



---

# The Initial Gaia Source List 1.0: Production and Plans

---

prepared by: R. L. Smart, R. Morbidelli, L. Nicastro,  
R. Pannunzio, A. Spagna  
affiliation : INAF - Osservatorio Astronomico di Torino  
approved by:  
reference: Gaia-C3-TN-OATO-RLS-001-01  
issue: 01  
revision: 0  
date: 02 Apr 2008  
status: Issued

## Abstract

We discuss the production of the first version of the Initial Gaia Source List (IGSL). This is only an approximate version of what will be the final product but it is still useful now as it is of the expected size and content so can be used to develop routines for the Gaia reduction pipeline, as a reference point for auxiliary catalogs and for simulations. We explore the probable role of the IGSL in the Gaia data reduction and other possible uses for the mission. Finally we discuss the immediate and long term future for the IGSL.

Up-to-date discussion, known problems and status of the IGSL can be found at Gaia wiki page [http://www.rssd.esa.int/wikiSI/index.php?title=CU3:\\_Auxiliary\\_Data:\\_Initial\\_Gaia\\_Source\\_List](http://www.rssd.esa.int/wikiSI/index.php?title=CU3:_Auxiliary_Data:_Initial_Gaia_Source_List)

## Auxiliary info

**Main WP:** GWP-S-335-11000

**Affected WPs:** GWP-S-335-12000

**Reviewed by:**

**Distribution:** Ulrich Bastian, Ronald Drimmel, Mario Lattanzi, Marica Sarasso

## Acronyms used in this document

The following table has been generated from the on-line Gaia acronym list:

Acronym	Description
2MASS	Two-Micron All Sky Survey
CCD	Charge-Coupled Device
CTI	Charge Transfer Inefficiency
CU	Coordination Unit (in DPAC)
GAIA	Global Astrometric Interferometer for Astrophysics (obsolete; now spelled as Gaia)
ID	Identifier (Identification)
IDT	Initial Data Treatment
IGSL	Initial Gaia Source List
LSPM	Lepine Shara Proper Motion Catalog
MDB	Main Database
NOMAD	Naval Observatory Merged Astrometric Dataset
POSS	Palomar Observatory Sky Survey
QSO	Quasi-Stellar Object
SDSS	Sloan Digital Sky Survey
SIM	Space Interferometry Mission
TN	Technical Note
TOC	Table of Contents
UCAC	USNO CCD Astrograph Catalogue
URAT	USNO star catalogue
USNO	United States Naval Observatory
WP	Work Package

## IGSL1 in a nutshell

A compilation catalog delivered in 12/2007

Number of Objects: 940463736

Astrometry: Positions, proper motions

Photometry: 4 magnitudes in GSC2 bands

Included Catalogs: Tycho2, CU3QSO, UCAC2, SDSS-DR6, GSC23

Wiki discussion: CU3:\_Auxiliary\_Data:\_Initial\_Gaia\_Source\_List

Catalog Location: <http://gaia.esac.esa.int/mdbexp/mdbexp/>

## 1 Introduction

The Initial Gaia Source List (IGSL) is envisioned to be a snapshot of all celestial objects before Gaia launches nominally to a magnitude limit of  $G=20$  but when possible the fainter limit of  $r=21.5$ . It is planned to include the best astrometric information available in large (e.g.  $>10000$  square degrees) sky surveys. It will also include photometry in two or more all-sky bandpasses and a rudimentary classification. Here we discuss the possible uses of the IGSL, the first attempt at the creation of this list - IGSL1, the plans for the immediate future and possibilities for the long term.

## 2 Uses of the IGSL

The IGSL will have many functions but it's main purpose will be to simplify the interconnection of the numerous auxiliary catalogs which provide mission critical calibration and reference data with the Gaia observations. All auxiliary catalogs will be matched to the IGSL before launch thus providing each object with a name. The Gaia observations need only then be matched via position to the IGSL to allow direct access via the name to all the auxiliary data. The process of pre-matching will allow a cleaning and homogenization of the auxiliary data and the number of mismatches will be minimized as part of this process. The general software development for the Gaia reduction will be simplified by the IGSL and many potential problems will be resolved before launch. How the IGSL will be used in the actual reduction process, as a simple external reference catalog or to populate the Gaia main database before launch, is still under debate (see wiki page [www.rssd.esa.int/SA-general/Projects/Gaia/wiki/CU3:\\_Auxiliary\\_Data:\\_Initial\\_Gaia\\_Source\\_List:\\_Using\\_It](http://www.rssd.esa.int/SA-general/Projects/Gaia/wiki/CU3:_Auxiliary_Data:_Initial_Gaia_Source_List:_Using_It)).

Another important use of the IGSL will be to help identify science alerts during the first 12-18 months of operation while the Gaia sky is still incomplete. These science alerts will include identification of solar system objects, novae, super-novae, and other large variables. The IGSL

will also allow significant data-mining via the cross matching of various catalogs, freeing the user from large scale positional cross matches. The flags available can be used to identify auxiliary catalogs or also non-Gaia related information for example the presence of 2MASS or Denis observations or the selection of an object as a SIM grid star thus enhancing the possibilities for data-mining. The IGSL will allow transit predictions for the purpose of charge-transfer history tracking, especially early in the mission when the Gaia sky is incomplete; this is critical to charge-transfer inefficiency (i.e. radiation damage) corrections. Indeed the faint limit requirements of the IGSL will be driven by CTI concerns. Finally a subset of the IGSL will be used as the Attitude star catalogue for Gaia before the main IDT cross-matching process.

### 3 Source Catalogs for IGSL1

A quick perusal of the catalogs currently available would suggest that the Naval Observatory Merged Astrometric Dataset (NOMAD, Zacharias et al. (2005)) is a perfect choice for a first cut IGSL. The adoption of NOMAD does not however diminish greatly the work required to develop the IGSL and for a number of reasons we decided to develop the IGSL from scratch. These reasons include:

- 1) A new internal production has allowed us to customize the parameters directly making them more compliant with the overall IGSL goals. For example we have homogenized the magnitudes and used different data from the various catalogs based on what is available.
- 2) In the IGSL1 we have included the SDSS results and future productions of the IGSL will use data from the new all sky surveys such as PanSTARRs (<http://pan-starrs.ifa.hawaii.edu/>) and Skymapper (<http://www.mso.anu.edu.au/skymapper/>) so a development of the IGSL from individual source catalogs is nevertheless required.
- 3) We have preferred to not use the HIPPARCOS data in the production of the IGSL as at the launch of the Gaia in 2012 the more precise proper motions of Tycho-2 will render positions from Tycho-2 more precise.
- 4) The dominant source of data for NOMAD is the USNO-B, we have preferred to use the GSC23 for reasons described below.
- 5) Future productions of the IGSL will not be simply a compilation catalog in that we intend to use all observations to improve the parameters of the included data, NOMAD like IGSL1 is a simple compilation of data from "best" catalogs currently available.

The source catalogs used to develop the IGSL1, in hierarchical order, are given below:

Tycho2: Høg et al. (2000),  $2.4 \times 10^6$  stars, magnitude limit  $R_F < 12$ . This catalog forms the backbone of all the major ground based catalogs currently available. Since Gaia will not be launched until 2012 the more precise proper motions of Tycho 2 will render its positions more precise than those of HIPPARCOS. Future versions of the IGSL may try to improve more upon these proper motions combining other independent information (e.g Hipparcos, UCAC, PM2000) with the Tycho-2 data but since all recent large astrometric catalogs use Tycho-2 this

has to be done with some care.

CU3QSO: Andrei (2007),  $1.2 \times 10^5$  QSOs, magnitude limit  $R_F < 22$ . This is a compilation of QSOs with precise positions developed as part of the CU3 auxiliary catalog development. This catalog has only positions and nominal B magnitudes, where available we have included the GSC23 measured magnitudes.

UCAC2: Zacharias et al. (2004),  $4.8 \times 10^7$  mostly stars, magnitude limit  $R_F < 17$ . This is the most precise astrometric catalog in the range  $V=11-16$  currently available. Future versions of the IGSL will try to improve positions and proper motions for the fainter objects from this catalog using GSC23 / SDSS observations.

SDSS: <http://www.sdss.org/dr6/>,  $1.2 \times 10^8$  objects, magnitude limit  $R_F < 22$  for this version. This catalog provides precise astrometry, photometry and classification for one quarter of the sky. We have included all objects to  $r < 22$  transforming the SDSS magnitudes to GSC23 bands when needed and including the GSC23 proper motions when available.

GSC2.3: Lasker et al. (2008), <http://www-gsss.stsci.edu/Catalogs/GSC/GSC2/GSC2.htm>,  $1.0 \times 10^9$  objects, magnitude limit  $R_F < 21.5$ . This catalog forms the bulk of the IGSL. We have preferred this over the USNOB catalog primarily because Torino, where the IGSL is produced, is a partner in the development of the GSC2 project, and therefore while the two catalogs have similar problems those for the GSC23 we understand while the USNOB problems are not as familiar to us. For example both catalogs suffer from a large number of false detections around bright stars (Barron et al. 2007), allowing and correcting for these requires an intimate knowledge of the scanning and reduction procedures. We also feel the GSC23 is superior to the USNOB in a number of important areas: for example in the USNOB less than 30 % of the objects have been assigned proper motions and these are relative not absolute, the magnitudes in the USNOB are not as precise as the GSC2.3 because of the higher number of calibrating sequences in the GSC2.3. To be fair the GSC2.3 also has a number of unique problems, primarily a magnitude dependant systematic error in the positions and proper motions, we are confident that this problem will be addressed in time for the second release of the IGSL. For this version the GSC23 proper motions have been scrambled as they have not yet been released.

A note on bright stars ( $V < 8$ ). All catalogs are missing some bright stars and to complete GSC23 in this region it was supplemented with the sky2000 catalog. This sometimes lead to double entries which we have attempted to clean. The IGSL should be complete in this region and most double entries have been removed except for some multiple entries of high proper motion objects that remain across the whole of the magnitude range.

## 4 Production of the IGSL1

All the source catalogs have been included in a mysql database and matched to the GSC2.3 using a nearest neighbor approach with a limit of 3". Once this database has been build the production of the IGSL1 is done with quite simple sql commands that include observations based on the following hierarchy:

**Positions** taken in the priority order Tycho2 - CU3QSO - UCAC2 - SDSS6 - GSC23.

**Proper Motions** taken in priority order Tycho2 - UCAC2 - GSC23.

**Magnitudes** It was decided at the beginning to provide magnitudes in the northern POSS-II bands of the GSC2.3. For the case of the SDSS where many objects did not have GSC2.3 observations in all bands, or Tycho2 objects where the GSC2.3 magnitudes were not published, for consistency we have included estimates of the GSC2.3 magnitudes taken from the transformations given in the appendix. We also provide a estimate of the Gaia G magnitude for the SDSS magnitudes using the transformations provided in Gaia-C5-TN-UB-CJ-041 and for the GSC2.3 magnitudes using the transformations in Gaia\_ML\_022.

All entries are provided with a sourceID as described in Gaia-C1-MN-IOA-FDA-002 and further developed in Gaia-CD-TN-ARI-BAS-020-01. For now we are not including any components. The name is therefore a combination of the source healpix level 12 ID (*healpix12ID*) and the running number in the healpix level 6 (*runningnumberinhealpix6*) bit combined via the relation  $SourceID = (healpix12ID \ll 32) + (runningnumberinhealpix6 \ll 7)$ .

## 5 The Catalog

The IGSL1 is currently available under the Gaia MDB in under CU3/Auxdata/IGSL. It is composed of 3 databases:

- 1) IntialGaiaSourceList (mdbcu3auxdataigslinitialgaiasourcelist) with the basic positional, photometric and classification data as listed in the appendix.
- 2) Igs1References with references to the various input catalogs and transformations used, how each position, proper motion and magnitude has been derived from the source catalogs is indicated by the associated code in the main catalog and explained in this database.
- 3) MatchIgs1 is a lookup table with the cross reference of the original catalog names (sourceParamID), the IGSL identifier (sourceId) and the code in the Igs1References indicating the parameter that the data is used for (sourceParam). Note: the SDSS, UCAC and CU3QSO the original catalog names are simple numbers so are stored directly in the sourceParamID. TYCHO-2 objects were identified by 3 numbers (TYC1, TYC2 and TYC3) and we have combined these into one complete number. GSC2.3 objects have original names that are combination of region names and running number, e.g. S00000001 is the first object in the region S0000000, we have converted the region names to the "official" numerical values in the htm level 6 structure, hence the smallest is S0000000 = 32768 and N3333333 = 65535 is the highest.

The IGSL can presently be accessed using the Data Explorer Tool <http://gaia.esac.esa.int/mdbexp/mdbexp/> under Remote Databases, MainDbAtESAC, maindb, igsllist. Simple sql queries and plots can be performed using this tool but the data is currently only indexed on the sourceid. If large subsets of the IGSL are required then using the GAIA extractor tool is the best method. There will be tools provided to cross match auxiliary catalogs, or at a minimum the subset of positions, included in the MDB. If it appears that many of the objects from a given auxiliary catalog are not already present in the IGSL (for example if the auxiliary catalog goes much deeper) then this auxiliary catalog should be provided to IGSL builders for direct inclusion.

## 6 Future

The current version of the IGSL1 has three large problems: 1) there has been no attempt to homogenize the astrometric information, 2) the base catalog GSC2.3 has a number of known systematic errors and 3) there are probably substantial numbers of false detections or multiple observations of large proper motion objects. It is however useful for software development, and for the inclusion of auxiliary catalogs which are expected to be submitted to the Gaia MDB via the normal procedures then matched directly to the IGSL by the CU3 GWP-S-335-11000 work package group. The IGSL1 can also be used today for simulations as a fair representation of what we expect Gaia to see. The inclusion of the SDSS data also provides the possibility to go beyond the nominal limit of Gaia in that part of the sky.

The next step in the production of the IGSL will be to develop procedures for the homogenization of the coordinate systems using a moving mean method or overlapping circles (Bucciarelli et al. 1993, available at [www.rssd.esa.int/SA-general/Projects/Gaia/wiki/images/b/b1/Generalization\\_of\\_the\\_moving\\_mean.pdf](http://www.rssd.esa.int/SA-general/Projects/Gaia/wiki/images/b/b1/Generalization_of_the_moving_mean.pdf)). This we expect to carry out for both positions and proper motions providing more consistent and precise parameters. We will also include other proper motion catalogs (SPM,NPM,PM2000) and catalogs of high proper motion stars (Lépine & Shara 2005) which are one of the major sources of double entries. We plan to provide a second release of the IGSL when the new version of the GSC2.3 is released which will be corrected for the systematic errors noted above. We also expect to produce a number of auxiliary catalogs for example stars with known parallaxes, and the other CUs are expected to provide catalogs of standard stars.

The future for the IGSL is dependant on the progress of the planned sky surveys. As large astrometric catalogs are produced they will be included in the IGSL and new versions released. The USNO have plans for a new release, UCAC3, late in 2008 and perhaps even a deeper version, URAT (Zacharias et al. 2006), before Gaia flies. There is a realistic hope that the Skymapper in the south and PanSTARRS in the north will produce major contributions before the launch of Gaia. All these will greatly improve the quality of the IGSL both in allowing a substantial cleaning of false objects and in the overall quality of the basic data. In particular

the resolution limit of the base catalog the GSC2.3 is quite low, 3-4 arcseconds, the use of orthogonal transfer CCDs will improve this situation and hence the completeness of the IGSL especially in the galactic plane and other crowded regions.

## 7 References

- Andrei A., February 2007, QSO Catalogue for Gaia; task description for GWP-S-335-13000, URL <http://www.rssd.esa.int/llink/livmlink/open/2731207>
- Barron J.T., Stumm C., Hogg D.W., Lang D., Roweis S., Sep. 2007, ArXiv e-prints, 709
- Høg E., Fabricius C., Makarov V.V., et al., Mar. 2000, A&A, 355, L27
- Jordi C., May 2007, Photometric relationships between Gaia photometry and existing photometric systems, URL <http://www.rssd.esa.int/llink/livmlink/open/2760608>
- Lasker B., Lattanzi M.L., McLean B.J., et al., 2008, AJ, submitted
- Lépine S., Shara M.M., Mar. 2005, AJ, 129, 1483
- Zacharias N., Urban S.E., Zacharias M.I., et al., May 2004, AJ, 127, 3043
- Zacharias N., Monet D.G., Levine S.E., et al., Nov. 2005, VizieR Online Data Catalog, 1297, 0
- Zacharias N., Laux U., Rakich A., Epps H., Jul. 2006, In: Ground-based and Airborne Telescopes. Edited by Stepp, Larry M.. Proceedings of the SPIE, Volume 6267, pp. 62670P (2006)., vol. 6267 of Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference

## Appendix A: IGSL1 Contents

Columns included in the IGSL1 along with type and units

PARAMETER	TYPE	COMMENT/UNITS
solutionId	BOOLEAN	Blank
creationDate	BOOLEAN	Blank
sourceId	BIGINT	Gaia ID
alpha	DOUBLE	Radians
delta	DOUBLE	Radians
alphaError	FLOAT	Radians
deltaError	FLOAT	Radians
alphaEpoch	FLOAT	Time
deltaEpoch	FLOAT	Time
sourcePosition	SMALLINT	In Igs1References
muAlpha	FLOAT	mu-mas/yr
muDelta	FLOAT	mu-mas/yr
muAlphaError	FLOAT	mu-mas/yr
muDeltaError	FLOAT	mu-mas/yr
sourceMu	SMALLINT	In Igs1References
magB_J	FLOAT	Mags, POSS B
magB_JError	FLOAT	Mags
sourceMagB_J	SMALLINT	In Igs1References
magV	FLOAT	Mags, POSS V
magVError	FLOAT	Mags
sourceMagV	SMALLINT	In Igs1References
magR_F	FLOAT	Mags, POSS R
magR_FError	FLOAT	Mags
sourceMagR_F	SMALLINT	In Igs1References
magI_N	FLOAT	Mags, POSS I
magI_NError	FLOAT	Mags
sourceMagI_N	SMALLINT	In Igs1References
magG	FLOAT	Mags
magGError	FLOAT	Mags
sourceMagG	SMALLINT	In Igs1References
class	BOOLEAN	0=point-like 1=extended
auxQSO	BOOLEAN	In CU3 QSO list

## Appendix B: Magnitude Transformations

Transformations used to determine magnitudes in GSC2 passbands when not available in the GSC2.3.

Tycho2 objects in the GSC2.3 have magnitudes replaced by the  $V_T$  and  $B_T$  of Tycho. To obtain blue ( $B_J$ , magB\_J), visual ( $V$ , magV) and red magnitudes ( $R_F$ , magR\_F) for the IGSL1 we have applied the below transformations:

$$\begin{aligned}
 B_J &= V_T + 0.00359667 + 0.655293 * (B_T - V_T) - 0.0429083 * (B_T - V_T)^2 \\
 V &= V_T - .09 * (B_T - V_T) \\
 R_F &= V_T - 0.0478440 - 0.577775 * (B_T - V_T) - 0.067737 * (B_T - V_T)^2 - \\
 &\quad 0.282895 * (B_T - V_T)^3 + 0.439489 * (B_T - V_T)^4 - 0.154056 * (B_T - V_T)^5
 \end{aligned}$$

Taken from a combination of Johnson - GSC2 transformations derived internally and Tycho - Johnson transformations published in the “Guide to the Tycho-2 Catalog” (Hog et al., <http://www.astro.ku.dk/~erik/Tycho-2/>).

For the SDSS magnitudes ( $g, r, i$ ) we use the following transformations derived internally and using the SDSS documentation at <http://www.sdss.org/dr5/algorithms/sdssUBVRITransformations.html>):

$$\begin{aligned}
 \text{Define :} \quad & gr = g - r + 0.093 + 0.166 \\
 B_J &= g + 0.0879778 + 0.24339 * gr - 0.0439303 * gr^2 \\
 V &= g - 0.5784 * (g - r) - 0.0038 \\
 R_F &= r - 0.178383 - 0.118655 * gr - 0.0194991 * gr^2 \\
 I_N &= i - 0.3780 * (i - z) - 0.3974
 \end{aligned}$$

For the Gaia G magnitudes we use either the SDSS-G transformations given in (Jordi 2007) or the following transformations in priority order derived internally for GSC23 magnitudes:

$$\begin{aligned}
 G &= R_F + 0.00751728 + 0.345256(B_J - R_F) - 0.230947(B_J - R_F)^2 \\
 &\quad + 0.283000(B_J - R_F)^3 - 0.115531(B_J - R_F)^4 \\
 G &= R_F - 0.0239712801 + 1.3159306049(R_F - I_N) - 1.4258466959(R_F - I_N)^2 \\
 &\quad + 0.4414260685(R_F - I_N)^3 - 0.0533889830(R_F - I_N)^4
 \end{aligned}$$