Highlights on Spanish Astrophysics X, Proceedings of the XIII Scientific Meeting of the Spanish Astronomical Society held on July 16 – 20, 2018, in Salamanca, Spain. B. Montesinos, A. Asensio Ramos, F. Buitrago, R. Schödel, E. Villaver, S. Pérez-Hoyos, I. Ordóñez-Etxeberria (eds.), 2019

Gaia source list evolution: Data Release 2 and beyond.

J. Castañeda[1], F. Torra[1], M. Clotet[1], J.J. González-Vidal[1], and C. Fabricius[1]

[1] Institut de Ciències del Cosmos, Universitat de Barcelona (IEEC-UB), C/Martí Franquès 1, E08028 Barcelona, Spain.

Abstract

The working catalogue of sources in *Gaia* evolves as new observational data is received from the spacecraft and it enters the iterative processing loop. The precision and accuracy of source parameters will improve with the addition of the new measurements but at the same time a change in source character and source identification can always occur as observations are added and linked to the already known sources. The source list is expected to stabilize in future *Gaia* data releases, specially once the spacecraft stops its operations, but some evolution of a substantial fraction of sources will take place up to the final data release (e.g., a stable source can turn into a variable from one data release to the next). In this paper, we describe the main causes that lead to the *Gaia* Data Release identification and major parameters updates, the nature of the updates and the best approach that the user can take to match the data of the current and subsequent data releases.

1 Introduction

The determination of the catalogue of sources from the processing of the observational data is an inherently complex issue in astronomy. For *Gaia* this issue is even more complex as new observations are received every day which must be linked to the sources so that further data reduction can take place. The process in charge of this task is the Cross Matching [6] which is executed at least once over all the accumulated observations for each scheduled *Gaia* Data Release (DR).

This work is organized as follows. In Section 2 we describe how the *Gaia* source list is determined, describing the main processes and techniques used and the source identification scheme adopted. In Section 3 we present some of the most common scenarios which trigger the updates on the source list and on the derived source attributes. Finally, Section 4 is devoted to summarize the major updates in the last published catalogue with respect the previous

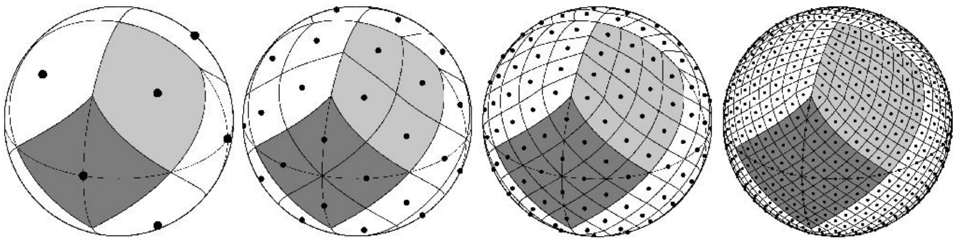

Figure 1: Orthographic view of *HealPix* partition of the sphere. From left to right the grid is hierarchically subdivided in subsequent level starting from 0 (Credit: [1]).

release, the expectations for the third release and the most important remarks regarding the tracing of sources across the published and future data releases.

2 Gaia Source list determination

The *Gaia* source list is determined just from the spacecraft detections, spacecraft attitude and orbit by means of classification and clustering techniques:

- Classification; mainly to censor the spurious detections [7].
- Clustering; to group all the detections from each individual object [6] [8].

The identified clusters of observations are then linked to the existing entries in the working source list. This working source list was initialized from on ground catalogues before *Gaia* launch and it has been updated progressively as new data has been processed. These changes are not limited to the source parameter updates - position, motion, parallax and magnitude among others - but also implies the addition of new entries and the removal of superseded or no longer matched entries.

To be able to track properly all these operations over the source list, all *Gaia* sources have been given unique source identifiers. This identifier is basically a numeric field assigned to each source to allow its unique identification and support its spatial arrangement. This numeric field basically codes a spatial *HealPix* index [1], the producer system and a sequence number. More specifically, the source identifiers have a level 12 *HealPix* index which is determined from the detections used in its creation time and it is encoded in the upper bits of the identifier.

The *HealPix* tessellation was adopted because of its mathematical properties and computation efficiency and its hierarchical numbering scheme is quite useful to speed up spatial queries or for massive job processing distribution. Fig. 1 shows the orthographic view of *HealPix* partition of the sphere.

Finally, it is worth pointing out that the *HealPix* index coded for a given source is never updated and discrepancies may appear when source position is updated. These discrepancies are more frequent in sources close to the the *HealPix* pixel boundaries.

3 Gaia Source list evolution

The Cross Matching task is executed periodically using the improved source parameters, spacecraft attitude, instrument calibrations and the updated censoring of spurious detections. Additionally, each new run starts from scratch ignoring any previous match solution so an independent new solution is obtained. Consequently, the new match solution may bring major updates in the observation assignations and the actual list of sources.

In this section, we describe some of the most common scenarios triggering updates of the source list by means of:

- New source creation: Fig. 2.
- Source list updates: Fig. 3 and 4.
- Source parameters updates: Fig. 5 and 6.

In the figures, we show the scenario and resolution in first *Gaia* Data Release (DR1) followed by the updated scenario and resolution in the second release (DR2). The symbols shown in the figures have the following meaning:

- Numbered circles represent observations from different scans; painted in blue the observations used in DR1, in yellow the DR1 observations affected by the updates on attitude and calibrations and in gray the observations classified as spurious detections and thus discarded in DR2.
- The stars represent source entries in the catalogue; in red the sources from DR1, in green the new sources in DR2 and in gray the superseded/deleted sources. Green stars with a red outline represents DR1 sources suffering major attributes updates.
- The black lines represent links between observation and sources.

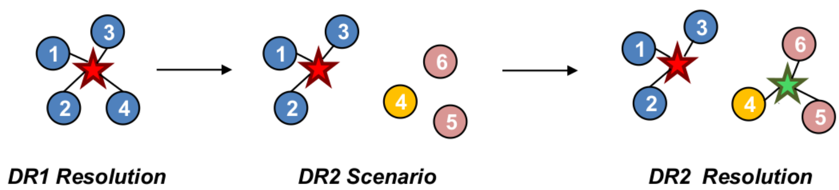

Figure 2: Example of new source creation. In this case the sky position of the observation belong to the fourth scan (4) used in DR1 has been updated and two new observations have entered the process. Due to the update of its position, the old observation (4) is no longer matched to the DR1 source and instead a new source is created together with the two new observations. As a consequence, the parameters of the source in DR1 will be slightly updated and a new source entry will appear in DR2. Legend in Sec. 3.

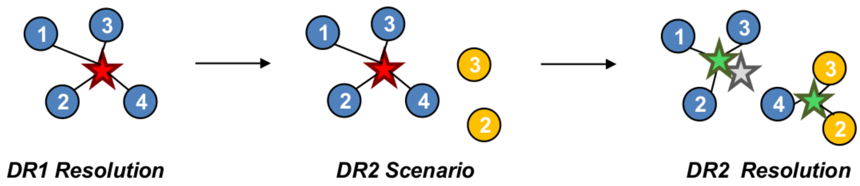

Figure 3: Example of new source creation by a split operation. In this case two observations from the same DR1 time period have been recovered i.e. reclassified as good detections. Consequently, for DR2 we have two scans with two nearby observations which is in general a clear indication of the presence of more than one source. In this situation, instead of creating just a new source and rearranging the observations, the policy in *Gaia* is to supersede the previous source and create two new entries so the evolution of all affected sources can be tracked across the different solutions. This split operation is only applied for observations and sources within a configured distance limit of ∼1-2 arcseconds. Legend in Sec. 3.

Figure 4: Example of new source creation by a merge operation. This case is basically the inverse operation of the split shown in Fig. 3. In DR1 we have two close observations in the same scan that triggered the creation of two sources. However, in DR2 one of those observations is no longer present i.e. discarded as spurious detections or suffering a large sky position update. In this situation, the policy in *Gaia* is to create a new source and supersede the previous sources with the same idea to be able to track the source list evolution across the different solutions. As in Fig. 3 case, this operation is limited to observations and sources within a configured distance limit of ∼1-2 arcseconds. Legend in Sec. 3.

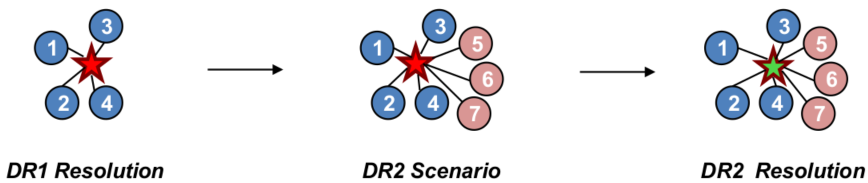

Figure 5: Example of source parameter updates. This example represents the most common scenario responsible of the source parameters updates across *Gaia* data releases. Each data release includes new observational data and/or new algorithms that revise and update previous solutions and the same source identifier can be present in DR1 and DR2 but the actual source features may be quite different. Legend in Sec. 3.

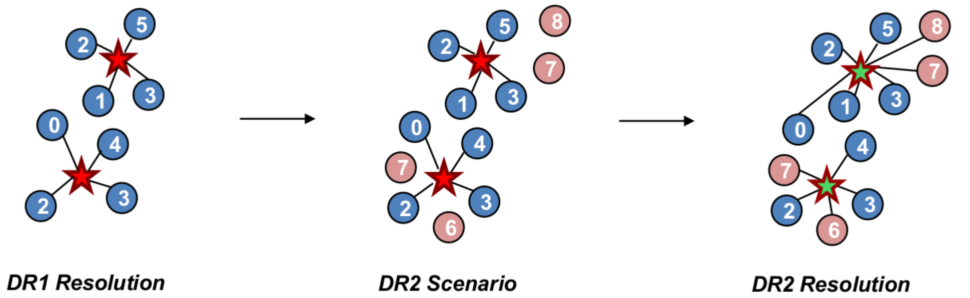

Figure 6: A more complex example of source parameters updates as the one shown in Fig. 5 with two sources involved. In this case we have new observations in DR2 that help the cross matching algorithm to resolve a source with a high proper motion (the one on the top right corner) which cause a rearrangement of the observation between the two existing sources. With this new match solution, the astrometry and photometry solution will be updated introducing a major revision on the source features. Legend in Sec. 3.

4 Conclusion

The working catalogue of sources in *Gaia* evolves as new observational data is received from the spacecraft and it enters the iterative processing loop. At the same time, the algorithms in charge of deriving the source parameters are also evolving taking advantage of a better knowledge of the instrument and the updates of the calibrations. Consequently, the source character and even the source identification are subject to change as observations are added and linked to the already known and published sources.

In 2016, the first *Gaia* catalogue (DR1 [3]) was published and the second release (DR2 [5]) has been recently made public this year. In DR2, ~25% of the DR1 source identifiers have been deleted and as a result ~75% have persisted. However, it must be mentioned that the attributes of ~70-80% of the sources brighter than 16^{th} magnitude that have persisted (whose source identifiers are still present in DR2) have been largely updated. These changes come from the updated censor of spurious detections and the rearrangement of the source matches procured by the clustering techniques implemented for DR2.

To provide support for the comparison of both releases, a table to trace sources from DR1 to DR2 has been provided in the *Gaia* Archive. This table can be used to identify which source identifiers from DR1 have changed or disappeared in the new release but also to identify possible replacements for the superseded or deleted sources.

In future releases, the source list is expected to become progressively more stable. In fact, source list changes between DR2 and DR3 are expected to be much less in terms of deleted/superseded sources:

- Just ~2-3% of DR2 source identifiers may be deleted in DR3.
- Less than ~4-5% of DR2 source identifiers may be superseded due to merge or split operations.

On the other hand, the source parameters may still show major updates and therefore the following remarks should be taken into account:

- Identifiers are not unique/maintained across different data releases.
- Never rely on a direct source identifier match or, even better, just treat the source identifiers in each release as being from completely different catalogues.

Acknowledgments

This work was supported by the MINECO (Spanish Ministry of Economy) - FEDER through grant ESP2016-80079-C2-1-R and MDM-2014-0369 of ICCUB (Unidad de Excelencia *María de Maeztu*)

References

[1] Górski,K. M., Hivon,E., Banday,A. J., et al. 2005, ApJ, 622, pp.759-771.

[2] Gaia Collaboration (Prusti et al.) 2016, A&A, 595, A1.

[3] Gaia Collaboration (Brown et al.) 2016, A&A, 595, A2.

[4] Fabricius, C., Bastian, U. Portell, J., et al. 2016, A&A, 595, A3.

[5] Gaia Collaboration (Prusti et al.) 2018, A&A, 616, A1.

[6] Clotet, M., González-Vidal, J.J., Castañeda, J., et al. 2016, in Highlights of Spanish Astrophysics IX, pp.634-639.

[7] Garralda, N., Fabricius, C., Castañeda, et al. 2016, in Highlights of Spanish Astrophysics IX, pp.646-651.

[8] Torra, F., Clotet, M., González-Vidal, J.J. et al. 2018, in Highlights of Spanish Astrophysics X (this volume).