Proposal for the object numbering scheme

prepared by: F. De Angeli, F. van Leeuwen, J. Hoar, W. O'Mullane

approved by: DPACE
reference: GAIA-C1-MN-IOA-FDA-002-3
issue: 3
revision: 0
date: Wednesday 21st March, 2007
status: Issued

Abstract

This is a proposal for the numbering of objects observed by Gaia.

Document History

Issue	Revision	Date	Author	Comment
3	0	16-03-2007	WOM	NESTED specified - Uli spotted this was missing.
2	4	07-02-2007	WOM	Ron’s intro para added
2	3	06-02-2007	WOM	Ron’s comments
2	2	06-02-2007	WOM	More DPACE comments, bib
2	1	02-02-2007	WOM	DPACE comments, keep IDs, Solar System
1	2	10-01-2007	WOM	Very minor text change - approve
1	1	09-01-2007	FDA	Changed document class to dpacdoc
1	0	21-08-2006	FDA	Incorporated comments from W. O’Mullane
D	0	18-08-2006	FDA	First draft

Contents

1 Introduction

This is a proposal for the numbering of objects observed by Gaia. The proposal is based on an idea developed by William O'Mullane aimed at defining a convenient numbering scheme for the object observed by Gaia during the whole mission.

All observations must be matched to sources in the MDB. If no corresponding object is found for a set of associated observations, then a new object must be created to which the observations can be matched. For non-moving objects (stars, galaxies, etc.) this crossmatching is done during IDT and refined by IDU. For moving objects (observations of solar system objects) associating the observations to single objects is more complicated and is done in further CU4 processing.

This note describes the object numbering scheme by which each MDB object will be uniquely identified. This object identifier is only intended to be regarded as a unique number assigned to a source. Although the proposed scheme includes spatial information, the nature of the generation process means that this might not be very accurate or reliable. In the Main Database, precise spatial index information will be provided, currently considered to be HEALPix and HTM identifiers.

This document proposes a numbering scheme composed of a spatial and sequence information. We consider the case of stellar [1] and solar system objects and outline some of the special cases.

2 Problem

The cross-matching process during the IDT and IDU will continuously update the catalogue of observed objects in the Main Database (MDB). A convenient way to keep track of objects throughout different versions of the MDB is needed by several tasks. It would be desirable to have a numbering system which can remain fixed over the whole mission.

This document describes MDB object numbering. In the final catalogue a different identification scheme could be adopted, at the expense of the direct link between source identification in the catalogue and the corresponding data in the Main Database from which it was derived.

3 Proposed solution

The idea is to take advantage of a tessellation scheme to divide the sky into small pixels and to assign to each source a numeric identifier which keeps track of its position in the sky.

[1] In this context this includes Galaxies, Quasars etc., practically all objects which are not Solar System.

The two most popular tessellation schemes are:

1. HEALPix: Hierarchical Equal Area and iso-Latitude Pixelisation, homepage `http://healpix.jpl.nasa.gov/`
2. HTM: Hierarchical Triangular Mesh, homepage `http://www.sdss.jhu.edu/htm/`

For a detailed comparison of the two schemes, we refer to the paper "Splitting the sky - HTM and HEALPix" O'Mullane et al. (2001), also available via ftp from `ftp://ftp.rssd.esa.int/pub/sciproj/womullan_082000.pdf`.

Here it suffices to say that the motivations for developing the two schemes were quite different and this obviously reflects in their functionalities. HTM was designed having in mind its main application of partitioning a point source catalogue and providing complex trigonometric queries in spherical space. The main goal of HEALPix was to support fast numerical analysis on data over the sphere. It was designed to be hierarchical and to have equal-area pixels with an iso-latitude distribution to facilitate numerical analysis. HEALPix has already been adopted for the Planck mission.

For Gaia we will have to handle a large point-source catalogue and we will want to produce statistical plots and analyses of all sky properties. These we consider are best facilitated by the HEALPix scheme. Moreover the conversion of a pixel ID to and from a sky coordinate is more efficient in HEALPix than in HTM[2]. For these reasons we shall concentrate on HEALPix.

For more details about the HEALPix framework please refer to Górski et al. (2005) as well as to the HEALPix homepage where further useful references are listed.

3.1 The numbering scheme

Whatever numbering scheme we decide to use for our pixelisation of the sky sphere, the object identifier will be a 64 bits integer, composed of two 32 bits integers. The upper 32 bits represents the index number and the lower 32 bits the sequence number. The former is the HEALPix pixel number, in the NESTED scheme, for stellar objects and simply -1 for solar system objects. The latter is just a monotonically increasing counter defined within the pixel for stellar objects and a global counter for solar system objects. The use of -1 for the solar system objects provides a simple way to flag these objects which will otherwise receive a sequence number exactly as done for stars.

[2] A comparison between the Java implementation of HTM and HEALPix shows that the first is considerably slower than the second one already for a HTM depth equal to 12 equivalent to $N_{side} = 1024$. Moreover the recursive HTM calculation is expected to become more and more time consuming when increasing depth while HEALPix calculation times do not change with N_{side}. More details on this can be found in O'Mullane et al. (2001)

Assuming a HEALPix scheme with number of sides equal to 4096, the total number of pixels would be slightly more than 200 million ($N_{pixel} = 12 \times N_{side}^2$). This implies a mean of 5 observed objects per pixel. With $N_{side} = 4096$ the index number can be stored in 4 Bytes. With a total of 200 million pixels, the area per pixel is equal to $2.04905E - 04$ square degrees (≈ 0.7 square arcminutes). The expected maximum star density (based on the Baade's Window estimates) is of the order of $3 - 5$ million stars per square degree, thus the maximum number of stars falling in each pixel should be approximately 1000. This should be compared with the total number of sequence numbers allowed in 32 bits, which is several orders of magnitude larger.

3.2 An example

Thus, an object with ($\alpha = 10.0°, \delta = 30.0°$) has a HEALPix id of 16377824 (at $N_{side} = 4096$). Let us assume that this object is the 31st object to be allocated to this pixel.

This object would have a index number of 16377824 and sequence number of 31. Converting these numbers to hexadecimal (in Java hexadecimal notation) `0x00F9E7E0` and `0x0000001F`. The object identifier is thus `0x00F9E7E00000001F` which translates to a 64 bits Java integer as 70342218459643935.

For a solar system object again with a sequence number of 31, the index number is -1 (`0xFFFFFFFF`) which results in an object identifier of `0xFFFFFFFF0000001F`, or -4294967265 (64 bits Java integer)

4 Processes assigning Object Identifiers

If a Gaia Initial Catalogue is used for the purposes of improving the accuracy of the Gaia data reduction during the initial year of the mission, a source table will be constructed using object identifiers generated according to this scheme. A record of the mapping from the initial catalogue to the Gaia Initial Catalogue will be kept for traceability purposes.

During IDT, the cross-matching process will match observations against the sources defined in the current source table; if sources are not matched against the source table, the IDT process will create a new source from the observation (or observations) and assign an object identifier for it. IDT can derive from the source table the next sequence number to be assigned - the case where source creation is a parallel and distributed process complicates the mechanism but remains feasible. IDT may be able to differentiate observations of moving (e.g. solar system) and fixed [3](e.g. stellar) objects using the AC motion during a transit. Moving objects could be ignored by IDT as it can not match them. CU4 would then be responsible for the -1 id assignation. IDU

[3]Objects with no perceived apparent motion on the sky during a single focal plane crossing.

will also perform a similar cross-match process, but at a higher level of accuracy, incorporating calibration and attitude data from the AGIS.

A few special cases should be considered. Objects which have previously identified as single objects may be subsequently classified as multiple sources. In this case we suggest to assign new IDs to all the components. In the opposite case, where a previously classified multiple system is re-classified as a single source, we suggest again to assign a new ID. In such a situation the previous IDs would not be physically deleted, an identifier should never be re-used. It is foreseen that IDT and IDU are the primary processes which allocate object identifiers; the mechanism by which other processing systems can flag objects which require special treatment must be determined and defined in the Main Database ICD. The exact mechanism of which process assigns the number is not the topic of this note and is not as trivial as one may think in a distributed environment.

Mapping tables will be created in the Main Database that will allow tracking the history of a source across Main Database versions.

5 References

Górski K.M., Hivon E., Banday A.J., et al., Apr. 2005, ApJ, 622, 759

O'Mullane W., Banday A.J., Górski K.M., Kunszt P., Szalay A.S., 2001, In: Banday A.J., Zaroubi S., Bartelmann M. (eds.) Mining the Sky, 638–+