude des Galaxies par Synthèse Evolutive
PI	Principal Investigator
PLM	Payload Module
PM	Project Manager
PPARC	Particle Physics and Astronomy Research Council
PPM	parts per million
PPN	Parametrised Post-Newtonian (formalism in General Relativity)
PRNU	Photo-Response Non-Uniformity
PS	Photometric System
PSD	Power Spectral Density
PSF	Point-Spread Function

PhD	Doctorate in Philosophy
QA	Quality Assurance
QC	Quality Control
QE	Quantum Efficiency (CCD)
QSO	Quasi-Stellar Object
RAL	Rutherford Appleton Laboratory
RAM	Random Access Memory
RAMOD	Relativistic Astrometric MODel
RB	Rubidium
REMAT	RElativistic Models And Tests (DU in CU3)
RGB	Red Giant Branch Branch (stars)
RGC	Reference Great Circle
RMS	Root-Mean-Square
ROEMER	Proposal for the Third Medium-Size ESA Mission (M3; not approved)
RON	Read-Out Noise (CCD)
RP	Red Photometer
RRFWG	Relativity and Reference Frame Working Group (obsolete)
RS	Reference System
RSSD	Research and Scientific Support Department (ESA)
RTN	Research Training Network (EU)
RV	Radial Velocity
RVS	Radial Velocity Spectrometer
SAG	Science Advisory Group (obsolete; superseded by GST)
SAN	Storage Area Network
SB	Spectroscopic Binary
SCRS	Spacecraft Reference System
SDPC	Spectroscopic Data Processing Centre
SDSS	Sloan Digital Sky Survey
SED	Spectral Energy Distribution
SEF	Standard Exchange Format
SEM	Space Environment Monitor (GOES)
SGIS	Spectroscopic Global Iterative Solution
SIM	Space Interferometry Mission
SIP	Science Implementation Plan
SK	Spare Kit
SM	Sky Mapper
SMP	Science Managment Plan
SNR	Signal-to-Noise Ratio (also denoted S/N)
SOC	Science Operations Centre
SPB	Senior Procurement Board

SPSS	Spectro-Photometric Standard Stars
SQMP	System Quality Management Plan
SRS	Scanning Reference System
SSMM	Solid State Mass Memory
SSO	Solar System Object
STAF	File Archive and Transfer Service (implemented at CNES)
STUFF	Computer code to generate mock catalogues of astronomical sources (E. Bertin)
SVM	SerVice Module
SVP	System Verification Plan
SVT	System Validation Test
SW	Software (also denoted S/W)
SdB	Satellite DataBase
TB	Tera Byte
TBD	To Be Defined (Determined)
TCB	Barycentric Coordinate Time
TDI	Time-Delayed Integration (CCD)
TM	Telemetry
TN	Technical Note
TT	Terrestrial Time
TWP	Top-level Work Package
UB	University of Barcelona
UCAC	USNO CCD Astrograph Catalogue
UK	United Kingdom
ULB	Université Libre de Bruxelles
UML	Unified Modeling Language
UPC	Universitat Politècnica de Catalunya
URD	User Requirements Document
US	United States
USA	United States of America
UV	Ultra Violet
VIM	Variability-Induced Mover (DMSA)
VISTA	Visible and Infrared Survey Telescope for Astronomy
VLTI	Very Large Telescope Interferometer (ESO)
VO	Virtual Object
VPU	Video Processing Unit
VSOP	Variations Séculaires de l'Orbite des Planètes
WBS	Work Breakdown Structure
WD	WinDow
WFE	WaveFront Error
WFS	WaveFront Sensor

WG	Working Group
WP	Work Package
WR	Window in RVS (covering a star spectrum)
XEUS	X-Ray Evolving Universe Spectrometer
XMM	X-ray Multi-mirror Mission (ESA; officially known as XMM-Newton)
ZAH	Zentrum fuer Astronomie Heidelberg (Centre for Astronomy, University of Heidelberg)

K References

[AH98] F. Arenou and M. Haywood. Simulations of perspective acceleration for gaia. Technical Report IWG-OPM-001, Observatory of Paris, 1998.

[ALtL05] S. G. Ansari, U. Lammers, and M. ter Linden. Gaiagrid : Its implications and implementation. In *Proc. Astronomical Data Analysis Software and Systems XIV*, volume 347, pages 429–. Astronomical Society of the Pacific, 2005.

[Ang05] F. De Angeli. The Gaia Software Toolbox - User guide. Technical report, IoA, 2005. `http://www.rssd.esa.int/SA-general/Projects/GAIA/wiki/index.php?title=CU1:_GaiaTools`.

[Bab06] C. Babusiaux. Gaia instrument and basic image simulator (gibis) overview. Technical report, October 2006. GAIA-C2-TN-OPM-CB-007-1.

[Bab07] C. Babusiaux. Requests for spacecraft and instrument specifications. Technical report, March 2007. GAIA-CD-TN-OPM-CB-010-2.

[BAC05] C. Babusiaux, F. Arenou, and F. Chéreau. New implementation of a star system in the gaia simulator. Technical report, November 2005. GAIA-C2-TN-OPM-CB-006.

[Bas06] U. Bastian. Extended Geometric Calibration Model for Gaia's Astro Instrument. Technical report, ARI, Heidelberg, October 2006. GAIA-ARI-BAS-011-3.

[BB02a] S. J. Bus and R. P. Binzel. Phase ii of the small main-belt asteroid spectroscopic survey: A feature-based taxonomy. *Icarus*, 158:146–177, 2002.

[BB02b] S. J. Bus and R. P. Binzel. Phase ii of the small main-belt asteroid spectroscopic survey: The observations. *Icarus*, 158:106–145, 2002.

[BBF$^+$06] M. Bellazzini, A. Bragaglia, L. Federici, E. Diolaiti, C. Cacciari, and E. Pancino. Absolute calibration of gaia photometric data. i. general considerations and requirements. Technical report, May 2006. GAIA-C5-TN-OABO-MBZ-001-1.

[BCd03] D. Bini, M. T. Crosta, and F. de Felice. Orbiting frames and satellite attitudes in relativistic astrometry. *Classical and Quantum Gravity*, 20:4695–4706, 2003.

[BCS06] C. Babusiaux, F. Chéreau, and P. Sartoretti. GIBIS User Guide. Technical report, July 2006. GAIA-C2-SP-OPM-CB-003-3, `http://gibispc.obspm.fr/~gibis/help/index.html`.

[Bec99] K. Beck. *Extreme Programming Explained: Embrace Change*. Addison-Wesley, 1st edition, 1999.

[B*et al*06] M. Biermann and *et al*. Proposal for an on-board data priority scheme. Technical report, July 2006. GAIA-CO-TN-ZAH-MB-13-1.

[BGG+06] D. Busonero, M. Gai, D. Gardiol, M. G. Lattanzi, and D. Loreggia. Chromaticity in all-reflective telescopes for astrometry. *A&A*, 449:827–836, April 2006.

[BGH+93] U. Bastian, G. Gilmore, J.L. Halbwachs, E. Høg, J. Knude, J. Kovalevsky, A. Labeyrie, F. van Leeuwen, L. Lindegren, J. W. Pel, H. Schrijver, R. Stabell, and P. Thejll. Roemer. Technical report, Lund Observatory, July 1993. Proposal for a Third Medium Size ESA Mission (M3), Lund 1993.

[BJ02] C. A. L. Bailer-Jones. Determination of stellar parameters with gaia. *ApSS*, 280:21–29, 2002. `http://www.mpia-hd.mpg.de/homes/calj/vilnius.html`.

[BJ03] C. A. L. Bailer-Jones. On the classification and parametrization of gaia data using pattern recognition methods. In U. Munari, editor, *Gaia Spectroscopy Science and Technology, ASP Conference Series*, volume 298, pages 199–208, 2003. `http://www.mpia-hd.mpg.de/homes/calj/monte_rosa_2002.html`.

[BJ04a] C. A. L. Bailer-Jones. Evolutionary design of photometric systems and its application to gaia. *A&A*, 419:385–403, 2004. `http://www.mpia-hd.mpg.de/homes/calj/hfd.html`.

[BJ04b] C. A. L. Bailer-Jones. Hfd filter system design for gaia. design of the h2b, h3b, h2m and h3m systems and general insights concerning mbp. Technical report, MPIA, June 2004. ICAP-CBJ-016.

[BJ04c] C. A. L. Bailer-Jones. Implementation proposal for the hfd-b1 and hfd-m1 photometric systems in the gaia design adopted at gst10. Technical report, MPIA, February 2004. ICAP-CBJ-015.

[BJ05] C. A. L. Bailer-Jones. Object classification and the determination of stellar parameters. In C. Turon, K.S. O'Flaherty, and M.A.C. Perryman, editors, *Proc. of The Three Dimensional Universe with Gaia*, pages 393–399, 2005. `http://www.mpia-hd.mpg.de/homes/calj/gaia_paris_2004_icap.pdf`.

[Bre82] P. Bretagnon. Thorie du mouvement de l'ensemble des plantes. Solution VSOP82. *Astronomy & Astrophysics*, 114:278, 1982.

[Bro03] A. G. A. Brown. Results of the second cycle of blind testing. Technical report, LEI, December 2003. ICAP-AB-004.

[Bro05] A. G. A. Brown. Gaia photometric data analysis. In Turon et al. [TOP05], pages 377–384.

[Bro06a] A. G. A. Brown. Photometry with dispersed images - overview of bp/rp data processing. Technical report, LEI, July 2006. GAIA-C5-TN-LEI-AB-009.

[Bro06b] A. G. A. Brown. Simulating prism spectra for the eads-astrium gaia design. Technical report, LEI, June 2006. GAIA-CA-TN-LEI-AB-005.

[Bro07] A. G. A. Brown. Calibration requests to industry. Technical report, LEI, Feb 2007. GAIA-CG-TN-LEI-AB-007-2.

[BTL93] B. Bucciarelli, L. G. Taff, and M. G. Lattanzi. A generalization of the moving mean. *J. Statist. Comput. Simul.*, 48:29, 1993.

[BWW04] D. R. Brett, R. G. West, and P. J. Wheatley. The automated classification of astronomical light curves using Kohonen self-organizing maps. *MNRAS*, 353:369–376, September 2004.

[CdFB$^+$04] M.T. Crosta, F. de Felice, B. Bucciarelli, M.G. Lattanzi, and A. Vecchiato. Proper frames and time scan for gaia-like satellites. Technical report, 2004. Technical note on Livelink.

[CJF$^+$06] J.M. Carrasco, C. Jordi, F. Figueras, G. Anglada-Escudé, and E.B. Amores. Towards the selection of standard stars for absolute flux calibration. signal-to-noise ratio for bp/rp spectra and crowding due to fov overlapping. Technical report, September 2006. GAIA-C5-TN-UB-JMC-001-2.

[CLM$^+$06] S. Casertano, M. G. Lattanzi, R. Morbidelli, R. Pannunzio, A. Spagna, D. Pourbaix, and A. Sozzetti. Double-blind tests program for astrometric planet detection with Gaia. Final report: Results for test T3. Technical report, Osservatorio Astronomico di Torino, April 2006. GAIA-C4-TN-OAT-SC-001.

[CM06] M. T. Crosta and F. Mignard. Microarcsecond light bending by jupiter. *Classical and Quantum Gravity*, 23:4853–4871, 2006.

[Cro03] M. T. Crosta. Methods of Relativistic Astrometry for the analysis of astrometric data in the Solar System gravitational field. *Ph.D. Thesis*, 2003.

[CSVS06] J.-F. Claeskens, A. Smette, L. Vandenbulcke, and J. Surdej. Identification and redshift determination of quasi stellar objects with medium-band photometry: application to Gaia. *MNRAS*, 367:879–904, 2006.

[DBL$^+$03] R. Drimmel, B. Bucciarelli, M. G. Lattanzi, R. Smart, and A. Spagma. All sky source counts to $R_F = 20$ using GSC-II. Technical report, April 2003. GAIA_ML_20.

[dBLSea06] J. de Bruijne, L. Lindegren, O. Svensson, and et al. Chromaticity in Gaia. Technical report, ESTEC, August 2006. GAIA-CA-TN-ESA-JDB-028-1.

[dBLV01] F. de Felice, B. Bucciarelli, M. G. Lattanzi, and A. Vecchiato. General relativistic satellite astrometry. II. Modeling parallax and proper motion. *A&A*, 373:336–344, 2001.

[dFCV$^+$04] F. de Felice, M. T. Crosta, A. Vecchiato, M. G. Lattanzi, and B. Bucciarelli. A general relativistic model of light propagation in the gravitational field of the solar system: the static case. *ApJ*, 607:580–595, 2004.

[Div03] ESA Publications Division. Project Phasing and Planning. Technical report, June 2003. ECSS-M-30B.

[dP06] F. de Felice and G. Preti. Ray tracing in relativistic astrometry: stellar positions, stellar motion and error budget. *Classical and Quantum Gravity*, 23:5467–5476, 2006.

[Dri02] R. Drimmel. Galactic absorption model: refined version. Technical report, August 2002. GAIA_SWG_RD_03.

[DS01] R. Drimmel and D. Spergel. *Astrophys. J.*, 556:181, 2001.

[dVC$^+$06] F. de Felice, A. Vecchiato, M. T. Crosta, M. G. Lattanzi, and B. Bucciarelli. A general relativistic model for the light propagation in the gravitational field of the Solar System: the dynamical case. *ApJ*, 653:1552, 2006.

[EA06] EADS-Astrium. Algorithms Performances Specification. Technical report, September 2006. GAIA.ASF.SP.PLM.00064, Issue 3, revision 0, 4.

[EB02] L. Eyer and C. Blake. Automated Classification of Variable Stars for ASAS Data. In C. Aerts, T. R. Bedding, and J. Christensen-Dalsgaard, editors, *ASP Conf. Ser. 259: IAU Colloq. 185: Radial and Nonradial Pulsationsn as Probes of Stellar Physics*, pages 160–+, 2002.

[EB05a] N. W. Evans and V. Belokurov. A Prototype for Science Alerts. In C. Turon, K. S. O'Flaherty, and M. A. C. Perryman, editors, *ESA SP-576: The Three-Dimensional Universe with Gaia*, pages 385–+, January 2005.

[EB05b] L. Eyer and C. Blake. Automated classification of variable stars for All-Sky Automated Survey 1-2 data. *MNRAS*, 358:30–38, March 2005.

[EC00] L. Eyer and J. Cuypers. Predictions on the Number of Variable Stars for the GAIA Space Mission and for Surveys such as the Ground-Based International Liquid Mirror Telescope. In L. Szabados and D. Kurtz, editors, *ASP Conf. Ser. 203: IAU Colloq. 176: The Impact of Large-Scale Surveys on Pulsating Star Research*, pages 71–72, 2000.

[EFN^{+}97] ESA, FAST, NDAC, TDAC, INCA, Matra Marconi Space, and Alenia Spazio. *The Hipparcos and Tycho Catalogues*. 1997. SP-1200.

[EG97] L. Eyer and M. Grenon. Photometric Variability in the HR Diagram. In *ESA SP-402: Hipparcos - Venice '97*, pages 467–472, 1997.

[EM05] L. Eyer and F. Mignard. Rate of correct detection of periodic signal with the Gaia satellite. *MNRAS*, 361:1136–1144, August 2005.

[EM06] C. Babusiaux E. Masana, X. Luri. Proposal for the definition of the gaia object generator (gog). Technical report, October 2006. GAIA-C2-SP-UB-EM-001-1.

[ERER05] L. Eyer, A. Robin, D. W. Evans, and C. Reylé. Implementation of variable stars in the galactic model. i. general concept. Technical report, June 2005. VSWG-LE-002.

[Exe06] ESA Executive. Announcement of Opportunity for the Gaia Data Processing. Technical report, November 2006. Gaia-AO.

[Eye98] L. Eyer. Les étoiles variables de la mission HIPPARCOS. *Ph.D. Thesis*, 1998.

[Eye02] L. Eyer. Search for QSO Candidates in OGLE-II Data. *Acta Astron.*, 52:241–262, 2002.

[Eye05] L. Eyer. Variability Analysis: Detection and Classification. In C. Turon, K. S. O'Flaherty, and M. A. C. Perryman, editors, *ESA SP-576: The Three-Dimensional Universe with Gaia*, pages 513–+, January 2005.

[Eye06a] L. Eyer. Astronomical Databases, Space Photometry and Time Series Analysis: Open Questions. In C. Sterken and C. Aerts, editors, *Astronomical Society of the Pacific Conference Series*, pages 15–+, April 2006.

[Eye06b] L. Eyer. The Gaia mission . Pulsating stars with Gaia. *Memorie della Societa Astronomica Italiana*, 77:549–+, 2006.

[FBD+06] L. Federici, A. Bragaglia, E. Diolaiti, M. Bellazzini, C. Cacciari, E. Pancino, G. Altavilla, P. Montegriffo, E. Rossetti, C. Jordi, F. Figueras, and S. Trager. Absolute calibration of gaia photometric data. ii. observing facilities for ground-based support (pilot program). Technical report, October 2006. GAIA-C5-TN-OABO-LF-001-1.

[FBJL+07] P. Re Fiorentin, C.A.L. Bailer-Jones, Y.S. Lee, T.C. Beers, and T. Sivarani. Estimating stellar atmospheric parameters from sdss/segue spectra. *A&A*, in press, 2007.

[FLS+05] A. Fienga, J. Laskar, J. L. Simon, H. Manche, and M. Gastineau. Imcce planetary ephemerides: Present and future. In C. Turon, K.S. O'Flaherty, and M.A.C. Perryman, editors, *Proc. of The Three Dimensional Universe with Gaia*, pages 293–296, 2005.

[gai] Gaia Parameter Database at `http://www.rssd.esa.int/\discretionary-Gaia/\discretionary-paramdb`.

[gai00] GAIA — Composition, Formation and Evolution of the Galaxy. Technical report, July 2000. Concept and Technology Study Report, ESA-SCI(2000)4.

[Gar06] D. Gardiol. Bright stars – considerations on the use of delayed gate activation. Technical report, 2006. GAIA-CH-TN-INAF-DG-001.

[GB04] M. Gai and D. Busonero. On-ground and in-orbit calibration working group – framework and activity plan for 2004. Technical report, 2004. Technical note on Livelink (CWG–003).

[GBGL05a] M. Gai, D. Busonero, D. Gardiol, and D. Loreggia. The Gaia focal plane to sky mapping: a sample of calibration issues. In Turon et al. [TOP05], page 433.

[GBGL05b] D. Gardiol, D. Busonero, M. Gai, and M. G. Lattanzi. Gaia performance on bright stars. Technical report, 2005. Technical note on Livelink (GAIA–ML–025).

[GGU+97] A. E. Gomez, S. Grenier, S. Udry, M. Haywood, L. Meillon, V. Sabas, S. Sellier, and D. Morin. Kinematics of disk stars in the solar neighbourhood. In *ESA SP-402: Hipparcos - Venice '97*, pages 621–624, 1997.

[GLM+04] D. Gardiol, D. Loreggia, S. Mannu, S. Mottini, L. Perachino, M. Gai, and M. G. Lattanzi. End-to-end optomechanical simulation for high-precision global astrometry. In *SPIE*, volume 5497, pages 461–470, 2004.

[GST^{+}02] L. M. González, I. Serraller, J. Torra, X. Luri, F. Figueras, C. Jordi, E. Masana, J. Cambras, and M. Mestres. GDAAS — Final Report. Technical report, May 2002. GMV-GDAAS-RP-001.

[GST^{+}03] J. Gray, A. S. Szalay, A. Thakar, P. Kunszt, C. Stoughton, D. Slutz, and J. Vandenberg. Data Mining in the SDSS SkyServer Database. *Distributed Data and Structures 4: Records of the 4th International Meeting, W. Litwin, G.Levy (eds), Paris France March 2002*, 2003.

[HdB06] E. Høg and J. de Bruijne. Gaia Sampling Scheme. Technical report, May 2006. GAIA-CA-TN-NBI-EH-176-2.

[Hec06] M. Hechler. Gaia Consolidated Report on Mission Analysis (CReMA). Technical report, ESA, European Space Operations Cenre, March 2006. GAIA-ESC-RP-0001, Issue 2.0.

[Her05] J. Hernandez. ICD for Gaia Main Database. Technical report, December 2005. GAIA-C1-SP-ESAC-JH-001.

[Hoa07] John Hoar. Gaia MOC SOC interface requirements document. Technical report, Jan 2007. GAIA-C1-SP-ESAC-JSH-008-1.

[Høg94] E. Høg. A new era of global astrometry and photometry from space and ground. Technical report, CUO, February 1994. Contribution at the G. Colombo Memorial Conference : Ideas for Space Research after the Year 2000.

[HOHL06] J. Hernandez, W. O'Mullane, C. Huc, and T. Levoir. Language recommendation for DPAC. Technical report, January 2006. GAIA-C1-TN-ESAC-JH-002.

[HRC97] M. Haywood, A. C. Robin, and M. Crézé. The evolution of the milky way disc. i. vertical structure and local constraints. *Astronomy & Astrophysics*, 320:428, 1997.

[IAU01] IAU. *Information Bulletin*, 88, 2001. (errata in IAU Information Bulletin, 89).

[JF06] C. Jordi and F. Figueras. Calculation of effective wavelength from BP/RP data for chromaticity calibration. Technical report, Univ. Barcelona, August 2006. GAIA-CA-TN-UB-CJ-040-1.

[JFC^{+}02] C. Jordi, F. Figueras, J. M. Carrasco, G. Taut, and F. Castelli. Gaia photometry: [α/fe] abundance variations. Technical report, June 2002. UB-PWG-008.

[JHBdB06] C. Jordi, E. Høg, A. G. A. Brown, and J. de Bruijne. Gaia spectrophometers: optimization study. Technical report, April 2006. GAIA-CH-TN-UB-CJ-037.

[JHBea06] C. Jordi, E. Høg, A. G. A. Brown, and et al. The design and performance of the gaia photometric system. *MNRAS*, 367:290–314, 2006.

[KB03] S. A. Klioner and R. Blankenburg. Technical report on the implementation of the Gaia relativistic model, available from the gaia document archive. Technical report, 2003. Technical note on Livelink.

[KLF$^+$95] J. Kovalevsky, L. Lindegren, M. Froeschle, F. van Leeuwen, M. A. C. Perryman, J. L. Falin, F. Mignard, M. J. Penston, C. S. Petersen, P. L. Bernacca, B. Bucciarelli, F. Donati, R. Hering, E. Hog, M. G. Lattanzi, H. van der Marel, H. Schrijver, and H. G. Walter. Construction of the intermediate HIPPARCOS astrometric catalogue. *A&A*, 304:34–+, 1995.

[Kli03a] S. Klioner. Requirements on the accuracy of Gaia orbit determination from the relativistic light deflection effects. Technical report, Lohrmann Observatory, Dresden, April 2003. Technical note on Livelink.

[Kli03b] S. A. Klioner. A practical relativistic model for microarcsecond astrometry in space. *AJ*, 125:1580–1597, 2003.

[Kli03c] S. A. Klioner. Proposal for the representation of the astrometric parameters. Technical report, 2003. Technical note on Livelink.

[Kli03d] S. A. Klioner. Technical report on the implementation of the Gaia relativistic model, amendment for version 1.0g. Technical report, 2003. Technical note on Livelink.

[Kli04] S. A. Klioner. Physically adequate proper reference system of a test observer and relativistic description of the Gaia attitude. *Phys.Rev.D*, 69:124001, 2004.

[Kli05] S. A. Klioner. Relativistic perturbations for lissajous orbits around l_2. Technical report, 2005. Technical note on Livelink.

[Kov98] J. Kovalevsky. First results from hipparcos. *ARA&A*, 36:99–130, 1998.

[KP03] S. A. Klioner and M. Peip. Numerical simulations of the light propagation in the gravitational field of moving bodies. *A&A*, 410:1063–1074, 2003.

[KS00] S. A. Klioner and M. Soffel. Relativistic celestial mechanics with PPN parameters. *Phys.Rev.D*, 62:024019, 2000.

[KV04] S. Kopeikin and I. Vlasov. Parametrized post-Newtonian theory of reference frames, multipolar expansions and equations of motion in the N-body problem. *Phys. Rep.*, 400:209–318, 2004.

[Lam03] U. Lammers. The Gaia parameter database – Technical User Manual. Technical report, ESA, August 2003. GAIA-UL-001.

[Lam04] U. Lammers. Gaia Telemetry Rate Simulations: A first look at the complete picture. Technical report, ESA/ESTEC, July 2004. GAIA-UL-008.

[Lam05] U. Lammers. Gaia Telemetry Rate Simulations: A results update. Technical report, ESA/ESTEC, August 2005. GAIA-UL-014.

[Lam06a] U. Lammers. First AGIS Results with 1 Million star/5 years simulated data. Technical report, April 2006. GAIA-C3-PR-ESAC-UL-018.

[Lam06b] Uwe Lammers. First AGIS results with 1 million star/5 years simulated data. Technical report, April 2006. GAIA-C1-PR-ESAC-UL-018-1.

[LB03] C. Larman and V. R. Basili. Iterative and Incremental Development: A Brief History. *IEEE Computer*, June 2003. `http://www2.umassd.edu/SWPI/xp/articles/r6047.pdf`.

[LCB97] Th. Lejeune, F. Cuisinier, and R. Buser. A standard stellar library for evolutionary synthesis. i. calibration of theoretical spectra. *A&ASS*, 125:229–246, 1997.

[LCJ$^+$05] M. G. Lattanzi, S. Casertano, S. Jancart, R. Morbidelli, D. Pourbaix, R. Pannunzio, A. Sozzetti, and A. Spagna. Detection and Characterization of Extra-Solar Planets with Gaia. In Turon et al. [TOP05], pages 251–258.

[LHHO06] U. Lammers, J. Hernandez, J. Hoar, and W. O'Mullane. ESAC Implementation of the Astrometic Global Iterative Solution. Technical report, January 2006. GAIA-C3-TN-ESAC-UL-015.

[Lin98] L. Lindegren. The scanning law for GAIA. Technical report, February 1998. SAG-LL-14.

[Lin00] L. Lindegren. Centroiding on GAIA CCD sample data. Technical report, October 2000. SAG-LL-32.

[Lin01] L. Lindegren. Proposed prototype processes for the Gaia Global Iterative Solution. Technical report, Lund Observatory, March 2001. GAIA-LL-34.

[Lin05a] L. Lindegren. A theoretical investigation of chromaticity. Technical report, Lund Observatory, November 2005. GAIA-C3-TN-LU-LL-064-1.

[Lin05b] L. Lindegren. Local plane coordinates for the detailed analysis of complex gaia sources. Technical report, Lund Observatory, October 2005. GAIA-LL-061.

[LJMD+07] T. Levoir, F. Jocteur-Monrozier, J.-C. Damery, W. O'Mullane, and J. Hoar. Product Assurance Plan for DPAC. Technical report, March 2007. GAIA-C1-SP-CNES-TL-001.

[LP94] L. Lindegren and M. A. C. Perryman. Gaia: Global astrometric interferometer for astrophysics. Technical report, Lund Observatory, September 1994. Supplementary Information Submitted to the Horizon2000+ Survey Committee.

[LPB+93] L. Lindegren, M. A. C. Perryman, U. Bastian, J. C. Dainty, E. Høg, F. van Leeuwen, J. Kovalevsky, A. Labeyrie, F. Mignard, J. E. Noordam, R. S. Le Poole, P. Thejll, and F. Vakili. Gaia: Global astrometric interferometer for astrophysics. Technical report, Lund Observatory, October 1993. Response to Call for Mission Concepts for Horizon 2000 Follow UP: Proposal for an astrometric interferometer as an ESA Cornerstone Mission.

[LR06] Uwe Lammers and Frederic Raison. Results from CY01-C1 testing campaign for the astrometric global iterative solution. Technical report, November 2006. GAIA-C3-PR-ESAC-UL-023-1.

[LSSC00] M. G. Lattanzi, A. Spagna, A. Sozzetti, and S. Casertano. Space-borne global astrometric surveys: the hunt for extrasolar planets. *MNRAS*, 317:211–224, 2000.

[MBJ05] F. Mignard and C. Bailer-Jones. A Model for the Gaia Data Processing Consortium. Technical report, July 2005. DACC-CHAIR-002.

[McD04] J. McDowell. Data models for the vo. *Toward an International Virtual Observatory: Proceedings of the ESO/ESA/NASA/NSF Conference Held at Garching, Germany, 10-14 June 2002, ESO ASTROPHYSICS SYMPOSIA. ISBN 3-540-21001-6*, 2004.

[Mig01a] F. Mignard. Detection of near earth objects with gaia. Technical report, Observatoire de la Côte d'Azur/CERGA, February 2001. SAG_FM_007.

[Mig01b] F. Mignard. Phase effect in the detection of neos with gaia. Technical report, Observatoire de la Côte d'Azur/CERGA, February 2001. SAG_FM_008.

[Mig03a] F. Mignard. Prototype of the ephemeris of the earth for Gaia. Technical report, Observatoire de la Côte D'Azur, February 2003. GAIA_FM_013.

[Mig03b] F. Mignard. Prototype of the ephemeris of the Gaia spacecraft. Technical report, Observatoire de la Côte D'Azur, March 2003. GAIA_FM_014.

[Mig05a] F. Mignard. Ephemeris requirements for Gaia. Technical report, Observatoire de la Côte D'Azur, September 2005. GAIA_FM_023.

[Mig05b] F. Mignard. Overall science goals of the gaia mission. In C. Turon, K.S. O'Flaherty, and M.A.C. Perryman, editors, *Proc. of The Three Dimensional Universe with Gaia*, pages 5–14, 2005.

[Mona] F. Jocteur Monrozier. Gaia cnes technological studies report. Technical report, Centre National d'Etudes Spatiales. GAIA-NT-31000-46-CN.

[Monb] F. Jocteur Monrozier. Java application profiling. Technical report, Centre National d'Etudes Spatiales. GAIA-NT-25000-0059-CN.

[Monc] F. Jocteur Monrozier. Optimisation form. Technical report, Centre National d'Etudes Spatiales. GAIA-DP-38300-21-CN.

[Mon03] M. Moniez. Does transparent hidden matter generate optical scintillation? *A&A*, 412:105–120, December 2003.

[Moo65] Gordon E. Moore. Cramming more components onto integrated circuits. *Electronics Magazine*, 1965.

[Mun99] U. Munari. High resolution spectroscopy over lambda lambda 8500-8750 for gaia. i. mapping the mkk classification system. *A&AS*, 137:521–528, 1999.

[Mun00] U. Munari. The diffuse interstellar band at 8620 : A good reddening tracer for gaia. In *Molecules in Space and in the Laboratory, Proceedings of a workshop held 2-5 June 1999 in Carloforte, Cagliari.*, volume 67, pages 179–. I. Porceddu, and S. Aiello. Bologna, Italy: Italian Physical Society, Conference Proceedings, 2000.

[NAS95] NASA/Science Office of Standards and Technology. Definition of the Flexible Image Transport System (FITS). Technical Report NOST 100-1.1, NASA/NOST, September 1995.

[OHH^{+}06] W. O'Mullane, J. Hernandez, C. Huc, T. Levoir, and F. De Angeli. Java Coding Standards and Guidelines for DPAC. Technical report, July 2006. GAIA-C1-SD-ESAC-WOM-005.

[OHHL05] W. O'Mullane, J. Hernandez, J. Hoar, and U. Lammers. Architecture of the AGIS System. Technical report, December 2005. GAIA-C3-SD-ESAC-WOM-004.

[OL05] W. O'Mullane and U. Lammers. Work Breakdown Structures for gaia DPAC. Technical report, September 2005. GAIA-C1-TN-ESAC-WOM-001.

[OLNS^{+}05] W. O'Mullane, N. Li, M. Nieto-Santisteban, A. Szalay, A. Thakar, and J. Gray. Batch is back: Casjobs, serving multi-tb data on the web. In *International Conference on Web Services*, July 2005. Also MS technote `http://arxiv.org/pdf/cs.DC/0502072`.

[O'M05] W. O'Mullane. Large scientific data systems - analysis of some existing projects and their applicability to gaia. Technical report, September 2005. GAIA-C1-HA-ESAC-WOM-003.

[ONHL07] Willam O'Mullane, Anne-Therese Nguyeni, John Hoar, and Tim Lock. Dpac software configuration management plan. Technical report, ESAC, Feb 2007. GAIA-C1-SP-ESAC-WOM-012-D.

[PB88] G. Francou P. Bretagnon. Planetary theories in rectangular and spherical variables. VSOP87 solutions. *Astronomy & Astrophysics*, 202:309, 1988.

[PdBG^{+}01] M. A. C. Perryman, K. S. de Boer, G. Gilmore, E. Høg, M. Lattanzi, L. Lindegren, X. Luri, F. Mignard, O. Pace, and P. T. de Zeeuw. GAIA: Composition, formation and evolution of the Galaxy. *A&A*, 369:339–363, 2001.

[Per04] M. Perryman. Estimation of Gaia Data Processing FLOPs. Technical report, August 2004. GAIA-MP-009.

[PFCT] Jordi Portell, Claus Fabricius, Javier Castaneda, and Jordi Torra. Technical report.

[PFea07] J. Portell, C. Fabricius, and E. Masana et al. Proposal of telemetry model for gaia-3. Technical report, March 2007. GAIA-CH-TN-UB-JP-001-1.

[PGF^{+}06] P. Protopapas, J. M. Giammarco, L. Faccioli, M. F. Struble, R. Dave, and C. Alcock. Finding outlier light curves in catalogues of periodic variable stars. *MNRAS*, 369:677–696, June 2006.

[Pic98] A. J. Pickles. A Stellar Spectral Flux Library: 1150-25000 Å. *PASP*, 110:863–878, July 1998.

[Pou02] D. Pourbaix. Precision and accuracy of the orbital parameters derived from 2d & 1d space observations of visual or astrometric binaries. *A&A*, 385:686, 2002.

[RBBdL06] A. Recio-Blanco, A. Bijaoui, and P. de Laverny. Automated derivation of stellar atmospheric parameters and chemical abundances: the MATISSE algorithm. *MNRAS*, 370:141, 2006.

[Ric01] H. Rickman. Reports on Astronomy. *Trans. IAU*, XXIV B, 2001.

[Ric04] G. T. Richards. Efficient Photometric Selection of Quasars from the Sloan Digital Sky Survey: 100,000 $z < 3$ Quasars from Data Release One. *ApJS*, 155:257–269, 2004.

[RRDP03] A. C. Robin, C. Reylé, S. Derrière, and S. Picaud. A synthetic view on the structure and evolution of the Milky Way. *Astronomy & Astrophysics*, 509:523–540, 2003.

[S0̈5] S. Söderhjelm. Census of Binaries the Big Picture. In C. Turon, K. S. O'Flaherty, and M. A. C. Perryman, editors, *ESA SP-576: The Three-Dimensional Universe with Gaia*, pages 97–+, January 2005.

[Sch05] J. Schneider. Can the Perturbation of a Stellar Motion in a Triple System Mimic a Planet. In Turon et al. [TOP05], pages 263–266.

[SCLS01] A. Sozzetti, S. Casertano, M. G. Lattanzi, and A. Spagna. Detection and measurement of planetary systems with GAIA. *A&A*, 373:L21–L24, July 2001.

[1] A.Defosseux (SILOGIC). Java benchmark report and recommendations. Technical report, Centre National d'Etudes Spatiales. GAIA-NT-31000-39-SI.

[SIK^{+}04] V. Smolčić, Ž. Ivezić, G. R. Knapp, R. H. Lupton, K. Pavlovski, S. Illić, D. Schlegel, J. A. Smith, P. M. McGehee, N. M. Silvestri, S. L. Hawley, C. Rockosi, J. E. Gunn, M. A. Strauss, X. Fan, D. Eisenstein, and H. Harris. A second stellar color locus: a bridge from white dwarfs to m stars. *ApJ*, 615:L141, 2004.

[Sim83] J.L. Simon. Thorie du mouvement des quatre grosses plantes. Solution TOP82. *Astronomy & Astrophysics*, 120:197, 1983.

[SKP*et al.*03] M. Soffel, S. A. Klioner, G. Petit, and *et al.* The IAU 2000 resolutions for astrometry, celestial mechanics, and metrology in the relativistic framework: explanatory supplement. *AJ*, 126:2687–2706, 2003.

[Sod04] S. Soderhjelm. Theoretical modelling of observational double-star distribution functions. Technical report, 2004. DMS-SS-05.

[Soz05] A. Sozzetti. Astrometric methods and instrumentation to identify and characterize extrasolar planets: A review. *PASP*, 117:1021–1048, 2005.

[Tea05] Gaia Project Team. Gaia Mission Requirement Document. Technical report, May 2005. GAIA-EST-RS-553.

[Tho84] D. J. Tholen. Asteroid taxonomy from cluster analysis of photometry. *Ph.D. Thesis*, 1984.

[TKBJ+07] P. Tsalmantza, M. Kontizas, C.A.L. Bailer-Jones, B. Rocca-Volmerange, R. Korakitis, E. Kontizas, E. Livanou, A. Dapergolas, I. Bellas-Velidis, A. Vallenari, and M. Fioc. Towards a library of synthetic galaxy spectra and preliminary results of the classification and parametrization of unresolved galaxies from gaia. *A&A*, submitted, 2007.

[TKK+06] P. Tsalmantza, M. Kontizas, R. Korakitis, B. RoccaVolmerange, E. Kontizas, E. Livanou, I. BellasVelidis, A. Dapergolas, C. A. L. BailerJones, A. Vallenari, and M. Fioc. A first library of synthetic galaxy spectra with pegase.2 code for gaia purposes. Technical report, Jul 2006. GAIA-C8-TN-UOA-PAT-001-1.

[tLdWG05] M. ter Linden, H. de Wolf, and R. Grim. Gridassist, a user friendly grid-based workflow management tool. In *2005 International Conference on Parallel Processing Workshops (ICPPW'05)*, volume icppw, pages 5–10. IEEE Computer Society, 2005.

[TOP05] C. Turon, K. S. O'Flaherty, and M. A. C. Perryman, editors. 2005.

[Vac02] M. Vaccari. Gaia galaxy survey - a multicolour galaxy survey with gaia. Technical report, February 2002. GAIA-CUO-105.

[VCdFL03] A. Vecchiato, M.T. Crosta, F. de Felice, and M.G. Lattanzi. The oato-pd ppn relativistic model for gdaas2: I. model overview. Technical report, 2003. Technical note on Livelink (RRFWG–TOPD–001).

[vL97] F. van Leeuwen. The hipparcos mission. *Space Sci. Rev.*, 81(3–4):201–412, 1997.

[vL05] F. van Leeuwen. Rights and wrongs of the hipparcos data. a critical quality assessment of the hipparcos catalogue. *A&A*, 439:805–822, 2005.

[VLB$^+$03] A. Vecchiato, M. G. Lattanzi, B. Bucciarelli, M. Crosta, F. de Felice, and M. Gai. Testing general relativity by micro-arcsecond global astrometry. *A&A*, 399:337–342, 2003.

[vLF05] F. van Leeuwen and E. Fantino. A new reduction of the raw hipparcos data. *A&A*, 439:791–803, 2005.

[VSHM03] D. Vande Putte, R. C. Smith, N. A. Hawkins, and J. S. Martin. A spectroscopic search for faint secondaries in cataclysmic variables. *MNRAS*, 342:151–162, June 2003.

[vT06] F. van Leeuwen and S. Theil. Gaia attitude modelling and reconstruction. Technical report, June 2006. GAIA-C3-TN-IOA-FVL-002.

[WE99] M. I. Wilkinson and N. W. Evans. The present and future mass of the Milky Way halo. *MNRAS*, 310:645–662, December 1999.

[Wil93] C. Will. *Theory and experiment in gravitational physics*. Cambridge University Press, 2 edition, 1993.

[WKBJ04] P. G. Willemsen, T. A. Kaempf, and C. A. L. Bailer-Jones. Identification and parametrization of spectroscopic binaries by medium band photometry. Technical report, Bonn, June 2004. ICAP-PW-003.

[WKBJ06] P. G. Willemsen, T. A. Kaempf, and C. A. L. Bailer-Jones. Estimates of the precision of stellar parameters derived from simulated rp/bp spectra. Technical report, DLR, April 2006. GAIA-C8-TN-DLR-PW-009.

[WL06] J. Warell and C.-I. Lagerkvist. Asteroid taxonomy in the gaia photometric system. *A&A*, submitted, 2006.

[WVT$^+$05] M. I. Wilkinson, A. Vallenari, C. Turon, U. Munari, D. Katz, G. Bono, M. Cropper, A. Helmi, N. Robichon, F. Thévenin, S. Vidrih, T. Zwitter, F. Arenou, M.-O. Baylac, G. Bertelli, A. Bijaoui, F. Boschi, F. Castelli, F. Crifo, M. David, A. Gomboc, A. Gómez, M. Haywood, U. Jauregi, P. de Laverny, Y. Lebreton, P. Marrese, T. Marsh, S. Mignot, D. Morin, S. Pasetto, M. Perryman, A. Prša, A. Recio-Blanco, F. Royer, A. Sellier, A. Siviero, R. Sordo, C. Soubiran, L. Tomasella, and Y. Viala. Spectroscopic survey of the Galaxy with Gaia- II. The expected science yield from the Radial Velocity Spectrometer. *MNRAS*, 359:1306–1335, June 2005.

[YIL06] E. Masana Y. Isasi and X. Luri. Gaia system simulator (gass) overview. Technical report, June 2006. GAIA-C2-TN-UB-YI-001-1.

[ZTT85] B. Zellner, D.J. Tholen, and E.F. Tedesco. The eight-color asteroid survey: results for 589 minor planets. *Icarus*, 61:355, 1985.

L Letters of Commitment

Letters of Commitment appended to the proposal

Country	Agency
Belgium	BELSPO
France	INSU
France	CNES
Germany	DLR
Germany	ZAH
Germany	Technische Universität Dresden
Greece	University of Athens
Greece	GSRT
Italy	ASI
Italy	INAF
Spain	MEC/PNE
Sweden	SNSB
Switzerland	University of Geneva
Switzerland	SSO
United Kingdom	PPARC

BELGIAN SCIENCE POLICY

Wetenschapsstraat 8 rue de la Science
B-1000 BRUSSELS
Tel. +32 2 238 34 11 ▪ Fax +32 2 230 59 12
www.belspo.be

Dr. François MIGNARD
OCA/Cassiopee
Le Mont Gros
BP 4229
06304 Nice Cedex 4
France

your reference	**our reference**		**enclosure(s)**
	\S\scienc\gaia\DPAC_AO_Belgium\06_WV_LoC.do c		
contact	**e-mail**	**telephone**	**date**
Werner VERSCHUEREN	verw@belspo.be	+32.2.23.83.589	21/11/2006

Subject : **GAIA – DPAC - Letter of Committment**

Dear Dr. Mignard,

I hereby send you, in your capacity of coordinator of the response of the scientific community to the ESA AO for the GAIA DPAC, our committments regarding the Belgian participation to this response.

BELSPO has committed itself to provide the following support via the ESA PRODEX Programme:

- 4 Institutes are supported: KUL (Leuven), ROB (Brussels), ULB (Brussels), ULg (Liège).
- 8 teams are supported within these 4 institutes, see the list below.
- each of the **4** institutes is provided with 1 **JAVA expert** dedicated to GAIA only.
- **6 postdoc** positions are provided to 5 teams.
- all 8 teams receive funding for **functioning costs** (travel, small equipment).
- all the support described is given for an initial period of **2 years (2007-2008)**; **extensions** are possible but are dependent on an assessment of the work done and still to be done near the end of 2008.

The support is distributed as follows:

- Work on QSO's within CU8
  ULg, Liège (Dr. Jean Surdej and Dr. Jean-François Claeskens)
  1 postdoc + shared use within ULg of 1 JAVA expert for GAIA + functioning costs.

- Variability characterisation and classification within CU7
  KUL, Leuven (Dr. Conny Aerts)
  1 postdoc + 1 JAVA expert + functioning costs.

- Variability characterisation and classification within CU7
  ROB, Brussels (Dr. Jan Cuypers and Dr. Peter De Cat)
  shared use within ROB of 1 JAVA expert for GAIA + functioning costs.

- Impact of variability and spots on astrometry within CU7
  ULB, Brussels (Dr. Alain Jorissen)
  shared use of the resources provided to Dr. Dimitri Pourbaix, see below.

- Management of CU4 + work on binaries within CU4
  ULB, Brussels (Dr. Dimitri Pourbaix)
  2 postdocs + 1 JAVA expert + functioning costs.

- Work on spectroscopic binaries within CU4 + Work on radial velocities of binaries within CU6 + Support to the treatment of WR stars within CU8
  ULg, Liège (Dr. Eric Gosset and Dr. Gregor Rauw)
  1 postdoc + shared use within ULg of 1 JAVA expert for GAIA + functioning costs.

- Work on radial and rotational velocities within CU6 + Work on the Extended Stellar Parametrizer and spectral modelling within CU8
  ROB, Brussels (Dr. Yves Frémat and Dr. Ronny Blomme)
  1 postdoc + shared use within ROB of 1 JAVA expert for GAIA + functioning costs.

- Work on Solar System objects within CU4
  ROB, Brussels (Dr. Thierry Pauwels and Dr. Peter De Cat)
  shared use within ROB of 1 JAVA expert for GAIA + functioning costs.

All support described above is evidently conditional to selection by ESA of the corresponding Belgian contribution to DPAC. Support to be provided beyond 2008 is conditional to the availability of the necessary funds in the Belgian PRODEX Programme enveloppe.

with best regards,

Dr. Werner Verschueren
Belgian ESA Delegation – Space Sciences and Exploration

Institut National des Sciences de l'Univers

INSU

3 rue Michel Ange - BP 287
75766 Paris Cedex 16
Téléphone : 0144 96 43 77
Télécopie : 01 44 96 53 50

November 24th , 2006

CNRS – INSU commitment to the Gaia DPAC

To whom it may concern,

The GAIA mission is a very ambitious and structuring project in Europe and in particular for the French community. It is aimed to answer some of the most difficult questions in modern astronomy: create a 3-D map of about one billion stars in our Galaxy to trace back its history.

Out of the nearly 300 DPAC members, the French astronomy community represents the single largest group with ~ 70 people working in French institutes, either with permanent position (most of them) of with temporary PhD or pos-doctoral position. This corresponds to 20-25 FTE involved in the preparation of the Gaia data processing. They are found primarily in the determination of the astrophysical parameters and in the analysis of spectroscopic data, complex objects and simulation, respectively in CU8, CU6, CU4 and CU2.

The Gaia mission has been recognised by the national astronomy advisory committees as one of the key astronomical space program with a major involvement of French institutes and this is clearly a priority for CNRS and INSU in this area of basic research.

As a consequence, we will support the efforts of the French teams by appointing in due time the necessary human and technical resources, and providing the financial support to cover their activities until the completion of the program. The amount of financial and human resources will be subjected to the annual budget appropriation procedures and to the national regulations for the hiring of public servants.

INSU is convinced that the proposal submitted by the DPAC to ESA is a great opportunity for European astronomical research and expects its approval by the ESA SPC early next year.

Yours faithfully.

Dominique LE QUEAU

Director

cnes

CENTRE NATIONAL D'ÉTUDES SPATIALES

Direction de la Prospective, de la Stratégie,
des Programmes, de la Valorisation
et des Relations Internationales
Programme Etude et Exploration de l'Univers

Dr. Fabio Favata
Astronomy and Fundamental Physic
Mission Coordinator
ESA D/SCI
8-10 rue Mario Nikis
75738 PARIS cedex 15

Paris, le 21 novembre 2006
DSP/EU/2006-0023872/RB/ml

Dear Dr. Favata,

The present letter first aims at confirming that Centre national d'études spatiales (CNES) as a technical centre wishes to contribute to the implementation of the Gaia ground segment. The detailed content of the CNES proposed activities is described in the Gaia Data Processing and Analysis Consortium (DPAC) proposal sent as a response to ESA's Announcement of Opportunity for the GAIA Data Processing (ESA AO n°2006-1106). That contribution is focused on :

- the technical coordination and the quality assurance activities for the Coordination Units Simulation (CU2), Object processing (CU4), Spectroscopic processing (CU6) and Astrophysics Parameters (CU8),
- the development of a hardware and software host infrastructure in which the CU4, CU6 and CU8 software will be integrated, validated and operated ; CNES will be the Data Processing Centres for these CUs,
- the deployment and the operation in the CNES computing centre of the GIBIS simulator,
- technical activities at the system level addressed by CU1 (responsibilities shared with ESAC and other groups).

At the national level, the CNES contribution will be implemented in close cooperation with the French scientific laboratories which propose scientific and technical contributions to the Gaia ground segment activities also described in the consortium proposal.

Should the consortium proposal be selected by ESA and pending a positive recommendation of our science programme committee, which is expected by March 2007, CNES will fund its own technical activities and as the French space agency will provide as appropriate a financial support to cover the activities of the French laboratories involved in the DPAC. The level of resources allocated to the project will be subject to our budget appropriation procedures and to the availability of funds.

Siège : 2 place Maurice Quentin - 75039 Paris cedex 01 - tél. : 33 (0)1 44 76 75 00 - **www.cnes.fr**
Direction des lanceurs : Rond-Point de l'Espace - Courcouronnes - 91023 Evry cedex - tél. : 33 (0)1 60 87 71 11
Centre spatial de Toulouse : 18 avenue Edouard Belin – 31401 Toulouse cedex 9 - tél. : 33 (0)5 61 27 31 31
Centre spatial guyanais : BP 726 – 97387 Kourou cedex - tél. : 33 (0)5 94 33 51 11

RCS Paris B 775 665 912 Siret 775 665 912 000 82 code APE 731 Z N° d'Identification TVA. FR 49 775 665 912

Re-affirming our commitment to the full success of the Gaia project, we hope that this proposal will be accepted and we look forward to hearing the results of the selection process.

The CNES points of contact for the implementation of this effort will be :

- for programmatic issues Dr. Jean-Louis COUNIL (CNES DSP/E²U) jean-louis.counil@cnes.fr
- for technical issues Mr. Claude HUC (CNES DCT/PS/VDO) claude.huc@cnes.fr

Yours sincerely,

Richard BONNEVILLE
Strategy and Programmes Directorate
Head, Space Sciences and Exploration office

Deutsches Zentrum
für Luft- und Raumfahrt e.V.

German
Aerospace Center

DLR

Space Management Space Science

DLR Space Management
Königswinterer Str. 522-524, 53227 Bonn, Germany

Prof. Dr. Francois Mignard
OCA/Cassiopee
Le Mont Gros
BP 4229
06304 Nice Cedex 4

Frankreich

Fax: 0033 4 92 00 31 18

Your letter	
Reference	ESA-D.SCI/DJS/22292
Our reference	RD-RX
Your correspondent	Alois Himmes
Telephone +49 2 28 4 47-	346
Telefax +49 2 28 4 47-	745
E-mail	alois.himmes@dlr.de
Bonn-Oberkassel,	28.11.2006

Announcement of Opportunity for the GAIA Data Processing and Analysis

Letter of Committment for the funding of German Institutes

Dear Professor Mignard,

the Astronomisches Rechenzentrum (ARI), Heidelberg, the Max-Planck-Institute for Astronomy (MPIA), Heidelberg, the Technische Universität (TU) Dresden, and the Astrophysikalisches Institut Potsdam (AIP) intend to participate in the Data Processing and Analysis (DPA) of the GAIA mission. Four groups of scientists under the leadership of Dr. Ulrich Bastian (ARI), Dr. Coryn Bailer-Jones (MPIA), Dr. Sergei Klioner (TU-Dresden), and Prof. Matthias Steinmetz (AIP), have joined the GAIA Data Processing and Analysis Consortium (DPAC) and will contribute to the DPA with specific workpackages in several Coordination Units as has been outlined in the proposal to ESA.

DLR is co-funding contributions of these institutes for the preparation of the DPA since 2005. DLR is committed to continue this funding for the workpackages agreed within the DPAC proposal and endorses the planned contribution of the institutes within the DPAC. The level of the DLR funding has been agreed between the institutes and DLR.

It is understood that this committment is subject to the relevant DLR funding procedures. The level of the DLR support will be subject to the availability of funds within the German budget allocation for the GAIA Data Processing and Analysis.

Sincerely

i. V. Dr. Thomas Galinski
Head Space Science

i. A. Alois Himmes
Space Science

Copy: Dr. Ulrich Bastian, ARI-Heidelberg	Fax: 06221 541 888
Dr. Coryn Bailer-Jones, MPIA-Heidelberg	Fax: 06221 528 246
Dr. Sergei Klioner, TU-Dresden	Fax: 0351 4633 7019
Prof. Matthias Steinmetz, AIP-Potsdam	Fax: 0331 7499 267

Königswinterer Str. 522-524
53227 Bonn
Germany
Telephone +49 2 28 4 47-0

Ruprecht-Karls-Universität Heidelberg
Zentrum für Astronomie
– Direktor –

Prof. J. Wambsganss, ARI/ZAH Uni HD, Mönchhofstr. 12-14, 69120 Heidelberg

Prof. Joachim Wambsganss
Direktor ZAH
Mönchhofstrasse 12 – 14
69120 Heidelberg
Tel.: +49 (0)6221-54 1800
Fax: +49 (0)6221-54 1802
e-mail: jkw@ari.uni-heidelberg.de

European Space Agency
Director of Science
8 – 10 rue Mario Nikis
75738 Paris Cedex 15
France

November 9, 2006

Letter of Commitment to support the Gaia data reduction

The ARI/ZAH has been actively supporting the preparations for the Gaia data reduction since 2002. It is planned to continue this work till the launch of Gaia, and also to contribute to the actual data processing during the Gaia scientific mission and until the completion of the final Gaia data products, i.e. about two to three years after the termination of Gaia's scientific operations.

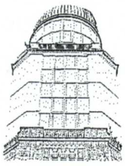
Part of the involvement of the ARI/ZAH is funded by the German space agency DLR, and we expect that this funding will be continued over the entire afore-mentioned timeframe. The in-house contributions of the ARI/ZAH presently correspond to 4.5 FTE of scientific staff and 2 FTE of technical staff, plus corresponding infrastructure and administrative overheads. We will keep at least this level throughout the afore-mentioned timeframe, and we are confident that we can increase it by 2 FTE of scientific staff within the next few years.

The commitments given in this letter express the firm intention of the ARI/ZAH. For legal reasons it is pointed out, however, that the commitments given here are subject to future budgets concerning the ARI/ZAH, which in turn are subject to legislation of the parliament of the state of Baden-Wuerttemberg.

Prof. Dr. Joachim Wambsganss

TECHNISCHE UNIVERSITÄT DRESDEN

Fakultät Forst-, Geo- und Hydrowissenschaften Lohrmann-Observatorium



Technische Universität Dresden, 01062 Dresden

To whom it may concern

Univ.-Prof. Dr. rer. nat. habil, director

Michael H. Soffel

Tel: 0351 4633-4200
Fax: 0351 4633-7019
E-Mail: Michael.Soffel@tu-dresden.de

Dresden, 5. November 2006

Subject: Letter of Commitment to support the Gaia data reduction

Dear Colleagues,

Lohrmann Observatory, TU Dresden has been actively supporting the preparations for the Gaia data reduction since 2002. It is planned to continue this work till the launch of Gaia, and also to contribute to the actual data processing during the Gaia scientific mission and till the completion of the final Gaia products, i.e. about two to three years after the termination of Gaia's scientific operations.

Part of the involvement of the Lohrmann Observatory, TU Dresden is funded by the German space agency DLR, and we assume that this funding will be continued over the entire afore-mentioned timeframe.

The contribution from in-house funds of the Lohrmann Observatory, TU Dresden presently corresponds to 1.0 FTE of scientific staff and 0.5 FTE of technical staff, plus corresponding infrastructure and administrational overheads. This level is intended to be kept throughout the afore-mentioned timeframe. We hope to be able to increase this level by 0.5 FTE of scientific staff within the next years.

The commitments given in this letter express the firm intention of the institute. For legal reasons it is pointed out, however, that the commitments given here are subject to the future budgets of the institute, which in turn are subject to legislation of the parliament of the state of Saxony.

Sincerely Yours,

Michael Soffel

M. Soffel

Postadresse (Briefe)
Technische Universität Dresden
01062 Dresden

Postadresse (Pakete u.ä.)
Technische Universität Dresden
Helmholtzstraße 10
01069 Dresden

Besucheradresse
Sekretariat:
Beyerbau, Zi. 147a,
Georg-Bähr-Str.1
01069 Dresden

Internet
http://astro.geo.tu-dresden.de

National and Kapodistrian
University of Athens
for Financial Affairs
and Development
30, Panepistimiou, st.
106-79, Athens Greece
tel. 00302103689760, 00302103689712,
fax. 2103689711, e-mail vrec-fin@uoa.gr

Athens 08 November 2006

To whom it may concern.

Subject : **Letter of Commitment of Supporting institution**

Dear Sir,
The National and Kapodistrian University of Athens (NKUA) is interested in participating at the preparation of ESA GAIA mission. A team of 9 researchers (5 permanent staff and 4 young collaborators) from our University, The National Observatory of Athens (three of its permanent staff) and The National Technical University (one Associate Professor) participate in the relevant activities since 2002.

1) The NKUA is running a programme funded by the General Secretariat of Research and Technology, based at the Dpt Astrophysics, Astronomy and Mechanics since 2004 for three years, coordinated by Prof. M. Kontizas. The funding has allowed the team to be fully equipped with several computers and the necessary software to enable them to participate to this project. The necessary rooms for the senior collaborators and the young PhD students (three) and Post-doc (one) are located in the Dpt of Astrophysics, Astronomy and Mechanics. The junior people are paid by the programme.

2) Prof. Kontizas is also the team leader of a Marie Curie, RTN focused on GAIA activities, which will support another PhD student from abroad, started October 2006 for the next three years. This project is also hosted by NKUA under the agreement already signed with EU.

Prof Kontizas is participating to a European Consortium for the Data Processing of GAIA as manager of a top working package for the next 6-8 years. Under these circumstances NKUA is committed to allow this team to work in the premises of our University and use the infrastructure allocated already to them for all the necessary period till the end of GAIA mission.

Sincerely Yours

D.N. Asimak

Prof. D N Asimakopoulos
Vice Rector

HELLENIC REPUBLIC
MINISTRY OF DEVELOPMENT
GENERAL SECRETARIAT FOR
RESEARCH & TECHNOLOGY
International Scientific & Technological
Cooperation Directorate
International Organisation Division

Athens, 20 November 2006
Our Ref.: 502 (ΔΙΕ)

To Dr. François Mignard
Chair of the DPACE
OCA/Cassiopee
Le Mont Gros
BP 4229
06304 Nice Cedex 4 – France

Subject: "GAIA DPACE – Support of the Greek participation"

With the present letter we would like to underline the interest of Greece to continue its participation in the GAIA Data Processing and Analysis Consortium (DPAC).

The Greek research team, lead by Mrs. Mary Kontizas, has received the necessary national funding and has developed the required infrastructure, in both personnel and equipment that will allow its smooth participation and active involvement in the Work Package for the "Unresolved galaxy classifier".

Thanking you for your cooperation,

Sincerely Yours

Olga Sterghiou
Head of the Division

agenzia spaziale
italiana

Prot-CE-CS-2006-088

Il Commissario Straordinario

ASI - Agenzia Spaziale Italiana
AOO-ASI - AGENZIA SPAZIALE ITALIANA
REGISTRO UFFICIALE
Prot. n. **0009414** - 18/12/2006 - USCITA

Rome, 15th December 2006

Mr. Fabio FAVATA

Coordinator for Astronomy and
Fundamental Physics Missions
European Space Agency - SCI-CA
8-10 rue Mario Nikis
75738 Paris Cedex 15
France
Telefax +31-71-565 4697

Dear Mr. Favata,

this is to confirm the Italian Space Agency strong interest in supporting the Italian contribution to the GAIA programme, and in particular the Italian Scientific team, led by Mario Lattanzi from INAF-Osservatorio Astronomico of Torino.
Starting from the very beginning of 2007, the Italian scientific community involved in the GAIA programme will be funded through a contract ASI is going to award to INAF for three years covering also a number of different programmes and activities. Moreover, in the preparation for the next coming national three years plan, covering the timeframe 2007-2009, a dedicated budget line would be created to support properly the Italian participation to GAIA.
This process in on going and should be concluded on time for the next SPC meeting, and in any case in time to give ASI the possibility to sign the MLA for GAIA.

Let me take the occasion to outline that the Italian Space Agency is expecting to get the proper visibility and recognition of its own role in the GAIA programme and therefore we are sure that this expectation will be properly taken into account in the finalization of the MLA.

Best regards,

Prof. Avv. Vincenzo Roppo

Viale di Villa Grazioli, 23 00198 Roma - Tel 06/8567821/800/819 Fax 06/8567468

INAF
ISTITUTO NAZIONALE DI ASTROFISICA
NATIONAL INSTITUTE FOR ASTROPHYSICS

Il Presidente

Roma, 4 dicembre 2006
Prot. n. 5192/06/AC/PI

Dr. David Southwood
Director of Science
European Space Agency
rue Mario Nikis 8-10
75738 Paris (Cedex 15)
France

e p.c.

Dr. Fabio Favata
Astronomy and Fundamental Physics Missions Coordinator
European Space Agency
rue Mario Nikis 8-10
75015 Paris
France

Dear Dr. Southwood,
In my capacity as President of the Istituto Nazionale di Astrofisica (INAF), I confirm the Institute long term commitment to support the proposal for the constitution of the GAIA Data Processing and Analysis Consortium (DPAC).
The importance of Gaia for INAF is emphasized in the Institute 3-Year Plan, and a Letter of Intent, expressing interest in the Gaia DPAC, was sent to ESA in 2005.
INAF contribution to DPAC is foreseen in providing scientific and technical personnel to the project including infrastructures, salaries and administrative overheads.

INAF involvement in DPAC is supported, at the institutional level, through a national scientific project dedicated to the Gaia data processing and analysis coordinated by Dr. Mario G. Lattanzi.

From 2007 until the publication of the final catalog, INAF will provide approximately 22 FTE/year to the DPAC in terms of research and technical staff (with an effective involvement of more than 50 permanent scientists and technicians).

Sincerely Yours,

Piero Benvenuti

viale del Parco Mellini, 84
I – 00136 Roma
tel. +39.06.35533-310/311/312
fax +39.06.35343154
e-mail: presidenza@inaf.it
www.inaf.it

MINISTERIO
DE EDUCACIÓN
Y CIENCIA

SECRETARÍA DE ESTADO
DE POLÍTICA CIENTÍFICA
Y TECNOLÓGICA

DIRECCIÓN GENERAL
DE INVESTIGACIÓN

Dr. F. Mignard
Chairman
Gaia Data Processing Consortium
Oca/ Cassiopee
Le Mont Gros
BP 4229
06304 Nice Cedex 4
France

Madrid, November 30th, 2006

Dear Dr. Mignard:

This is to confirm that providing a significant contribution to the Gaia mission is one of the highest priorities of the Spanish *Programa Nacional del Espacio* (PNE) for coming years. This contribution will materialize in:

- Gaia simulator
- Photometric data processing
- Core data processing
- Automatic classification tools
- Providing computer facilities to support processing activities

Most of these activities are currently funded by PNE, which uses a three-year slot for financing. Spanish data processing centers CESCA (*Centre de Supercomputació de Catalunya*) and BSC (*Barcelona Supercomputing Center*) are only partially financed by PNE for Gaia related activities, while these facilities are supported by specific governmental agreements.

According to the rules of the PNE, the whole contribution to GAIA will be managed by the Spanish Lead Scientist, presently Dr. J. Torra. Recently, his proposal has been resolved positively and the funding requested for the years 2007-2009 has been granted. Beyond that, a proposal to the PNE should be presented and will be resolved on a competitive basis after proper evaluation, and depending on budgetary conditions.

Finally, I can confirm you that PNE is willing to negotiate with ESA a bilateral agreement regarding hardware facilities.

Sincerely yours,

Fdo: Ramón López de Arenosa Díaz

Jefe del Departamento de
Tecnologías de la Producción y Comunicaciones

2006-12-01 Dnr 238/06

Prof. Francois Mignard
Chair of Gaia Data Processing and Analysis
Consortium (DPAC)

Letter of Endorsement

The Swedish National Space Board (SNSB) is aware of the on-going and planned participation by Swedish scientists in the Gaia Data Processing and Analysis Consortium (DPAC).

It is our understanding that the Swedish participation is led by Prof. Lennart Lindegren in Lund (CU1[1] and CU3), Dr. Andreas Korn in Uppsala (CU8) and Claes-Ingvar Lagerqvist, also in Uppsala (CU4 and CU8). It is further understood that the work mainly concerns scientific preparations and software development.

SNSB has, advised by its Space Research Advisory Committee, favourably reviewed both the scientific and programmatic aspects of the Swedish participation in the Gaia project. To this end SNSB is committed to allocate reasonable resources to allow Swedish research groups to contribute in a significant way to the DPAC work. However, due to limited funds a prioritisation of SNSB's financial support will be necessary, based on an assessment of the importance of the proposed participation for the realisation of the Gaia project and to the fulfilment of the goals of the Gaia mission. The following priorities are presently adopted and will be implemented in an appropriate way:

1. Participation in CU1 and CU3
2. Participation in CU8
3. Participation in CU4

The above commitment is made subject to the present understanding of the project needs and the availability of appropriated funds.

Yours Sincerely,

Lennart Nordh
Director, Space Research

Copy: Lennart Lindegren
Andreas Korn
Claes-Ingvar Lagerqvist

---

[1] CU = Coordination Unit within DPAC

**UNIVERSITÉ
DE GENÈVE**

FACULTÉ DES SCIENCES
Département d'astronomie

Observatoire de Genève
51, chemin des Maillettes
Sauverny
CH-1290 Versoix
Switzerland

Tél. +41 22 379 22 00 – Fax +41 22 379 22 05
www.unige.ch/sciences/astro

Prof. Gilbert Burki
Directeur

Ligne directe : +41 22 379 23 90
Gilbert.Burki@obs.unige.ch

Dr François MIGNARD
Chair of the DSPACE
OCA/Cassiopée
Le Mont Gros, BP 4229
06304 Nice Cedex 4
France

Sauverny, 27 November 2006

Concerne: Support to GAIA DPAC CU-7

Dear Chairman of DPACE, Dear Dr MIGNARD,

The Geneva Observatory, including the Integral Science Data Center (ISDC) as an attached institute, confirms a strong local interest to participate actively to the Gaia space mission, and especially to the management of the Coordination Unit 7, *Variability Processing.*

The following ressources will be guaranteed by the Direction and Staff Members of Geneva Observatory (Department of Astronomy of Geneva University):

- the scientific position of the CU7 manager, Dr Laurent EYER;
- the complements to the salaries of two PhD students, according to the Geneva University regulations. The main part of these salaries will be covered by Swiss National Foundation grants (through subsidy requests) and by the EU financement;
- the necessary offices and working facilities for the collaborators associated to the Gaia project at Geneva, in particular those founded by Prodex and FNS grants;
- the necessary computational facilities for these collaborators, including their upgrading;
- the administrative and technical support;
- the access to the Geneva facilities at ground-based telescopes (photometry and spectroscopy), via the local Program Committee, prior and during the mission.

The Coordination Unit 7 is already working hard since several months, in collaboration with several groups in Europe. This effort will be continued at Geneva, towards a successful Gaia Mission.

Sincerely yours,

Prof. Gilbert BURKI

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Home Affairs FDHA

State Secretariat for Education and Research SER
Swiss Space Office

CH-3003 Bern, SER, **GAIA CU-7**

Dr François Mignard
Chair of the DPACE
OCA/Cassiopée
Le Mont Gros, BP 4229
06304 Nice Cedex 4, France

Reference: 912.14D2
Your Ref.:
Our Ref.: FR
Official in charge: Jakob Frauchiger
Bern, 24 November 2006

Support to GAIA DPAC CU-7

Dear Chairman of DPACE, Dear Dr. Mignard,

This letter of support is to express the awareness of the Swiss Space Office of your response to the ESA Announcement of Opportunity for the GAIA Data Processing. Assuming that the proposal you submit as the team coordinator of the Coordination Unit 7 (CU-7 and associated Data Processing Centre) of the Data Processing and Analysis Consortium (DPAC) shall be selected, the Swiss Space Office will enter in procedures to evaluate the allocation of an appropriate part of resources to the part of the activity to be conducted in Switzerland by the PI and CU-7 Manager & Scientific Coordinator Dr. Laurent Eyer at the Geneva Observatory.
Until now 0.7 M€ of funding has been allocated to pre-developments and concept studies for GAIA variability data processing. For future phases of the CU-7 data processing ground segment we expect further requests for a continuation of Prodex funding and national resources to step in. According to our best estimates of cost and risk at the time of signature of this letter, for the full duration of the project, further requests of the Swiss CU-7 PI of the Geneva Observatory will be positively regarded; however, the allocation of funding will be subject to national evaluation criteria and the availability of funds.

Sincerely

State Secretariat for Education and Research SER

J. Frauchiger

Jakob Frauchiger
Scientific Programmes Manager

Copy: - Dr. Laurent Eyer, Swiss PI CU-7

State Secretariat for Education and Research SER
Jakob Frauchiger
Hallwylstrasse 4, CH-3003 Bern
Phone +41 31 322 99 67, Fax +41 31 322 78 54
Jakob.Frauchiger@sbf.admin.ch
www.sbf.admin.ch

PPARC

The UK's Strategic Science Investment Agency

Particle Physics and Astronomy Research Council
Polaris House, North Star Avenue, Swindon
Wiltshire SN2 1SZ United Kingdom

Professor Richard Wade
Director Programmes & Deputy Chief Executive

Tel +44 (0)1793 442028 Fax +44 (0)1793 442036
Email richard.wade@pparc.ac.uk

23 November 2006

To whom it may concern:

UK – PPARC commitment to GAIA DPAC

This letter is provided as endorsement for the programme of work towards the GAIA DPAC as proposed by the UK consortium, led by the Institute of Astronomy, Cambridge. PPARC has considered the proposal in detail, fully reviewing the costs, staff resources and schedule. As a result PPARC Council has, at its meeting in June 2006, given approval to negotiate for the UK's role in the Data Flow System of the European Space Agency's (ESA),

It is expected that the proposed UK contribution will:

- design, develop, document, implement and test the software required for the photometric data processing for all GAIA CCDs;
- design and build the data handling system to process the photometric data;
- design, develop, document, implement and test the software required for a very significant part of the spectroscopic data processing and calibrations;
- design, develop, document, implement and test the software for the in-flight CCD calibrations;
- provide the managerial structure required to plan and coordinate the above activities and to ensure the highest standards on the delivered products.

PPARC has made provision for a cost to PPARC of up to £10M for the period from 2006 to 2012 (approximately 3 months after launch). This will cover both hardware and staff effort costs, together with appropriate contingency

Council expects that this level of commitment will enable the UK to maintain the PI-ship for this programme which is a key element of the mission.

Council will meet again on December 11th to confirm these funds, with the expectation that full commitment will be released on completion of the ESA assessment exercise and appropriate contractual discussions in 2007.

I hope that this evidence of support will assist you in the consideration of the UK proposal.

Regards

M Index