Project Periodic Report

Grant agreement no: **606740**

Project acronym: **GENIUS**

Project full title: **"Gaia European Network for Improved data User Services"**

Funding scheme: SPA.2013.2.1-01 Exploitation of space science and exploration data.

Call FP7-SPACE-2013-1

Type of Action Collaborative project

Duration: 42 months

Final Period Report

Period covered: from 01/10/2015 to 31/03/2017

Project coordinator: Dr. Xavier Luri

University of Barcelona

Tel: (+34) 934039834

Fax: (+34) 934021133

E-mail: xluri@fqa.ub.edu

Project website address:

For the GENIUS project: `http://genius-euproject.eu`

For the general public: `http://www.gaiaverse.eu`

For internal network update: `http://gaia.am.ub.edu/twikigenius`

3.1 Publishable summary

3.1.1 Summary description of project context and objectives

GENIUS was designed to boost the impact of the Gaia astrometric mission, a European breakthrough in astrophysics. Gaia is an ESA Cornerstone mission launched in December 2013 and is producing the most accurate and complete 3D map of the Milky Way to date. A pan-European consortium named DPAC is in charge of the Gaia data processing, of which the final result is a catalogue and data archive containing more than one billion objects. The archive system containing the data products is located at the European Space Astronomy Centre (ESAC) and is already serving to the community the first product of the mission, the Gaia Data Release 1 (Gaia DR1).

The design, implementation, and operation of this archive is a task that ESA has opened up to participation from the European scientific community. GENIUS has significantly contributed to this development based on the following principles: an archive design driven by the needs of the user community; provision of exploitation tools to maximize the scientific return; ensuring the quality of the archive contents and the interoperability with existing and future astronomical archives (ESAC, ESO, ...); cooperation with the only other two astrometric missions in the world, Nano-JASMINE and JASMINE (Japan); and last but not least, the facilitation of outreach and academic activities to foster public interest in science in general and astronomy in particular. GENIUS has fit seamlessly into existing Gaia activities, exploiting the synergies with ongoing developments. Its members have actively participated in these ongoing tasks and have provided an in-depth knowledge of the mission as well as expertise in key development areas. Furthermore, GENIUS has had the support of DPAC, several Gaia national communities in the EU member states, and has fostered the cooperation with the Japanese astrometric missions already mentioned.

3.1.2 Work performed during the third year of the project

The work performed in the last period of the project covers four main areas (corresponding to the four main GENIUS work packages):

User community: the analysis of the archive requirements, from the general user needs to more specific analysis of several areas specifically targeted by GENIUS, has been applied to the actual archive design and deployment. The operational archive, available at `https://gea.esac.esa.int/archive/`, is based on this work. Furthermore, before the start of its operations it was tested by a group of beta-testers with the help of GENIUS, ensuring that the requirements were properly covered.

The cooperation with the nano Jasmine and Jasmine team has continued, with representatives of these missions participating in joint meetings, including a final GENIUS meeting in Japan. The installation of a mirror of the Gaia archive in Japan has been completed.

Archive system design: following the analysis and design work carried out in GENIUS during its first and second year, the features and tools developed are now part of the operational Gaia archive. Details can be found in section 3.2.2.3.2.

Tools for data exploitation: all the GENIUS developed VO tools are now fully operational and being used by the scientific community. This includes TopCat, VOSA and Clusterix, described in more detail in Section 3.2.2.4.2. The visualization system has been deployed at the Gaia archive, including a visualization server and an integrated visualization environment. This system has been intensively used by the scientific community during the first months of operations of the archive. Finally, the development of a data mining framework and tools has been completed, and is available at the CSUC partner, as well as a lightweight client for development in local computers. Large scale experiments are being run using the public Gaia DR1 data. The framework has been presented to ESA to consider the options of integration in the main archive for future releases. Details can be found in Section 3.2.2.4.2.

Validation: during this period the validation of the Gaia Data Release 1 has been completed. This was a mandatory task before the public release of data and the GENIUS contribution has been crucial for its success. The results of the validation have been published in Astronomy and Astrophysics. The validation framework has been completed and is fully operational, and is now being deployed for the validation of the upcoming Gaia Data Release 2. The tests developed by GENIUS are also being upgraded and will be used for this upcoming validation. More details can be found in Section 3.2.2.5.2.

3.1.3 Expected final results and their potential impact

As explained in the DOW, GENIUS results naturally lead to significant outreach into the public domain. Without a coordinated project like GENIUS, one risk is that Gaia would be just another specialised star catalogue (albeit an extremely precise one). The full potential of the 3D (6D) information can be realised only from the exploration and visualization tools developed within GENIUS, not from the Catalogue alone. Furthermore the impact on society goes beyond outreach only with, for example, surveillance activities. We are developing software to confirm and automate the alerts and combine ground-based with space-based Gaia data for detected Solar System objects, including the potentially hazardous Near-Earth Objects. Ephemerides for Solar System objects and the Celestial Reference Frame are other Gaia products that will also be used in areas far beyond specialised astrophysics. Needless to say, GENIUS was a pan-European project that fostered enhanced working relationships and collaboration between European research and higher education establishments, and as such continues to impact society in a fundamental and positive way.

Regarding economic impact, innovation in the use of Information Technology for research and development programmes often leads to mutually beneficial commercial partnerships. Development of enterprise-level Database Management Systems has benefited from scale-out deployment of billion-row astronomical datasets - e.g. UEDIN has in the recent past collaborated with Microsoft Research; in GENIUS WP3 we worked closely with ESAC and their commercial DBMS provider. Physical science research and development projects in IT often result in training of developers who subsequently go out into industry and commerce, with resulting economic benefits (WFAU at UEDIN has trained and/or employed developers who have gone on to employment within the commercial IT sector, e.g. Google).

Regarding education Gaia has already started to realise a 3D 'journey through our Galaxy' introducing astrophysics to a new generation of students and inspiring the next generation of researchers to enter the physical sciences. Tools for teaching purposes have been developed within WP4 - there are established precedents for Hipparcos which paved the way for Gaia. For postgraduate training, actions continue to be coordinated through GREAT `http://www.great-esf.eu`.

Finally, the scientific impact of Gaia, and therefore GENIUS, is clear. At the end of the nineteenth century, the first large international astronomical collaboration, the "Carte du Ciel", was conceived with the goal of providing "a legacy of the exact status of the sky at the end of the nineteen century". This massive project, which contributed to the origin of the International Astronomical Union, was the realization for sky maps of the potential power of photography, the new technology at that time.

One century later, Hipparcos, the European Gaia precursor, was the first experiment to use space technology for pinpointing the positions of (a limited) number of stars. Hipparcos had a significant impact on astrophysics, as assessed by the number of refereed publications derived from it which was in the range of 150 to 200 per year in the first years after the publication of its catalogue. One can expect that the Gaia impact will be much higher, given the larger number of objects and the additional types of data. Gaia operates on the same principles as Hipparcos: the measurement time of a star transit on the Gaia CCD is transformed into 1D epoch measurements, then into 2D thanks to the various scan orientations of the satellite, and finally into 3D information through the measurement of the parallactic motion of stars. In that sense Gaia represents an extraordinary means by which to convert time into space through its more than one billion star Catalogue, all the more so because Gaia has measured the velocity and the physical properties of the observed sources thereby increasing the dimensionality of the observables to more than 6.

Only time will truly tell, but it is already clear that Gaia represents the European legacy mission at the beginning of the twenty first century, being not simply an ESA cornerstone, but also a cornerstone in the historical quest to measure the size of the local universe, and the astrophysical record of its observable content.

GENIUS represents an essential part of the Gaia project, namely the dissemination of the results of the biggest astronomical survey to date (as a matter of fact, several surveys in one: astrometric, photometric and spectroscopic) to the scientific community and the general public. Since it is intended to provide and help visualise the results to the community, GENIUS represents the concrete and visible part of the huge work being undertaken by the 430+ European DPAC scientists and engineers, not to mention the work done by European industry. For this simple reason the impacts of the results from GENIUS are secured. Indeed, it is through the work undertaken within the GENIUS project that the full scientific potential of the Gaia catalogue and data archive will be unlocked. Hence GENIUS represents a clear and timely added value to the Gaia mission and data processing through various synergistic approaches:

- having gathered the different fields of expertise in the community to provide advanced requirements going well beyond usual queries to data archives;
- having played a significant role in distributing the data to the whole astronomical community thereby enhancing the visibility and impact of Gaia;
- having developed visualization and data mining tools to allow the most effective archive analysis;

- having combined Gaia with ground-based data, thus extending the interpretation capabilities across archives and wavelength domains.

Although the GENIUS project was focused on the Gaia data archive, the research and development within this project has benefited also other data archives, be they from space or ground-based experiments. Part of this benefit arises naturally through the push for interoperability with other archives, while the public dissemination of the GENIUS results has been used to enhance other existing archives or to prepare future data archives. Gaia is an European mission funded through ESA and with industrial partners in all ESA member states. Likewise DPAC is also a European effort which is funded through contributions from the various national funding agencies. The scale and complexity of the effort to bring Gaia into being necessitates this European approach. Likewise the science community has come together on a European scale in order to exploit the Gaia data. Examples are the GREAT Research Networking Programme and the Gaia-ESO survey `http://www.gaia-eso.eu/`, a European effort to gather complementary ground-based data. In the same spirit the effort to develop and deploy an advanced archive that will do justice to the exquisite data collected by Gaia can only be achieved by gathering together the relevant expertise, which no single institute or country harbours, from across Europe.

Furthermore, beyond Europe this project has cemented the Gaia collaboration with the only other astrometric missions in the world: the Japanese Nano-JASMINE and JASMINE missions, maximizing the synergies between the two projects and fostering the collaboration between two established space powers, Europe and Japan.

3.1.4 Project web site

In this last period the GENIUS portal has been a reference for the Gaia outreach, becoming a fully operational multilingual web site for the dissemination of Gaia and GENIUS activities: `http://www.gaiaverse.eu`. The GENIUS project web `http://genius-euproject.eu/` has been updated and linked to the Gaiaverse portal.

For internal project purposes (documentation, tracking and management) a site based on the cooperative Twiki environment was set up at `https://gaia.am.ub.es/Twiki/bin/view/GENIUS/`.

3.1.5 General contacts for the GENIUS project

Primary Coordinator Contact (GENIUS coordinator):

* Xavier Luri xluri@fqa.ub.edu

Other coordinator contacts:

* Lola Balaguer lbalaguer@fqa.ub.edu

* recerca.europea@ub.edu

* Xavier Gutierrez xgutierrez@fbg.ub.edu

3.1.6 Project partners

- Universitat de Barcelona (UB) - COORDINATOR Xavier Luri <xluri@fqa.ub.edu>
- Centre National de la Recherche Scientifique (CNRS) Frederic Arenou <frederic.arenou@obspm.fr
- The University of Edinburgh (UEDIN) Nigel Hambly <nch@roe.ac.uk>
- Universiteit Leiden (UL) Anthony Brown <brown@strw.leidenuniv.nl>
- Consorci de Serveis Universitaris de Catalunya (CSUC, former CESCA) Gorka Roldan <gorka.roldan@csuc.cat>
- Istituto Nazionale di Astrofisica (INAF) Ricky Smart <smart@oato.inaf.it>
- Instituto Nacional de Técnica Aeroespacial (INTA) Enrique Solano esm@cab.inta-csic.es
- Université de Geneve (UNIGE) Laurent Eyer <laurent.eyer@unige.ch>
- Université Libre de Bruxelles (ULB) Dimitri Pourbaix <pourbaix@astro.ulb.ac.be>
- Fundacao da Faculdade de Ciencias da Universidade de Lisboa (FFCUL) Andre Moitinho <andre@sim.ul.pt>
- University of Bristol (UBR) Mark Taylor <m.b.taylor@bristol.ac.uk>
- The Chancellor, Masters and Scholars of the University of Cambridge (UCAM) Nick Walton <naw@ast.cam.ac.uk>
- National University Corporation, Kyoto University (KU) Yoshiyuki Yamada <yamada@amesh.org

3.2 Core of the report: Project objectives, work progress and achievements, and project management

3.2.1 Project objectives for the period

The general objective of the GENIUS project was to contribute to the design and implementation of the Gaia archive, the key to the scientific exploitation of the Gaia data in the context of CU9. In this last period the Gaia archive is already operational, integrating GENIUS contributions. Therefore the objective for this final period was to complete all the deliveries and consolidate the GENIUS legacy for the Gaia archive.

- **Tailor to user needs**

  During the first year of GENIUS the work concentrated on the definition of (a priori) requirements for the archive and in the second year the requirements were being implemented in the archive. Now in this last year and a half of GENIUS work, the requirements have been implemented in the operational Gaia archive, available at `https://gea.esac.esa.int/archive/`. The goal has been to consolidate these requirements for the future evolution of the archive, integrating during the last phase of the project the experience gathered from beta testers and from real operations.

  Related milestones for the last period: MS9, MS12, MS16

- **Optimum archive system**

  Like in the above case the goal in this period has been to consolidate the previous design work on architecture issues and integrate it as much as possible into the actual operational archive. Furthermore, we also made progress in the design of some additional advanced features that cannot be integrated into the archive for the moment but that will be left as a GENIUS legacy for future upgrades.

  Related milestones for the last period: MS9, MS12, MS13, MS14, MS16

- **Tools for exploitation**

  The goal for GENIUS exploitation tools in this period has been to go from the prototyping stage to fully operational tools. In the case of visualisation tools and VO tools the goal was that they were available for the scientific exploitation of the first Gaia release (September 2016) and became part of this release ecosystem of data exploitation resources. In the case of the data-mining framework, the goal was to provide a fully functional system at the end of GENIUS, but the integration in the archive is not scheduled until a future third release. Thus, this framework will be a GENIUS legacy that will serve as the basis for the development of the Gaia data mining, in the archive but also in other environments.

  Related milestones for the last period: MS12, MS16

- **Validation**

  The goal in this period has been to apply the validation framework and validation tests designed and implemented in previous periods to the first Gaia data release. This application had already started with provisional versions of the data (the TGAS solution) but has now been applied to the full final dataset. Furthermore, this framework and tests will be the basis for the validation of future Gaia data releases, and in fact the preparation for these upcoming tasks is one of the goals reached at the end of the project.

  Related milestones for the last period: MS12, MS14, MS16

- **Dissemination**

  In this final period of GENIUS all the dissemination work, and in particular the Gaiaverse portal, is fully integrated into the activities of DPAC and ESA. Its application to the dissemination of the first Gaia data release was a key objective for this period, as well as the handover of the resources to the Gaia community.

  Related milestones for the last period: MS12, MS16, MS17

3.2.2 Work progress and achievements during the period

3.2.2.1 Work Package 1 Management

Lead Partner: UB

Contributing partners: -

3.2.2.1.1 Overview of WP objective

This package provided the overall administrative management of GENIUS.

3.2.2.1.2 Summary of progress made

This work package included the administrative tasks to fulfil the EC requirements and rules as well as the global administrative tasks inside the consortium, including financial management, intellectual property management and project documentation. These tasks were carried out by the GENIUS coordinator assisted by a hired project manager.

Task 1.1 - Global administrative tasks

Task leader: UB

As planned, the last period was marked by the publication of the first data release from Gaia, that is, the first public Gaia archive. The coordination work with the full DPAC consortium in general and the archive team at ESAC in particular was very intense. In particular the dual role of several of the GENIUS members, also acting as coordinators and task managers in DPAC, allowed a fluid coordination with the global mission activities.

Specifically in this work package, the coordinator Xavier Luri and the hired project manager Lola Balaguer continued the GENIUS management tasks: coordination of EC reporting, maintenance of the TWiki internal information system, management of the WebEx teleconferencing system and financial management of the project, as well as tracking of the project technical requirements and schedule. Monthly teleconferences have been held with the representatives of each node in order to track progress and two joint meetings with the full Gaia DPAC were organised: one in Leiden 2015, one in Sitges 2017, with an attendance of above 200 participants in each case. Both meetings were combined with GENIUS plenary meetings and served for the coordination of the GENIUS–DPAC activities and to carry out scheduled internal GENIUS reviews.

During this reporting period the external advisory board carried out the mid-term review (November 2015, milestone MS8) that provided an independent assessment of the project advancement from external experts. Also, the final external review took place on 27 March 2017 and together with the above mentioned meeting in Sitges covers milestone MS15.

Finally, the management in this period also covered the wrap-up of the project, including the final periodic report and the final review meeting.

All deliverables have been successfully produced with some minor delays in the delivery of the documentation.

3.2.2.1.3 Highlights

- Joint DPAC-GENIUS plenary meeting in Leiden, 16-20 November 2015 ((see `https://gaia.am.ub.es/Twiki/bin/view/GENIUS/PlenaryMeeting2015`)
- Completion of GENIUS External Advisory Board Mid-Term Review (see `https://gaia.am.ub.es/Twiki/bin/view/GENIUS/MidTermReviewMeeting2015`)
- Successful deployment of GENIUS contributions at the Gaia archive in the First Gaia Data Release 1, 14 September 2016
- Joint Gaia-Jasmine plenary meeting in Tokyo, 6-8 December 2016 ((see `http://www.jasmine-galaxy.org/meetings/gaia-jasmine.html`)
- Joint DPAC-GENIUS plenary meeting in Sitges, 23 - 27 January 2017 ((see `https://gaia.am.ub.es/Twiki/bin/view/GENIUS/CU9SitgesPlenary`)
- Final GENIUS External Advisory Board Review (see `https://gaia.am.ub.es/Twiki/bin/view/GENIUS/FinalReviewEAB`)
- Final project review meeting held in Brussels, 19th April 2017 (see `https://gaia.am.ub.es/Twiki/bin/view/GENIUS/FinalReviewMeeting2017`)
- Full report on dynamics of the gender balance on Gaia DPAC and the GENIUS project (Deliverable 1.11)
- All Deliverables and Milestones successfully achieved

3.2.2.1.4 Deviations and impact on tasks and resources

There have been no deviations in the administrative management tasks.

3.2.2.1.5 Use of resources

The following table lists the person-months per participant in the last period of the project in WP1[1].

	UB
WP1	10.4

[1] Provisional numbers until final report from financial offices of each node

3.2.2.2 Work Package 2: tailoring to the end user community

Lead Partner: UL

Contributing partners: INAF, FFCUL, UCAM, KU

3.2.2.2.1 Overview of WP objective

Unlocking the full potential of the Gaia catalogue and archive is not straightforward and will require an ambitious and innovative approach to data publication and access. A key aim of GENIUS was to ensure that the corresponding technical developments were driven by and focused on the scientific needs of the astronomical community that are using the Gaia catalogue. That is, the Gaia catalogue and data archive should be tailored to not only the needs of the scientific end user, but also the interested amateur or curious member of the general public. Tailoring was done by capturing the end users' scientific requirements and turning those into specifications on the basis of which the Gaia data archive, catalogue and data access methods were built. This issue was recognised by the Gaia community and a first round of requirements gathering amongst the scientific users was completed in 2012, coordinated by the Gaia Archive Preparations group (GAP). This process was non-trivial because of the often vague nature of the scientific requirements. It is easy to state that we want to compare a multi-billion particle N-body simulation to the entire Gaia catalogue but how would this be done in practice and what requirements does that place on the way the Gaia data is published and made accessible? In this work package these top-level requirements were analysed with the goal of turning them into detailed requirements. These requirements were cast in a language that both the scientists and the archive developers understood. The GAP requirements-gathering process revealed a number of advanced requirements (the Grand Challenges) that go well beyond the normal queries to data archives, and which required research in order to work them out in detail. Implementing these requirements could add very significant value to the Gaia data archive, while the expertise built up in this work package can be employed to enhance the value of other existing or future archives. The requirements for these Grand Challenges were researched in this work package.

3.2.2.2.2 Summary of progress made

Task 2.1 - Technical coordination

Task leader: UL

Brown managed WP2 through the supervision of the GENIUS postdocs working with him in Leiden and by staying in touch with the other WP2 contributors. Brown participated in the regular GENIUS teleconferences and also the GENIUS kick-off and plenary meetings.

Three GENIUS postdocs were hired in Leiden:

- Gráinne Costigan, May 5 2014 – April 30 2016
- Arkadiusz Hypki, September 1 2014 – November 30 2016
- Davide Massari, January 1 2017 – March 31 2017

Costigan worked on T2.2, T2.5, and T2.6, while Hypki worked mainly on T2.3. Massari worked on T2.2.

Task 2.2 - Analysis and working out of requirements gathered by GAP

Task leader: UL

Contributing partners: FFCUL, UCAM, KU

UL Contribution

Costigan and Walton (UCAM) coordinated the beta-testing of the Gaia archive ahead of Gaia DR1. This was done through regular teleconferences with the group of beta-testers during which the feedback was collected based on the experience of using the pre-release version of the archive. Although not all the issues uncovered could be addressed (mainly for lack of time), it was extremely useful to go through the beta-testing process. The infrastructure for the next round of beta-testing ahead of Gaia DR2 (April 2018) is now in place, with the testing planned to start much earlier.

Massari revisited the archive requirements with the experience of using Gaia DR1 in mind. He uncovered no missing requirements. The user requirements will be monitored continuously and adapted if needed in the light of the feedback received from users of Gaia data releases.

KU contribution

A discussion on the calibration effort for the Small-JASMINE mission was held with with European partners, Michael Biermann and Wolfgang Löffler.

The Gaia DR1 data products were made available to the Japanese community through the facilities available at the National Astronomical Observatory of Japan: `http://jvo.nao.ac.jp/portal/gaia.do`. A demonstration of the functionalities of the archive was provided during the Gaia-JASMINE meeting in December 2016.

KU took the lead in organizing a joint Gaia-JASMINE meeting in Tokyo from December 6 to 9 2016. During the meeting the GENIUS participants shared their experience from the data processing and the release process for Gaia DR1. This was followed by a discussion on the synergies between Gaia and the JASMINE missions and options for collaboration. One day during the meeting was dedicated to discussion of the prospects for collaboration between the Gaia/DPAC teams and the Japanese radio-astrometry community.

UCAM contribution

As part of the GENIUS science requirements update (see `http://great.ast.cam.ac.uk/Greatwiki/GaiaDataAccess`) the community was asked to provide additional usage scenarios detailing how they wished to use the GENIUS archive, and outline the functionality required.

An update requesting additional input was announced (in a presentation by Walton) at the GREAT plenary (see `http://great.ast.cam.ac.uk/Greatwiki/GreatMeet-PM9`) at the European Astronomical Society's 'European Week of Astronomy and Space Science' conference, held in Athens, Greece, July 2016.

The scenario collection period remains open after the end of GENIUS with analysis of the new scenarios to occur in advance of the development of the Gaia Access System enhancements scheduled to support the Gaia Data Release 2. This analysis will take place spring 2017 (reflecting the rescheduling of the Gaia Data Release 2 to April 2018).

The interim document 'Conclusion of requirements update gathering exercise' (D2.5) was delivered in November 2015. This will be updated in advance of Gaia DR2 as a DPAC CU9 deliverable.

Year 3/4 activity also involved assessment of alerts-related requirements, for implementation as part of the alerts testbed. See above on the contributions to the archive beta-testing.

FFCUL contribution

The FFCUL work was already completed by the time of the GENIUS second-year report.

Task 2.3 - Confronting complex models with complex catalogues

Task leader: UL

Hypki worked on an example implementation of an advanced access facility for the Gaia archive, the so-called BEANS package developed by himself. The corresponding paper was submitted (`https://arxiv.org/abs/1603.07342`) but has not yet been accepted for publication. Unfortunately this effort thus did not result in a finished product that might be included in future Gaia data releases.

Task 2.4 - Seamless data retrieval across archives and wavelength domains

Task leader: INAF

The activities for Task 2.4 were closely connected to CU9 cross-match activities in the near-infrared/optical/UV range and benefited from the developments in connection to that effort. A description of the Task 2.4 can be found in GAIA-C9-TN-ASDC-PM-011.

The multi-wavelength cross-match work was started in July 2015 when M. Fabrizio joined the INAF/ASDC/OATo group. He started from the radio domain by making a census of the existing and future radio surveys (GAIA-C9-TN-ASDC-MFA-001). The work was continued in the sub-millimetre/infrared and high-energy (X- and gamma-ray) domains. He prepared two additional technical notes (GAIA-C9-TN-ASDC-MFA-003, GAIA-C9-TN-ASDC-MFA-004). These notes describe the larger and more popular surveys available in the literature, in particular those for unresolved and/or compact sources. For each survey Fabrizio collected the relevant information about the instrument(s) adopted and the observed sources, especially those important for the cross-match to Gaia sources. In addition he provided a detailed description of a data model for every catalogue. During this period, he also made a census of existing cross-match techniques (GAIA-C9-TN-ASDC-MFA-002). Indeed, the goal of task 2.4 was to understand the requirements on the concept of seamless data retrieval across different archives, in particular combining Gaia data with both large and small catalogues in the different wavelength domains and the work Fabrizio did follows this track. The last part of Fabrizio's work was dedicated to the definition (requirements) of a prototype interface for the interactive cross-match tool for which the agreement is that it should be ready for the third Gaia release. Fabrizio prepared a technical note describing the web interface (GAIA-C9-TN-ASDC-MFA-005). At the end of June 2016 Fabrizio finished his one-year contract for Genius task 2.4 in ASDC.

Task 2.5 - The living archive

Costigan finalised and submitted a technical note on the requirements for implementing the concept of a 'living archive' for the Gaia mission (GAIA-C9-TN-LEI-GCO-003, corresponding to deliverable D2.6). The findings resulted in a recommendation (by the DPACE chair) not to pursue the living archive approach in the context of the Gaia mission. Note that some living-archive aspects are implemented in the Gaia archive, such as the sharing of user-provided/constructed tables and the availability of the VOSpace.

Task 2.6 - Reprocessing of archived (raw) data

Task leader: UL

Costigan worked on the gathering of requirements for the long-term archiving of the Gaia data products and the processing software. The goal would be to facilitate the (partial) reprocessing of raw or intermediate-level Gaia data, which would be aimed at improving the standard DPAC data products. Costigan contacted all the DPAC Coordination Unit leaders and several other experts within DPAC in order to get their inputs on what aspects of the Gaia data and processing software should be preserved for the longer term. The inputs were recorded in a technical note (GAIA-CU9-TN-LEI-GCO-006), which corresponds to deliverables D2.7 and D2.8. The technical note will serve as the basis for further preparations by the Gaia Mission Manager and DPACE chair for the long-term archiving of the Gaia data and the processing software.

3.2.2.2.3 Highlights

- The list of requirements for the Gaia archive was consolidated over the course of the GENIUS project and forms a solid basis for the further evolution of the Gaia archive and its data-access facilities.
- The realisation of the infrastructure for the beta-testing of the Gaia archive ahead of each data release.
- Precomputed cross-matches between Gaia DR1 and other large optical surveys, a facility that was very well received by the astronomical community.
- State-of-the-art cross-match algorithms were developed which will be used as part of the realisation of the cross-archive data-access facilities.
- The results from task 2.6 provide a solid basis for further preparations for the long-term archiving of the Gaia data and the processing software.

3.2.2.2.4 Deviations and impact on tasks and resources

None.

3.2.2.2.5 Use of resources

The following table lists the person-months per participant in the last period in WP2[2].

	UL	INAF	FFCUL	UCAM	KU
WP2	31.2	9	2	4	0

[2]Provisional numbers until final report from financial offices of each node. FFCUL: PMs not accounted for in previous report

3.2.2.3 Work Package 3: Aspects of archive system design

Lead Partner: UEDIN

Contributing partners: INAF

3.2.2.3.1 Overview of WP objective

The objective of this workpackage was to design, prototype and develop aspects of the archive infrastructure needed for the scientific exploitation of Gaia data. The design and technology choices were motivated by the real user requirements identified by WP2 – in particular, the massive, complex queries defined by the Grand Challenges – and by other initiatives, such as the GREAT project, and were made with full recognition of the constraints imposed by the ESAC archive system, with which the GENIUS deliverables must interface effectively. Prototypes were prepared and tested in cooperation with the end-user community and with the ESAC Science Archive Team through the DPAC CU9. A core principle was the adoption of Virtual Observatory standards and the development of the VO infrastructure required to enable ready interoperation with the other external datasets needed to release the full scientific potential of Gaia.

3.2.2.3.2 Summary of progress made

Task 3.1 - Technical coordination

Task leader: UEDIN

At the highest level, WP coordination continued via the usual collaboration tools (DPAC Wiki, SVN and Main Database Dictionary Tool) along with regular teleconferences in addition to face–to–face meetings and reflected the complete integration of the GENIUS activities within the wider DPAC CU9 developments. During the reporting period WP3 coordination contributions were made:

- to 6 DPAC CU9 teleconferences;
- to 2 internal review meetings (Leiden 2015 and Sitges 2017);
- and to several teleconferences and face–to–face meetings on aspects of archive architecture, e.g. data model for DR1 and DR2 and component-integration meetings with ESAC Science Data Centre (ESDC) developers

Coordination between GENIUS WP3, the wider DPAC CU9 WPs 920 (documentation), 930 (archive architecture), 970 (sub-WPs concerning data mining and cross matching) and ESDC was particularly valuable during the six month run–up to Gaia DR1 in September 2016. Although DR2 falls after GENIUS, these activities have continued recently as DPAC gears up for that release in one year's time.

Milestone achieved within this reporting period:

- (MS16) GENIUS products availability

Task 3.2 - Aspects of archive interface design

Task leader: UEDIN

The main activity during the reporting period was to adapt and integrate UEDIN's TAP/ADQL autocomplete library into the Gaia archive. As reported previously this end–user UI feature enhancement was deployed in the GENIUS archive prototype demonstrator. Work involved was recasting the software as a JavaScript library and changing the end–point metadata resource from VO Support Interface (VOSI) to TAP_SCHEMA. The developer concerned (Voutsinas) then spent a week at ESAC Science Data Centre working with their development team to integrate the software. Other activities included significant contributions to the Gaia DR1 Data Model, end–user documentation for presentation through the Gaia archive interface, and bug fixing and enhancements to the ADQL parser employed in the Gaia archive (and also archives at other European Data Centres).

Milestones achieved and Deliverables within this reporting period:

- (MS12) Prototype archive tools open to the user community[3]
- (D3.4) UEDIN TAP autocomplete:
  `https://github.com/stvoutsin/tap-autocomplete`

Task 3.3 - VO infrastructure

Task leader: INAF

Contributing partners: UEDIN

INAF IA2 completed the development on the IA2TAP Table Access Protocol implementation. Work has focused on smoothing the source code and adding a simpler script–based installation of the service. The KVM solution devised in the previous project period has been changed into a Docker contained one to have a quicker, and more portable, out-of-the-box TAP instantiation that requires only the external (MySQL) resource database for the data to be deployed. A TAP_SCHEMA Manager web application has been added to the set of tools (and is also included in the Docker release) to allow a graphical, easy-to-use, means for the user to populate the TAP_SCHEMA component at the core of the TAP services. This last application, among the other capabilities, features a UCD suggestion and validation dialogue as well as VO-Units validation.

UEDIN continued their work representing Gaia/GENIUS interests within various IVOA fora. The development team also specified the entire set of UCD1+ content descriptors for Gaia DR1, having previously worked on the required enhancements to the DPAC Data Modelling tool to enable

[3] see for example `http://genius.roe.ac.uk`

this to happen. As part of wider testing activities, Hambly wrote an ADQL 'cookbook' in support of GDR1[4] and as a culmination of these VO–related activities the team has written a formal VO compliance report for the Gaia archive in which we found the level of VO compliance to be excellent.

UEDIN made representations at the IVOA meeting in IVOA Stellenbosch, 8–13th May 2015 to discuss evolution of TAP, ADQL[5] and VOTable standards.

Milestones achieved and Deliverables within this reporting period:

- (D3.3) IVOA compliance document:
  `https://gaia.esac.esa.int/dpacsvn/DPAC/CU9/docs/WP930_Architectur` `nonECSS/TechNotes/GAIA-C9-TN-IFA-NCH-040/GAIA-C9-TN-IFA-NCH-040.pdf`
- (D3.4) IA2TAP deployed web services and code:
  `http://ia2.inaf.it/index.php/14-projects/39-eu-genius`
- (MS9) User prototype archive review
- (MS12) Prototype archive tools open to the user community[6]

Full documentation for these deliverables is available from the GENIUS Twiki[7]

Task 3.4 - Data Centre Collaboration

Task leader: UEDIN

Activities during the reporting period have consisted of finalising the Distributed Query Processing subsystem along with deployment in Docker containers (see next section) and extensive testing. The software has been containerised for ease of deployment in any Data Centre, not least ESDC for Gaia data. Presently, the software provides the underlying infrastructure in the GENIUS prototype demonstrator (see document deliverable D3.6). The report and analysis document (document deliverable D3.5) describes extensive testing of this system and also the Gaia archive for a wide range of queries harvested from real–world usage scenarios and ADQL queries gathered during data-centre operations.

Milestones achieved and Deliverables within this reporting period:

- (D3.5) Data Centre report and analysis document:
  `https://gaia.am.ub.es/Twiki/pub/GENIUS/DeliverablesGenius/Deliverable3_5.pdf`

[4] https://gaia.ac.uk/science/gaia-data-release-1/adql-cookbook
[5] https://github.com/stvoutsin/taplib
[6] see for example http://genius.roe.ac.uk
[7] https://gaia.am.ub.es/Twiki/pub/GENIUS/DeliverablesGenius/GENIUS_Deliverable_3.4.pdf

- (D3.6) TAP+ code and documentation:
  `https://gaia.am.ub.es/Twiki/pub/GENIUS/DeliverablesGenius/GENIUS_Deliverable_3.6.pdf`
- (MS13) Stress test[8]
- (MS14) Loading of actual Gaia data (see for example archive prototype linked under MS12 above)

Task 3.5 - Cloud–based research and data-mining environments

Task leader: UEDIN

At the time of writing of the GENIUS proposal (5 years ago) the state–of–the–art research environment was the CANFAR[9] system operated by the Canadian Astronomy Data Centre. As originally proposed task 3.7 in WP3 envisaged the somewhat ambitious plan of developing something similar for European astronomy and specifically aimed at 'grand challenges' for Gaia data. During the development programme it has become clear from investigations carried out within both WPs 3 (architecture) and 4 (data-mining tools) that bespoke systems are better designed around operating system level virtualisation as opposed to the more traditional hardware virtualisation, not least because OS–level virtualisation, a.k.a. containerisation, is more lightweight and hence efficient. In the IT world 'Docker' is emerging as the leading third–party system for containerisation. During the final reporting period we have extensively tested Docker in the context of use for deployment of astronomical software and demonstrated its use in this way by containerising two of our main deliverables as detailed above, namely the IA2TAP client–side VO publishing software, and the Distributed Query Processing infrastructural components. In addition to those SW deliverables we have written a detailed report of this experience for publication in Astronomy & Computing.

- (D3.7) VM research environment report:
  `https://gaia.am.ub.es/Twiki/pub/GENIUS/DeliverablesGenius/Deliverable3_7.pdf`

3.2.2.3.3 Deviations and impact on tasks and resources

Contributions to this WP were originally envisaged from INTA and CNRS. In the case of INTA the decision was made during the final year of the programme to devote that effort to WP4. In the case of CNRS, the work plan for WP3 concerned VO applications for studying solar system objects (SSO), but Gaia Data Release 1 contained no SSO data so the effort was expended elsewhere in WP5. As these pieces of work were independent there was no impact on any other WP3 task or their resources.

[8] see for example `https://gaia.am.ub.es/Twiki/pub/GENIUS/DeliverablesGenius/Deliverable3_5.pdf`

[9] `http://www.canfar.net`

3.2.2.3.4 Use of resources

The following table lists the person-months per participant in the last period in WP31[10].

	UEDIN	CNRS	INAF	INTA
WP3	21.4	0	8	0

[10]Provisional numbers until final report from financial offices of each node

3.2.2.4 Work Package 4 Tools for data exploitation

Lead Partner: UB

Contributing partners: INTA, FFCUL, UBR

3.2.2.4.1 Overview of WP objective

The overall goal of this work package was the development of powerful data-exploitation tools allowing for scientific use of the Gaia archive beyond the basic queries provided by the main archive interface. In this period the archive has become fully operational with the Gaia Data Release 1 and therefore the tools have moved from a prototype stage to full operational status. The final objective has been to leave them as a robust legacy of the project for the future use of the Gaia archive.

Visualisation tools

For the visualisation task the grand objective for this period was the implementation and testing of the interactive visual exploration web service running at the Gaia archive and ready for DR1. The service was adapted to the large size and complexity of the Gaia archive. In addition to the technological developments, the service produces visual contents, including images in specialised formats (e.g. HiPS) that are intelligible representations of the huge information content of the archive. The design has been based on the work done in the previous periods. Namely:

- Defining the list of requirements and feasible use cases to be covered by visualisation.
- Defining the architecture to support the visualisation requirements.
- Identifying open-source visualisation tools/libraries to be used or extended to support the graphical view of the Gaia archive.
- Defining the proper data models for the visualisation of the requirements.

Data mining

For this task the objective of this period was the completion and testing of the data-mining prototype for the archive. The Gaia Data Analytics Framework (GDAF) is the final result of the GENIUS work and constitutes an important part of the GENIUS legacy. It has been presented to ESA and its integration in the Gaia archive is being evaluated, as well as its installation in other data centres. The installation at CSUC is in use and the Gaia UB team plans to build on this infrastructure for its exploitation of the Gaia data.

VO tools

This task included the development of several tools working with the archive using the VO standards. One of them, TopCat, has been ready since the last reporting period, and the goal for this period was to complete and deliver other planned VO tools. This ensemble of VO tools constitutes another of the GENIUS legacies that is already being used for the exploitation of the Gaia data.

3.2.2.4.2 Summary of progress made

Task 4.1 - Technical coordination

Task leader: UB

The tasks in this work package (visualisation, data mining and VO tools) were mostly independent from each other. Therefore, the coordination by the UB has focused on tracking the progress of each of them with the respective coordinators in each institution (during the monthly teleconferences or in specific meetings), complemented with global updates during the general meetings. More detailed tracking is provided inside each task report in the following paragraphs.

Task 4.2 - Visualisation tools

Task leader: André Moitinho (FFCUL)

• Developed from scratch the Gaia Archive visualisation web service:
• Conceptual design and implementation of the web-based visualisation client, with mock-ups of the visualisation panels and user interfaces.
• Complete redesign of the server-side software: more scalable algorithms for computing visualisations and serving levels of detail at much higher speeds than the previous approach.
• Large effort expended in integration at ESAC. Many teleconferences with the GACS team at ESAC.
• Data ingestion and visualisation creation rehearsals test datasets (GUMS and IGSL).

The system was ready for DR1: it was deployed and made publicly available simultaneously with the opening of the archive for DR1.

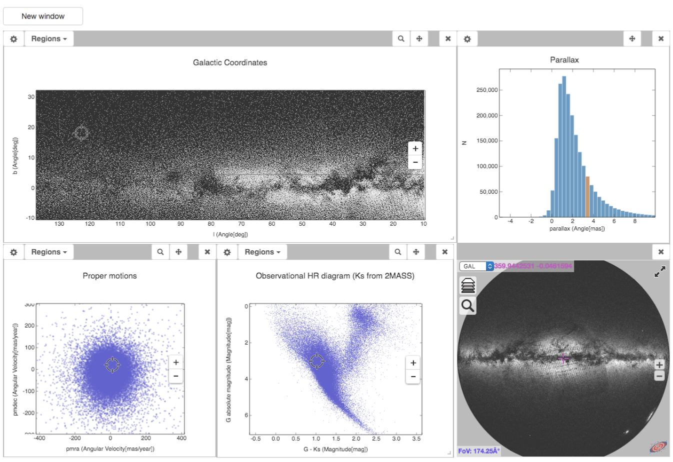

FIGURE 4.1: The Gaia visualisation portal, integrated with the main Gaia archive at ESAC

Several visualisations of the archive were made available, including scatter plots, histograms, FITS maps and HiPS maps (based on density and on luminosity). The DR1 sky map that symbolised DR1 was also produced in this WP. Since DR1, the visualisation software has been further improved and now provides extended features such as 2D binned plots, improved 3D scatter plots, and session saving and restoring. This new version has not yet been deployed at the Archive Portal. Finally, a Portuguese translation of the gaiaverse portal was created.

All items in the deliverable list were delivered:

GENIUS Deliverable 4.4. prototype of exploitation tools: Visualisation

GENIUS Deliverable 4.5. User Manual for prototype of exploitation tools: Visualisation

GENIUS Deliverable 4.6 Deployment of exploitation tools on the first actual Gaia archive: Visualisation

Milestone: System ready for DR1.

Task 4.3 - Data mining

Task leader: UB

Contributing partner: INTA

GDAF (Gaia Data Analytics Framework) is a data-mining framework proposed for use by the Gaia archive. During the last year of GENIUS activity we have formally defined the components of the system, their interactions and basic usage, together with some use cases and recipes. By the end of GENIUS, GDAF release version was v0.1, that is, a basic infrastructure with all functional systems well configured and some preliminary scientific use cases and a pool of scientists identified to act as a 'super-user' community to continue with the system definition. Tasks developed during the last year were focused on several working lines:

1. Data Lake

   As a Big Data platform it is essential to design a well-structured data-lake architecture, this is, how information will be available, how will it be stored and in what format, metadata availability, scalability, etc. In addition, an ETL (Extract Transform and Load) process is necessary in order to ingest data into the system from different sources. In our case, possible sources are the archive data themselves, but also external catalogues in ASCII or similar formats. An ETL tool has been developed to perform such conversion, clean-up and ingestion.

2. Resource manager

   YARN (Yet Another Resource Negotiator) is the resource-management layer in GDAF. It is configured to offer two different execution queues, high priority (with most of the available resources) and low (with less resources).

3. Big Data Engine

Apache Spark is the Big Data engine integrated in GDAF. It is currently the most efficient tool for performing distributed computing, and also offers interesting APIs to access and process data. Spark 2.1.0 is deployed in GDAF.

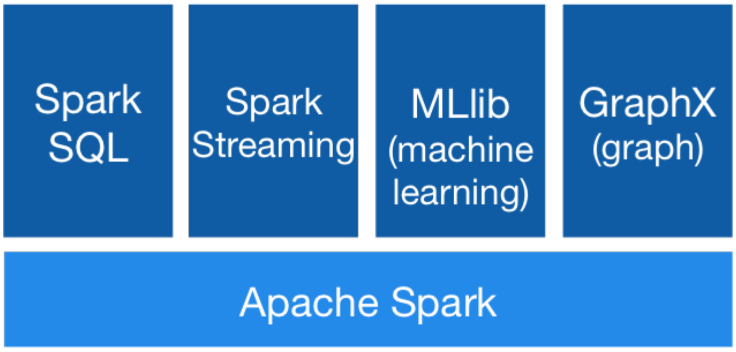

FIGURE 4.2: Spark API diagram

Among the available Spark APIs GDAF relies mainly on: Spark SQL, which allows the user to dynamically load, filter, transform and store data; and MLlib, where parallel implementations of Machine Learning techniques are available.

4. Distributed computing services

   Most of the features presented here rely on Apache Hadoop, a framework of services for distributed-processing capabilities. This framework is integrated in GDAF as a skeleton. GDAF v0.1 uses the 2.7.0 version.

Summarizing, the overall architecture is as follows:

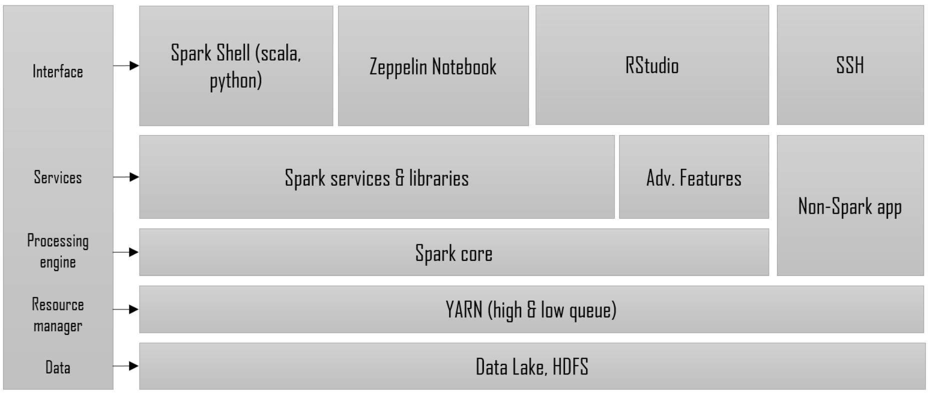

FIGURE 4.3: GDAF architecture

For interfacing with the system in GDAF v.0.1 Python and Scala shells are available. In addition, we have started to use Apache Zeppelin notebook and RStudio, although they have not been included in the formal release GDAF v0.1 as more testing and stability are necessary before making these options available for the tester community.

As stated previously Apache Spark provides a collection of machine-learning techniques such as several regression methods and unsupervised classification algorithms. However, external libraries have been integrated into the system such as SOM (Self Organizing Maps) and DBScan clustering to offer a more diverse utilities set.

As a debug environment, GDAF is also available in a Virtual Machine Image that can be executed using Oracle Virtual Box. That means that the user can interact with a miniaturised version of the system, developing code that can be executed later in a production environment. Both GDAF and VM version are synchronised in terms of software versions and data available.

Use case example

As an example, we developed a DBScan process on a canonical example, using both Scala and Python shells. Using a Scala shell (invoked as spark-shell) we load data and perform the distributed processing:

```
// Import necessary libraries
import org.apache.spark.mllib.clustering.dbscan.DBSCAN
import org.apache.spark.mllib.linalg.{Vector, Vectors}

// defining variables
val (src, dest, maxPointsPerPartition, eps, minPoints)=("/tmp/moons.txt","/tmp2",4,5,5)

// Loading data
var data = spark.sparkContext.textFile(src)

// transforming data to be consumed by DBScan
var parsed=data.map(s=>s.split(' ')).map(s=>Vectors.dense(s(1).toDouble,s(2).toDouble))
// Running the DBScan
var model = DBSCAN.train(parsed,eps = eps,minPoints = minPoints, maxPointsPerPartition = maxPointsPerPartition)

// Storing DBScan data
model.labeledPoints.map(p =>  s"${p.x},${p.y},${p.cluster}").saveAsTextFile("/dbscan-2")
```

FIGURE 4.4: Scala shell where DBScan is executed

In this figure, we see a brief recipe showing how to perform a DBScan on a canonical dataset (moons clustering shape). Once the clustering has been executed, results are stored in the '/dbscan-2' directory.

Once data have been saved they can be retrieved for plotting in a Spark python shell, making use of full Python plotting capabilities. Resulting clusters can be seen in figure 4.6:

The entire executing process takes only some seconds and moreover, the improvement versus traditional methods is focused on productivity. With a very few lines of code the process can be executed over a large amount of data. In fact, the size of the data to be processed is transparent to the user. It can range from a few MB to PB of data with the same code thanks to the Spark architecture and the scalability factor.

Conclusions and ways forward

For the coming releases we will stabilise Zeppelin as the preferred interface for the system. The astronomical community will access GDAF through a web notebook with all system features accessible from one single access point.

```
import numpy as np
import matplotlib.pyplot as plt
d=spark.read.csv("/dbscan-2")
d.registerTempTable("cloud")
d1=spark.sql("select _c0,_c1 from cloud where _c2=1")
d1.count()
x0=d1.select("_c0").collect()
y0=d1.select("_c1").collect()
plt.plot(x0, y0, 'ro')
d2=spark.sql("select _c0,_c1 from cloud where _c2=2")
x1=d2.select("_c0").collect()
y1=d2.select("_c1").collect()
plt.plot(x1, y1, 'bo')

d0=spark.sql("select * from cloud where _c2!=1 and _c2!=2")
x2=d0.select("_c0").collect()
y2=d0.select("_c1").collect()
plt.plot(x2, y2, 'go')
plt.savefig("/tmp/file.png", dpi=480)
```

FIGURE 4.5: Python shell where data is loaded for plotting

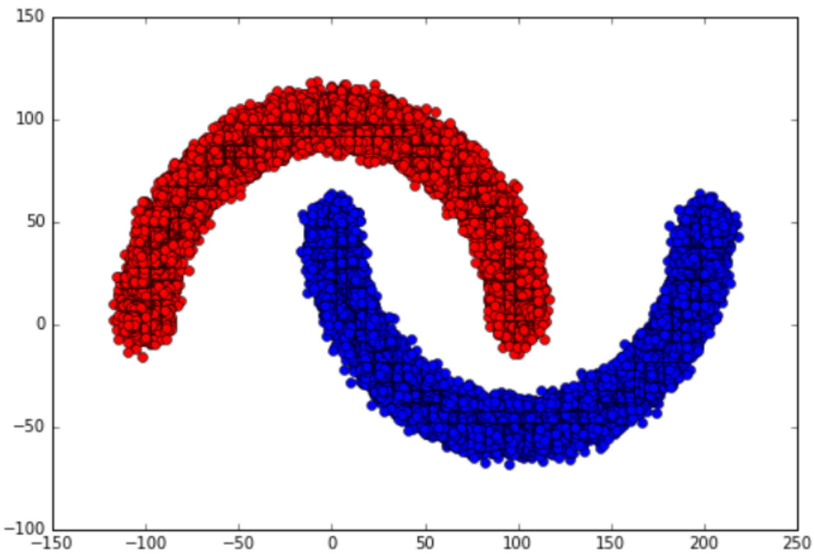

FIGURE 4.6: Python plot of the two clusters found

RStudio will also be available to write data processing pipelines on the archive using the R language. However, Zeppelin can also integrate R language into its programming, so RStudio GDAF integration has lower priority than Zeppelin.

GDAF will be a tool for the scientific community so more important than technical implementations is the use that astronomers will make of the tool and the added value to their research. With that in mind, for the following releases we will emphasise the scientific value of the platform.

Deployment on other Data Centres: the tool will be defined in such a way that it can be deployed in any other data centre. It is independent of the hardware infrastructure. It is highly scalable and no specific architecture developments have been done. The use of Docker will be evaluated in order to boost the deployment capabilities in other infrastructures. The exploration of this service has been done in the Architecture Work Package.

Task 4.4 - VO tools and services

Task leader: INTA

Contributing partners: UBR

<u>INTA contribution</u>

Jiménez-Esteban:

- Optimisation of the SVO Filter Profile Service. This service (used by VOSA) contains standardised, detailed descriptions of the most common photometric systems. A review of the documentation available for each of the filter sets, in particular concerning the units (energy or photons) and the components included in the filter definition (filter / detector / optics / atmosphere), has been carried out.
- Definition of a robust criterion to estimate the goodness of fit in VOSA in order to avoid the current need for visually inspection.

López del Fresno: Building of a new Virtual Observatory compliant tool (Clusterix, `http://clusterix.cab.inta-csic.es/`). Clusterix is a web-based application to calculate the grouping probability of a list of objects using proper motions. The tool also takes advantage of the Virtual Observatory to gather parallaxes, radial velocities and proper motions from VO services and to estimate temperatures, gravities and metallicities using VO tools like VOSA. Clusterix is a collaboration between INTA-CSIC and the University of Barcelona (UB) that is based on a previous joint effort between Marsaryk University (Czech Republic) and UB. Clusterix runs in a three-step process. First, the user selects the objects of interest by name/coordinates in a catalogue, in the Webda database or by uploading a file with positions and proper motions. In a second step the user chooses the "cluster" and "field" regions. Circles and polygons can be used as figures to represent those regions. Basic parameters for the membership determination can also be customised in this step. Finally Clusterix returns a table containing the list of objects selected in the previous step with the associated membership probability. This table can be sent, using the SAMP protocol, to any VO tool for its further analysis.

Regarding Clusterix, the list of tasks accomplished in the period January 2016 - March 2017 is as follows:

- Beginning of the design and development phase.
- Preparation of a preliminary prototype version for the progress meeting held in Barcelona on 2016 April 22.

- Preparation of an operational version for the XII Scientific Meeting of the Spanish Astronomical Society.
- Implementation of the "query-by-object" capability.
- Possibility of using circles or polygons to define the "cluster" and "field" regions.
- VO services update: APOGEE DR13 (radial velocities), PPMXL, and UCAC4 (proper motions) are now available from Clusterix.
- Implementation of new visualisation capabilities.
- Look and feel improvement.
- Code optimisation. Increase of the speed of searches in VO services.
- Bug identification and fixing.
- Scientific evaluation of the tool using both real and synthetic clusters.
- Generation of the associated documentation.

Non-refereed publications:

- *VOSA New Release*. Enrique Solano. IVOA Newsletter. December 2016 `http://ivoa.net/newsletter/016/`.

Outreach:

- Press release. *¿Cómo estudiar miles de millones de datos y no morir en el intento?* `http://www.cab.inta-csic.es/es/noticias/299/como-estudiar-miles-d`

UBR contribution

The whole of the effort allocated for UBR within GENIUS was performed during the first year. Taylor/UBR remains connected to GENIUS developments via involvement in Gaia CU9 funded separately from the GENIUS project; the remaining allocated GENIUS travel funds have assisted in communicating with the GENIUS project over the rest of its lifetime.

3.2.2.4.3 Highlights

- The visualisation portal for the Gaia archive has been developed from scratch with support from GENIUS. It is currently a unique service providing an environment for interactive visual exploration of a data set with over one billion objects with many dimensions. We know of no other service currently providing this in a scientific context.

The service is scalable, having been able to handle over 7500 users in the afternoon of the day of DR1. This was made possible by GENIUS. There is also a legacy here: the current service is expansible. It will be the platform that will host increased functionalities for future data releases (e.g. 3D, data overlays and more). The system has captured the interest of other collaborations and may be adopted by other missions and surveys.

The Gaia source density map has become the iconic image of Gaia Data Release 1 and was made possible by the GENIUS support. It was a worldwide success featured in all important media. It also has a legacy: it has inspired new ideas for visualising the next data releases. These ideas would not have appeared with this first stepping stone. Indeed, just a few months ago, it was not obvious how to transform a table of stellar data into an engaging image. We (the DPAC) all knew that Gaia is a great mission, but conveying that to the public without an image always required a lot of explanation. And conceiving an image was not straightforward: after all, Gaia observes in the optical band and the final product is a catalogue, not images of the sky, so there didn't seem to be much promise there. In the end, the image with the scanning stripes - highlighting the presence of the Gaia satellite - became a source of wonder and inspiration for the public and also for the developers. Finally, the GENIUS funding was crucial for consolidating the Lisbon team as a group recognised for the quality of their work in scientific visualisation.

- The TopCat developments for GENIUS have made it the tool of choice for the DPAC consortium. Being able to read the native gin format of the Gaia data, it is used all over the consortium. TopCat, along the rest of the tools, is now being used for the scientific exploitation of the Gaia data.
- GDAF is a very robust platform designed for Gaia that will be the basis for further developments leading to data-mining systems for Gaia, in the main archive and also in data centres hosting copies of the Gaia data

3.2.2.4.4 Deviations and impact on tasks and resources

To overcome administrative problems described in previous reports, and with the PO agreement, the INTA funds for this period were transferred to the UB; these funds were then used to hire the personnel to cover the INTA tasks in this work package and described above. These personnel have been working at INTA premises under the direction of Enrique Solano.

3.2.2.4.5 Use of resources

The following table lists the person-months per participant in the last period in WP4[11].

	UB	INTA	FFCUL	UBR
WP4	41.3	2	20.2	0

[11] Provisional numbers until final report from financial offices of each node

3.2.2.5 Work Package 5 Tools for data validation and analysis

Lead Partner: CNRS
Contributing partners: INTA, KU, FFCUL, UNIGE, ULB

3.2.2.5.1 Overview of WP objective

Despite the precautions taken during the acquisition of the satellite observations and when building the data-processing system, it is a difficult task to ensure perfect astrometric, photometric, spectroscopic, and classification data for a catalogue of one billion sources built from the intricate combination of many data items for each entry. However, several actions have been undertaken to ensure the quality of the Gaia Catalogue through both internal and external data-validation processes before each release.

This work package has produced such tools, based on the actual validation needs and on the characteristics of the archive system, thus making them as efficient as possible. Furthermore, the validation process relies on methods and tools that can also be used, with little or no adaptation, for the scientific analysis of the catalogue.

3.2.2.5.2 Summary of progress made

Task 5.1 - Technical coordination

Task leader: CNRS

Documents issued: Arenou et al. (FA-067), Arenou et al. (FA-070), Arenou et al. (FA-073), Shih (IS-006)

Benefiting from an organisational structure parallel to the DPAC CU9 one, this work package has put in place a technical coordination which first relies on the coordination of the CU9 validation sub-work packages themselves. For this purpose, regular teleconferences with sub-WP managers have been organised, one every about 1.5 months, the minutes of which can be found in the CU9 WP940 Wiki page[12]. The telecon meetings with validation sub-WP managers were organised as follows:

- Telecon meeting #15 with validation WP managers (2015-12-14)
- Telecon meeting #16 with validation WP managers (2016-02-08)
- Telecon meeting #17 with validation WP managers (2016-03-11)
- Telecon meeting #18 with validation WP managers (2016-04-22)
- Telecon meeting #19 with validation WP managers (2016-07-12)
- Telecon meeting #20 with validation WP managers (2016-11-15)

[12] http://wiki.cosmos.esa.int/gaia-dpac/index.php/WP940

- Telecon meeting #21 with validation WP managers (2016-12-19)
- Telecon meeting #22 with validation WP managers (2017-03-14)

In addition progress meetings were also organised within the sub work packages, sometimes very regularly (every one or two weeks) within Task 5.2 (CU9 WP942).

The validation group is too large to organise large teleconferences, and thus, plenary meetings were also organised once per cycle (every six months). These plenary meetings were organised as follows:

- Splinter meeting, and validation plenary #3 in Leiden DPAC meeting (2015-11-19)[13]
- Plenary validation meeting #4 in Paris (2016-05-26/27)[14]
- Splinter meeting, and validation plenary #5 in Barcelona DPAC meeting (2017-01-26)[15]

Some fundamental activities of the technical coordination were:

- to organise the input data, to run and deliver the tests done by the Work Packages. One key issue was ensuring the integration of the validation software at ESAC, which is needed as the software will not run on the Gaia data in distributed sites (non-disclosure restrictions on Gaia data).
- to maintain a common software environment (aka ValidationTools). The rationale for this is obviously that many tests are being implemented, and working on a more than a billion star Catalogue cannot (only) be done interactively. Thus that the data access, runs, configurations, and reporting should be done in a consistent way, so that the repetition of the validation tests on simulated data and then on successive data releases should run smoothly, while allowing the computing resources to be optimised as much as possible.
- to develop as much as possible the common tools which may be needed by the various WPs, mainly statistical ones (aka StatisticalTools).

Technically, this work package benefits from the structures put in place for CU9, and more generally DPAC, namely:

- The WP940 pages of the DPAC Wiki[16];

[13] https://wiki.cosmos.esa.int/gaia-dpac/index.php/DPAC_Consortium_Meeting:2015:Splinter_CU9_Validation_for_Gaia_DR1
[14] https://wiki.cosmos.esa.int/gaia-dpac/index.php/CU9:Plenary4_Validation
[15] https://wiki.cosmos.esa.int/gaia-dpac/index.php/DPAC_Consortium_Meeting:2017:Splinter:CU9_Validation_for_Gaia-DR2
[16] http://wiki.cosmos.esa.int/gaia-dpac/index.php/CU9:940

- The DPAC svn repository[17] for both code and documents;
- Hudson [18], which allows for checking the status of the JUnit tests of the validation after each software modification.
- Jira, for software bug tracking,
- Nexus library repository, all directories being defined in a homogeneous way (conventions on files, tests names, etc.) and configured to use Ivy, etc,

While the CU9 WP940 activity focused during the first year on the full definition of the validation tests, described in a Validation Test Specification (VTS) document[19], the role of this GENIUS work package was to develop the software implementing these tests. The current status of the test implementation (described in the following Tasks) is shown Figure 5.1.

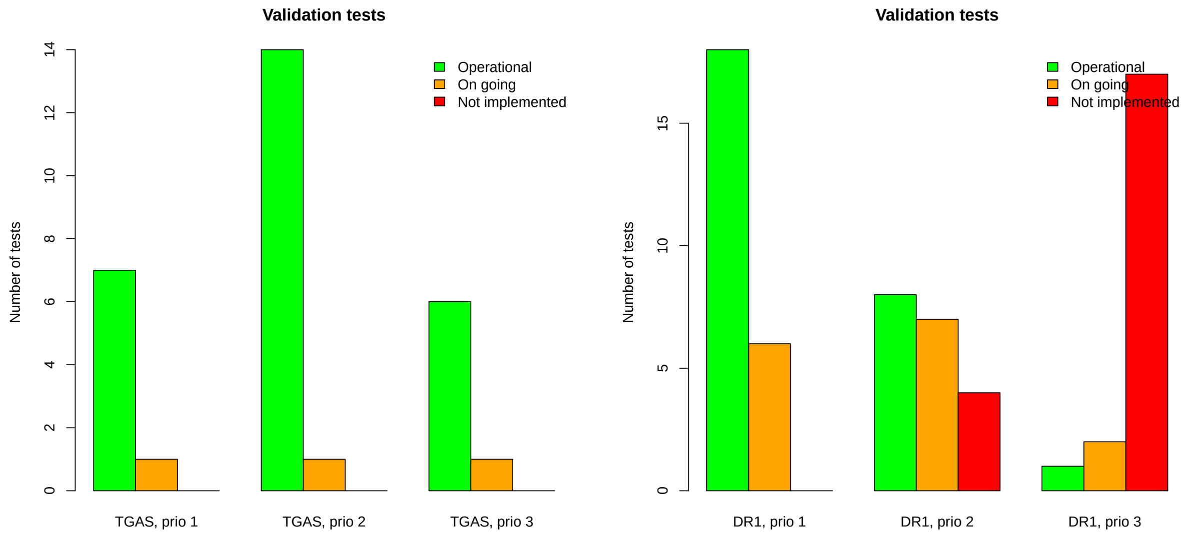

FIGURE 5.1: Software implementation of the validation tests, current situation for TGAS (with various priorities) and release I. Each test in a given release is also applied to the next ones, so that the tests indicated for release TGAS also apply e.g. to GRD2.

As indicated above, a common software environment has been developed, the ValidationTools project[20], which allows a consistent interface with the database or gbin files for all the validation users. In addition, a local platform with DB and TAP for tests has been set up in Meudon (CNRS).

In principle, this Task includes the coordination and supervision of the activities to be carried out. It was found that the coordination required the development of common software, and the responsibility of this Task was extended to the supervision of the overall framework allowing the tests to be run in a homogeneous way. A consequence of this in terms of FTE is that a fraction of the

[17] `http://gaia.esac.esa.int/dpacsvn/DPAC/CU9/`
[18] `http://gaiahud.esac.esa.int/view/CU9/`
[19] `http://gaia.esac.esa.int/dpacsvn/DPAC/CU9/docs/WP940_Validation/ECSS/VTS/`
[20] it is described in several documents, see Shih (IS-001), Shih (IS-004), Shih (IS-005)

CNRS involvement in each Task 1 to 6 has been devoted to putting in place this common environment which is used by the various Tasks. For example, within this coordination work, it was necessary to generate simulated data in order to test the software[21]. Then, thanks to the work done at ESAC, it was possible to apply the validation tests on preliminary Gaia data, the TGAS preliminary solution. The description of the tasks below and results applied on this solution illustrates the progress which has been made thanks to GENIUS/CU9 activity, through the development of the required software.

Task 5.2 - Looking for trouble: definition of problem cases, validation scenarios and tools

Task leader: CNRS

Documents issued: Arenou et al. (FA-067), Arenou et al. (FA-070), Arenou et al. (FA-073)

This work package defined validation scenarios, and implemented the corresponding tests, after basic verifications of the Catalogue content were performed to ensure that the field contents are as expected and that all fields are within valid ranges and fields present as indicated. Blind automated tools for fulfilling these tests have thus been developed in a close collaboration of CNRS with the UB.

Many tests have been implemented or refined during this last year, and this work package has undergone a lot of developments. The results are described in Arenou et al. (FA-073) and deliverable D5.9.

The test suite WP942-VAL-110 verified whether the data provided in the Gaia-DR1 release conformed to their specifications, either as explicitly stated in the CU9 DM or as implicitly understood by the validation group. Five of the tests failed, affecting up to a quarter of the catalogue.

The test suite WP942-VAL-120 verified whether the data from different fields provided in the TGAS release were mutually consistent, given expectations about relationships between fields. All but two tests failed, though the failure of WP942-VAL-120-020 was not considered a problem for Release 1.

The WP942-VAL-130 test suite verified whether the data in different versions of the Gaia catalogue agreed. The tests could not be completed due to technical difficulties.

The WP942-VAL-210 test suite verified whether the data provided in the TGAS release contained duplicate sources. One test failed due to apparent multiple detections of the same object.

The test suite WP942-VAL-310 verifies the consistency of the parameter distribution in the Gaia catalogue, over the whole sky. This package was responsible for verifying the distribution of the astrometric parameters and their median error distribution consistency, standard error distribution continuity as a function of magnitude at gate transitions, high proper motions and negative parallaxes ratio, and the dependence of the position error ratios on magnitudes, etc. As only TGAS sources from GDR1 (pre- and post-filtering) are considered five-parameter solutions while the other new Gaia sources only have two parameters solved for this release, all test cases defined

[21] described in Arenou & Michalik (FA-066)

have been executed for TGAS sources, but not all of them have been run for the complete Gaia catalogue (including TGAS as well).

The WP942-VAL-320 test suite verified the completeness in bright magnitudes in the TGAS data. Only one test, verifying completeness was implemented for GDR1 (pre- and post-filtering).

The WP942-VAL-340 test suite verified the accuracy of the errors reported in the TGAS data. Only one test, verifying parallax errors, was implemented for GDR1 (pre- and post-filtering). This test found that the parallax errors are accurate to within the limits of the test.

Task 5.3 - Simulation versus reality: from models to observables

Task leader: CNRS

Documents issued: Mor et al. (RMC-001), Arenou et al. (FA-070), Arenou et al. (FA-073) and deliverable D5.5 (delivery of model-based validation tools)

The aim of WP943 validation tests was to verify Gaia data through comparison with a data set generated from a realistic model of the Milky Way. The latter is provided by a population synthesis model of the Milky Way based on hypotheses from a probable scenario of formation and evolution of the Milky Way, and on stellar models, combined with empirical constraints and dynamical considerations. The tests are based on comparisons of different moments of the distributions over magnitude, colours, proper motions and parallaxes, and there variations as a function of latitude. Comparisons are done between model expectation and Gaia data. Validation tests for Gaia and TGAS (Gracia et al., GGA-038) data are described in detail in Arenou et al. (FA-061) and summarised in Houri Ziaeepour et al. (HZ-001).

The Besançon Galactic Model (BGM) simulates physical and dynamical properties of stars in the Milky Way in a self-consistent manner. In addition, it includes phenomenological models for extinction. The model is designed to generate fake stars and to predict their properties. But these sources do not correspond to real stars in the sky. Therefore, only statistical properties of sources in a region or the whole sky can be compared with observed catalogues. All the validation tests in WP943 (Task 5.3) are consequently based on comparisons of statistics of physical quantities. Several mock catalogues are generated by using various versions of BGM, namely GUMS, GOG, and BMGBGT.

As described in the D5.5 deliverable, we used two different reference models for comparison with data: BGMBTG_2_08 and the new BGMBTG_4.08.dat produced by Mor et al. (2015) from a revised Besancon Galaxy Model simulation with a modified kinematical model (see Mor et al, 2015 for details). It appeared that the first simulated catalogue (BGMBTG2) gives a better agreement with AGISL for the proper motions, while the second is more suited for parallaxes, due to improved scale lengths and extinction models. Hence, for testing proper motions we used BGMBTG2 and for parallaxes BGMBTG4. The results are described in deliverable D5.9.

Task 5.4 - Confronting Gaia to external archives

Task leader: CNRS

Contributing partner: INTA, KU

Deliverable D5.6 of GENIUS, Delivery of prototype of external validation tools
Documents issued: Ruiz-Dern et al. (LL:LRD-001), Arenou et al. (FA-070), Arenou et al. (FA-073) and deliverable D5.7 (delivery of external validation tools)

One of the first uses of the Gaia data is the cross-matching to external archives (e.g. to obtain the absolute luminosities in various wavelength ranges). Defining the tools to allow this is thus mandatory on the "scientific" side; on the "validation" side, however, cross-matching is of importance as it allows us to show the consistency between Gaia data and external data, and perform "external" validations.

The aim of the Task 5.4 tests was to check that Gaia data are coherent compared to other known catalogues, in order to determine to whether or not to accept them for publication. They concern several types of objects (simple stars, variables, binaries, multiples systems, quasars,...) as well as different kind of parameters (parallaxes, proper motions, radial velocities,...). The priorities have been selected depending on the information received on each Release.

To ensure that the tests are correctly working and return correct results, we tested them on simulated and real data. For that purpose we used the simulation of the *Tycho-Gaia* Astrometric Solution, the Hipparcos catalogue, and the Initial Gaia Source List (IGSL). These external validation tools that we have developed have then been applied to Gaia DR1 data. This is described in deliverable D5.7 and LL:FA-070, LL:FA-073 documents.

KU contribution

Work has been done to compare validation of JASMINE satellites data to Gaia data with the goal of helping both:

JASMINE will face small difficulties related to non-single objects and different bands:

- Either unresolved binaries: binaries (or larger multiplicity) with intermediate periods
  - Unresolved photocentre position is different in a different band
  - Motion of the photocentre is different too (different orbit scale)
- Or optical doubles with low separations (x-JASMINE angular resolution being smaller than Gaia)
  - In dense fields such as the Galactic bulge
  - Unresolved photocentre of blended stars with a different astrometry

Astrometry using a first epoch (nano-J) or a reference catalogue (small-J) with the risk of bad cross-matching for resolved, and low separation, binaries. Available Gaia simulations (within GOG) may help to characterise JASMINE properties in this respect.

Concerning validation tests, a large fraction of Gaia validation tests can be applied to JASMINE, e.g, for astrometry:

- Parallax zero-point: comparisons to other astrometry, photometry, etc.
- Parallax precision: internal (deconvolution), external (astro/photo) comparisons
- Correlations: measured with true or better astrometric or photometric reference

Task 5.5 - Data demining: outlier analysis

Task leader: CNRS

Contributing partner: FFCUL

Documents issued: Findeisen (KF-004), Antiche (ELA-026), Antiche & K.; Leclerc (GAIA-CU9-SP-UB-ELA-025) and deliverable D5.4 (delivery of statistical tools)

Outliers being by definition objects which deviate from an assumed model, it would be surprising if a mission such as Gaia planned for deciphering the complex structure of the Galaxy exhibited no outliers departing from our current knowledge.

The goal of this Task was thus to develop tools which will allow identification of outliers, or at least substructures which could then prove to actually be due to artefacts, not real structures. Two different developments approaches were followed.

To understand whether the statistical properties of the Gaia DR1 data set are consistent with expectations, we compared the distribution of the data (and in particular their degree of clustering) to suitable simulations for all 2D subspaces. In the case of TGAS, the comparison data is the simulation designated as ‘Simu-AGISLab-CS-DM18.3cor’, while for Gaia DR1 it is GOG18.

To this end, we used the Kullback-Leibler divergence (KLD)

$$p_{\mathrm{KLD}} = -\int \mathrm{d}^2 x p(\mathbf{x}) \log[p(\mathbf{x})/q(\mathbf{x})], \tag{5.1}$$

where $\mathbf{x}$ is a (sub)space of observables, $p(\mathbf{x})$ is the distribution of the observables in the data set, and $q(\mathbf{x})$ is a comparison distribution. When $q(\mathbf{x}) = \Pi_i p_i(x_i)$, i.e. the product of the marginalised 1D distribution of each of the observables, the KLD gives the mutual information. This expression shows that the mutual information is sensitive to clustering or correlations in the data set, with a high degree leading to high values, while in their absence p_{KLD} would be zero.

We thus computed p_{KLD} for more than 300 subspaces for the data, as well as for the simulations. In both cases, we used a range for the observables defined by the data after clipping the top and bottom regions by 3σ. Since the simulated and the observed data can have different distributions without this necessarily implying a problem in the data, we preferred to work with the relative mutual information rankings. If the structure is similar in data and simulations, we expect the rankings to cluster around the one-to-one line, while if a subspace shows very different rankings this would imply very different distributions. Such a subspace (or observable) is flagged for further inspection. This is important since the number of subspaces is very large.

The comparison to the simulations is sensitive to global issues (across the whole sky), while there could potentially be systematic problems in the data restricted to small localised regions of the sky. Therefore, we also compared the values of the mutual information obtained for different regions of the sky (e.g. symmetric with respect to the Galactic plane) and with a similar number of observations.

FFCUL contribution

The first panoramas of the Galaxy from GDR1 data were produced in May 2016; these were source-density maps. Denser regions are coded in brighter tones. Several resolutions and colour scales were tested in order to detect structures in the data that might require further inspection. This work relied on visualisation tools still in development and ended driving several adjustments and fixes in the software.

Thorough trials revealed an unexpected artefact. The Figure 5.2 is an image with a contrast/brightness enhanced zoom over the south equatorial pole. As can be seen, there is a very obvious effect of tiling in circular rings around the south pole. The tiles are visibly delineated by lines of higher density. It was found that it could not be an effect of the HEALpix tessellation used in determining the density of sources: it would not form rings and would not be almost circular like the one in the pole. Further analysis showed that it was an effect of the overlap of the IGSL Schmidt tiles used for the initial astrometric calibration. This would mean that due to the astrometric distortions in the positions of the IGSL sources, there are many non-cross-matches in the plate overlaps producing double or multiple counts.

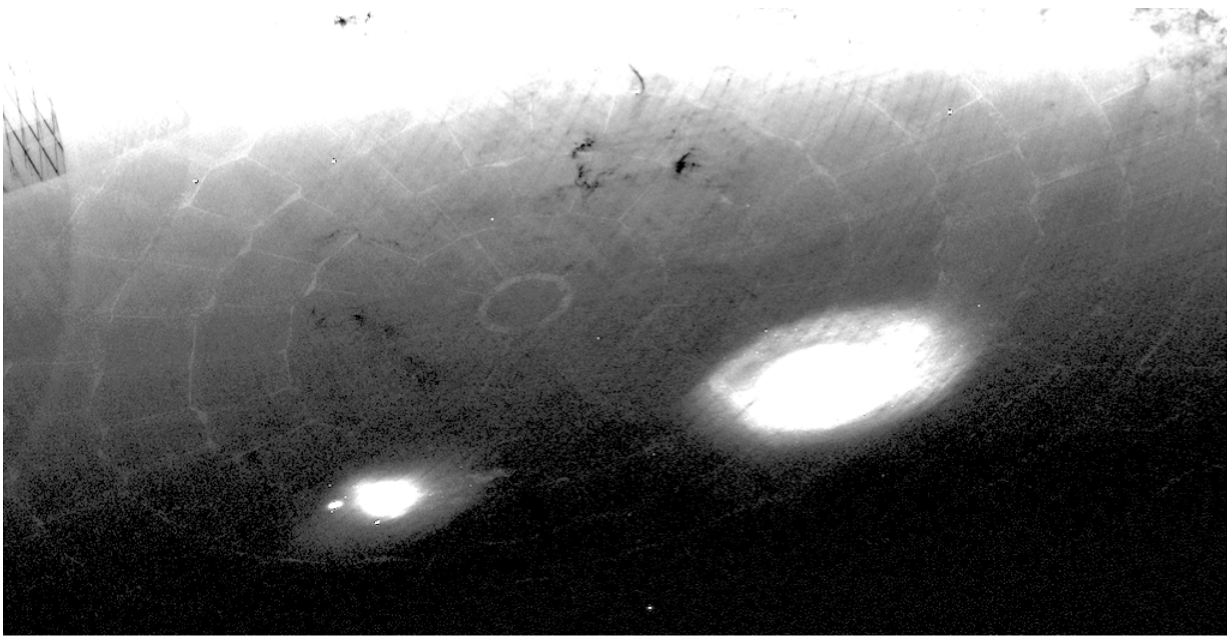

FIGURE 5.2: A contrast/brightness enhanced zoom over the south equatorial pole. See text for details.

Further analysis concluded that this multiplicity affected a number as high as 3% of the candidates in DR1. These findings triggered, in Validation, a filtering of multiple sources. The filtering performed well leading to a better DR1 free of the artefact. This was an important contribution to the role of validation in the data release.

Task 5.6 - Transversal tools for special objects

Task leader: CNRS

Contributing partners: UNIGE, ULB

Documents issued: Spoto et al. (FSP-001) and deliverable D5.8 (delivery of special-object tools)

Solar System Objects

Centuries of observation has permitted the construction of a large set of known SSOs. These are gathered in different databases accessible to the interested parties, e.g. at the Minor Planet Center (MPC) and astorb. What these archives typically provide is a set of osculating elements (Osc) which describe the trajectory of an object in space at a given epoch, namely a conic; this is usually an ellipse for an asteroid but can be a parabolic or hyperbolic orbit in the case of some SSOs (comets for example).

These Osc are determined by using observations and adjusting a conic to the data. Observations are performed continuously either by amateur astronomers, through dedicated programmes for SSOs or serendipitously, because when observing the sky there are frequently SSOs which appear in the images. Thus a continual flux of new observations is available for different objects so that it is possible to add these, little by little, into the existing database which is centralised at the MPC. As more observations are added for a given object the preliminary orbit can be improved. Thus the Osc are updated on a regular basis.

The Osc, usually given as the properties of a conic, viz. the eccentricity, the semi-major axis, the inclination of the orbital plane to some reference plane, the ascending node, the argument of perihelion and the mean anomaly can be transformed into a state-vector in the nominal reference system, i.e. three spatial coordinates and three velocity components at the given epoch. Thus the Osc, or the equivalent state-vector, should be understood to be the initial conditions of the SSO which can be used to integrate the equations of motion, i.e. the differential equations which theoretically govern the movement of the SSO in the Solar system. This then provides a way to predict the position of an object at some time in the future, and by symmetry to determine the position of an object in the past.

Validation of the Gaia SSO data starts with prediction of the positions of the SSO that are known from external databases and then comparing these positions with those resulting from the data reduction in the Gaia processing. From this comparison the six sets described previously can be assembled and then the relevant statistics can be computed. Additionally validation can be performed by comparing the dates of actual observations with those predicted, in particular to confirm that no observations are missing.

Stellar clusters

Another aspect of "special objects", the work on stellar clusters, has been done at Osservatorio Astronomico di Padova. Detailed completeness tests were run within globular clusters using HST data specifically reduced for the study of those crowded fields. The completeness was studied as a function of the stellar density observed in the HST data. Different crowded regions present different degrees of completeness, depending on the number of observations in that region. In addition, holes are found around bright stars (typically for $G < 11 - 12$ mag) and entire stripes are missing. In less crowded regions, such as in the field around NGC 5053 where stellar densities are under 1 million per square degree, the completeness is very high.

Parallax accuracy was also tested using distant clusters. Unfortunately, the small number of tracers available in this experiment did not allow us to draw a map of the bias by averaging values in coordinate space. The slight variation in zero point between a $l > 200^\circ$ and a $l < 200^\circ$ groups can then be interpreted either as random variations caused by the uncertainties on the reference values or as local variations of the parallax zero point (of the order of a few tenths of mas on a scale of several degrees).

A global zero-point test was performed from the Δ values obtained for individual clusters, restricting the sample to objects distant enough so that their internal dispersion in proper motions is negligible compared to the uncertainty on the proper motion of individual stars. The expected all-sky average of this quantity should be zero if no bias is present. A clustering test allows us to verify whether outliers are randomly distributed or clustered in problematic areas in the sky. No significant differences in proper motions were found.

Testing of G photometry using clusters was also done. The residuals clearly showed systematics at a 10 mmag level related to the presence of gates.

Multiple Stars

Document Arenou F. et al. (DP-055) reports all the issues and unexpected behaviour of the NSS scientific chain during OR5 stage 3. This analysis was carried out through the work package GENIUS WP560 multiple stars (for deliverable D5.8).

Concerning now the DR1 data: owing to the preliminary nature of the DR1 data release it was found that the effective angular resolution of the DR1 data (not the angular resolution of the Gaia instrument itself, which is as expected) is degraded, with a deficit of close double stars.

In addition to the above general tests, a specific test has also been done on known double and multiple systems from the Hipparcos new reduction (HIP2) and the TDSC in order to detect any possible bias between single and non-single stars. A difference in behaviour between those different subsets with respect to the *single-star* samples was looked for, using various parameters: the parallax and proper motion residuals (TGAS-external), and the TGAS errors, goodness of fit, and excess noise (source-modelling errors). Mainly acceleration solutions are expected to show large discrepancies between their proper motions in TGAS and those from Hipparcos or TDSC. Another source of discrepancy may be the fictitious difference created by the comparison of TGAS and Hipparcos proper motions for close systems for which only the photocentre was observed by

Hipparcos. For example, it was found that the excess noise, which is about 0.5 mas on average except for very bright stars, did not exhibit significant differences between single stars, primaries, and secondaries; on the contrary, unresolved systems had significantly degraded solutions with about 1.2 mas excess noise on average in the $7 \lesssim G \lesssim 12$ mag range.

Several other tests have also been done on secondary components, checking whether the separation or position angle with respect to the primary component had any adverse effect. In the past, during the validation of early preliminary Gaia data, it had been found that proper motions of many secondaries below $2''$ separation had a large discrepancy (up to a 80 mas/yr amplitude) compared to TDSC. Noting that $2''$ divided by the time span between Hipparcos and Gaia (2015-1991) gives about 80 mas/yr, it was deduced that the cross-matching of some close double stars had been deficient: most probably an incorrect first epoch position had been used for the Tycho-Gaia astrometric solution (TGAS), e.g. the Tycho position for the A component was associated with the observations of the B component because it was closer to it, depending on the position angle of the system, and vice versa.

Variable Stars

Gaia is particularly interesting for stellar variability studies since it provides a remarkable time-domain survey, which will help to better characterise the already known variables and will even detect new ones. GDR1 includes light curves for a selection of Cepheids and RR Lyrae stars. Several tests were developed to validate the data compared to ground-based surveys.

Additionally, objects with intrinsic or extrinsic variability may also affect the Gaia data analysis, cf. Eyer & Grenon (2000). For instance, the instrument and/or the data processing can also introduce false variability that might be interpreted as real. This aspect has been taken into consideration to implement a set of tests capable of verifying that no significant statistical biases are present in GDR1.

3.2.2.5.3 Highlights

The WP5 succeeded in developing a common framework, the rationale being to automate most tests. The access was standardised for all sub-WPs for both TAP interfacing or gbin flat files (for all “serial” tests when a too large fraction of rows would otherwise be returned with TAP). Thanks to the GENIUS help, about 130 000 source code lines were produced within the contract duration.

The WP5 made a Validation Test Specification (VTS) document describing all tests : objective, release, test data. Then it implemented these validation tests, ran them at ESAC or (sometimes) locally, initially using simulated data then the true data, generated a Validation Test Report (VTR) receiving the outcome of the tests done on data.

So, obviously, the main highlight of WP5 is the participation to the DR1 validation. In particular, WP5 has put into light a few problems before DR1, which lead to the data filtering before publication:

- of duplicate sources (36 745 480 sources have been marked for removal and 35 968 602 have been flagged as having been duplicated).

- of sources with too small number of G observations (746 292 sources have been marked for removal)
- of sources with obvious colour problems (164 446 sources marked for removal

Once all filtering had been done the final numbers for DR1 were the following: 1 180 112 861 sources before filtering, 1 142 679 769 after filtering (37 433 092 removed, 3.2%), and 35 951 041 have been flagged as formerly having duplicate sources associated to them.

Then, the main findings were the following. First, concerning the TGAS data:

- Astrometric accuracy
  - A global negative parallax zero point (about -0.04 mas) is consistently found with many independent estimation methods
  - Large scale spatial variations of the parallax zero-point may reach at least a 0.3 mas amplitude. Locally, much larger biases may be expected on a significant fraction of the sky. The bias is dependent on the number of observations, with colour and magnitude effects perhaps present
  - Concerning proper motions, significant differences with Hipparcos and Tycho-2 have been found which clearly originate from the external Catalogues
  - Proper motion of secondary components in close double systems are biased by up to several mas/yr,
- Astrometric precision
  - Overall, the formal precision appears correctly estimated
- Completeness
  - A large fraction of high proper motion stars are missing
  - Some stars matched as Tycho-2 objects are much too faint
  - A significant fraction of Tycho-2 stars are not in TGAS

Concerning the full Catalogue data:

- Photometry
  - There is an uneven distribution of magnitude which may be related partly to gate problems.
  - There is a significant number of photometric outliers with $G < 17$
  - In crowded areas there is also poor photometric precision, though not larger than expected, due to blending
- Completeness

  - Before filtering, duplicate sources was one of the main problems of DR1.
  - The ‘angular resolution’ of the DR1 Catalogue does not look better than the ground-based catalogues
  - The magnitude completeness in very dense regions is not worse than expected
  - The surrounding of bright stars does not show, apparently, too many spurious stars or ghosts
- Other
  - A few cosmetic errors were found in the data fields

3.2.2.5.4 Deviations and impact on tasks and resources

Initially, it was anticipated that Solar System Objects (SSO) could participate to the DR1 data release. For various reasons (too large detection threshold, spurious objects) the difficulty to discriminate between SSO and stars prohibited the run of SSO algorithms and consequently the validation of SSO data. The effort was thus brought on the orbit computation for DR2+ validations.

3.2.2.5.5 Use of resources

The following table lists the person-months per participant in the last year in WP5[22].

	CNRS	**INTA**	**UNIGE**	**ULB**	**FFCUL**	**KU**
WP5	29.3	0	0	0	2	1

[22]Provisional numbers until final report from financial offices of each node

3.2.2.6 Work Package 6: support activities

Lead Partner: UB

Contributing partners: CSUC, UCAM

3.2.2.6.1 Overview of WP objective

This work package aimed to provide support activities needed for the development of the tasks in the rest of WPs:

1. The provision of simulated data mimicking the actual Gaia catalogue; this mock-up data was be used for testing the system, from technical tests to user trials for validation.

2. The provision of a testbed for science alerts; the prototypes of the science alerts system was installed in this for testing and validation and made accessible to the test users.

Note: The contributions related to the development and implementation of the basic infrastructure for the community portal were moved to WP7.

3.2.2.6.2 Summary of progress made

Task 6.1 - Technical coordination Task leader: UB (J. Torra/E.Masana)

The two tasks in this WP were mostly independent and therefore the coordination has been centred on the tracking of progress through teleconferences, meetings and reports.

Task 6.2 - Simulated catalogue data

Task leader: CSUC (formerly CESCA)

Contributing partner: UB

In this reporting period two full simulations were generated using the GOG simulator. The main use of these simulations has been the validation of the Gaia DR1 data: they have been used for comparison with the real data to assess if the observed distribution of parameters followed the predictions from the galaxy models contained in the simulator. The first simulation was used during the preparatory validation work, carried out using Gaia preliminary data, and the second simulation, with an updated kinematic model requested by the validation team, was used for the final validation of the Gaia DR1 data.

The validation environment and the GOG simulator have been maintained during the project and will be available in the future for further Gaia releases through the Gaia UB team.

Task 6.3 - Science alerts testbed

Task leader: UCAM

Overview

The Gaia flux-based science-alerts stream is issued to the community through the science-alerts processing carried out at the Cambridge Photometric Data Processing Centre (DPCI). The science-alerts processing issues basic information for each flux alert via the VOEvent system to the community in a timely fashion (with alerts being produced 1-2 days after observation by Gaia). The alert packet contains basic characterisation information for each event, including parameters such as estimated alert object type, and more advanced classification for certain objects such as supernovae (SNe). For these, inherent Gaia photometric data is used to provide additional information concerning SNe alerts including class, epoch, redshift, reddening.

The testbed work carried out in WP-630 developed the interfaces required to connect the real time science alerts classification processing to the main Gaia data products. The first release of historical alerts will be made in Gaia Data Release 2[23]. Thus, as the mission evolves, and more knowledge is accumulated about objects measured by Gaia as it successively scans the sky, there will be opportunity to cross-reference new alerts against previous knowledge of that sky point as well as previous alerts against new information. For instance, irregular outburst events may show multiple times during the Gaia mission. Identification will be improved through correlation with earlier Gaia knowledge. The testbed provides linkages to external data resources provided through GENIUS, in particular via interfaces to the archive development through WP300. The alerts testbed plugs in to the portal testbed developed in WP720. With the termination of the GENIUS WP630 testbed activity, the full functionality will be deployed for community use - providing increasingly enhanced access to science alert data from the initial 2015 release.

Summary of progress in year 3/4

Year-3 activities included further requirements analysis. The testbed involves the integration of the real time alerts from the Gaia Alerts stream (from the Gaia/DPAC/CU5), for longer term duration within the Gaia Archive (developed through Gaia/DPAC/CU9 and GENIUS).

With the early operations of Gaia, the CU5 Gaia Alert stream was activated, with a validation phase running through to June 2015 (see `http://gaia.ac.uk/selected-gaia-science-alerts`). This was the initial Gaia Alerts testbed - *D6.2 Deployment of first pblic science alerts prototype*[24].

The second testbed, *D6.3 Deployment of second public science alerts prototype* was released as part of the full operational alert system in early 2016[25]. This followed a pause in operations of the photometric science alerts system (July-November 2015) to allow for the implementation of enhancements to the Alerts Pipeline, required to address issues arising in the alerts validation phase. For instance, the alerts pipeline is now more robust to false positive alerts resulting from spurious detections by Gaia around bright stars.

The third testbed, *D6.3 Deployment of third public science alerts prototype*, was released in Oct 2016. Figure 6.2 shows the alerts main page as of November 2016. This includes an improved suite of publishing tools. The Gaia Science Alerts Publisher was described in Delgado et al, 2016,

[23] See `http://www.cosmos.esa.int/web/gaia/release` for schedule
[24] See alerts release at `http://gaia.ac.uk/selected-gaia-science-alerts`
[25] See the alerts interface at `http://gsaweb.ast.cam.ac.uk/alerts/`

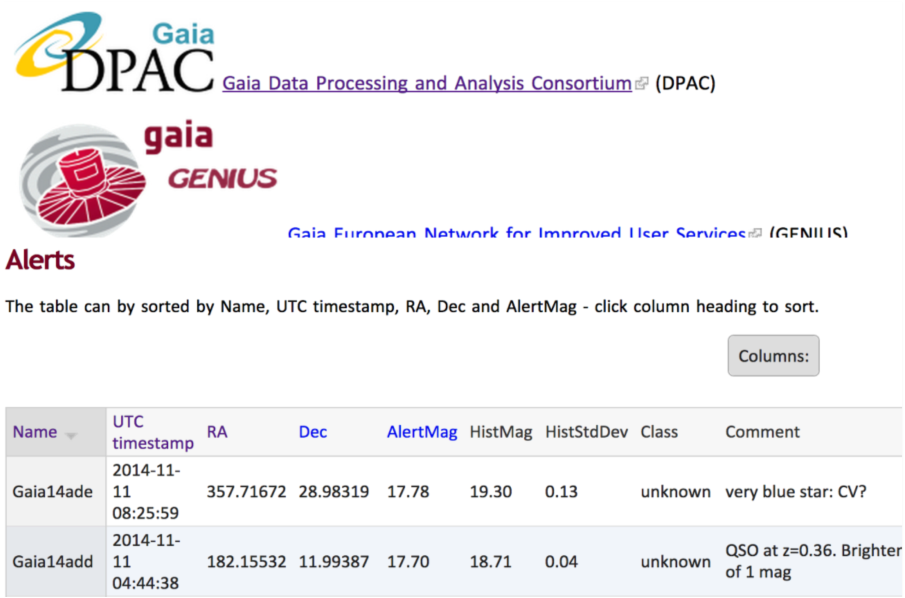

FIGURE 6.1: The Science Alerts portal acting as the initial testbed for the CU9/GENIUS alerts archive

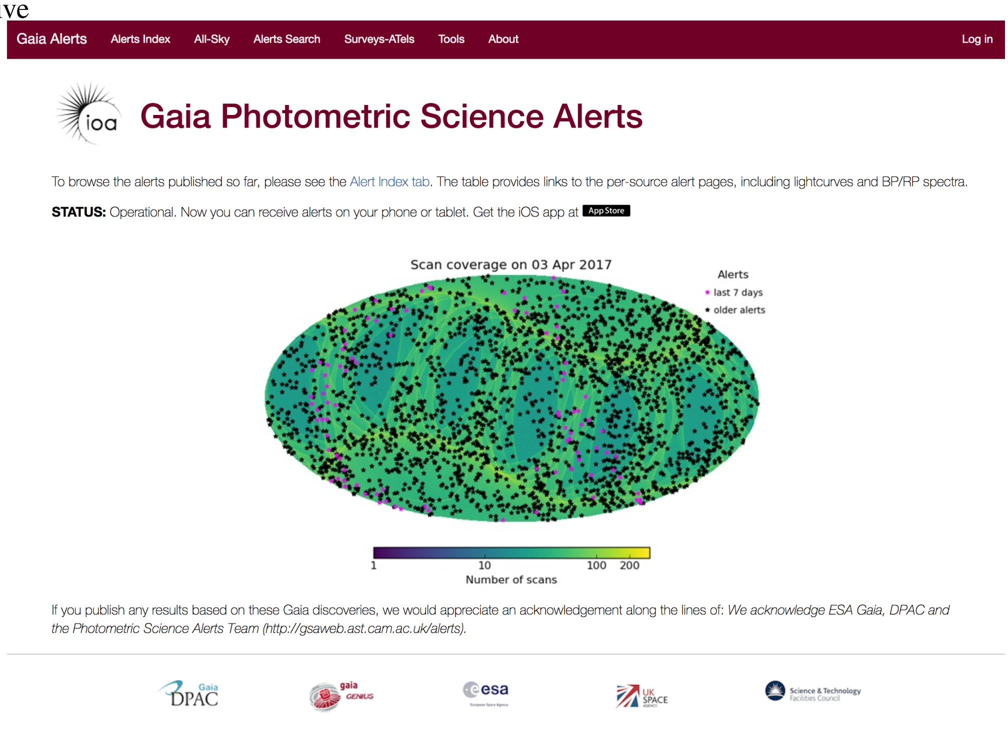

FIGURE 6.2: Deliverable D6.3 Release alert publishing page

PASP Proc ADASS XXVI.

Note that the Gaia Alerts are currently only available through the alerts testbed at `http://gsaweb.ast.cam.ac.uk/alerts/`. It is the intention to publish all alerts accumulated to date through the main Gaia archive at `http://archives.esac.esa.int/gaia/` commencing with Gaia DR3.

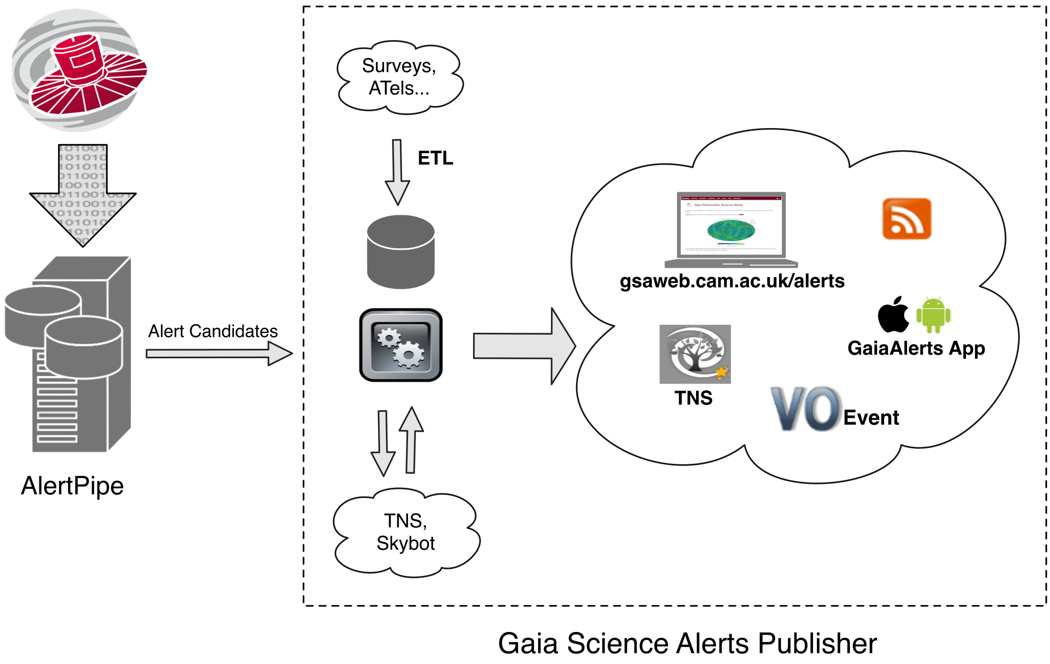

FIGURE 6.3: Gaia alerts publisher

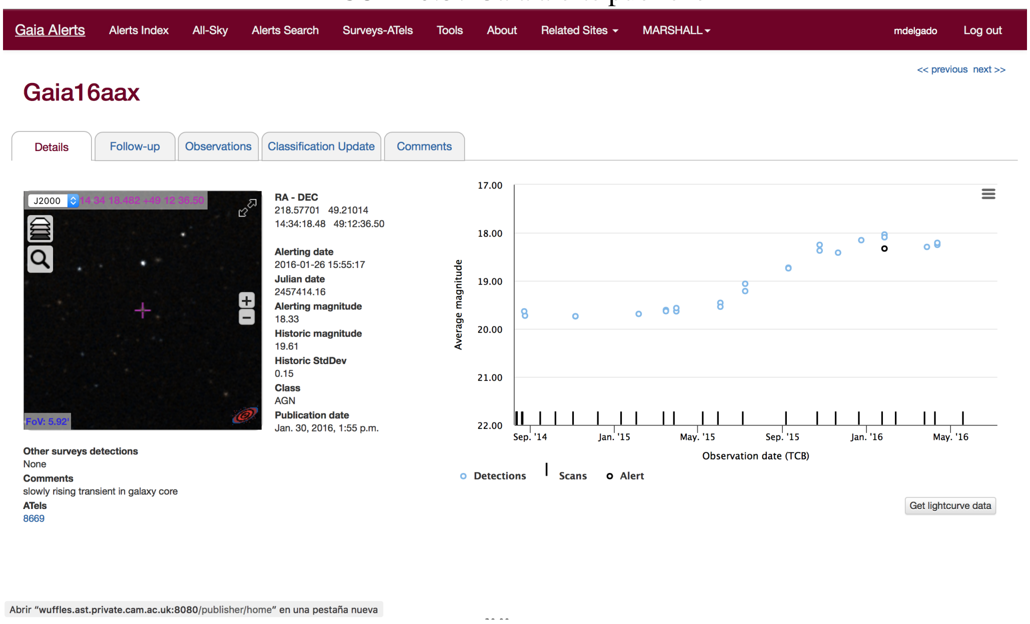

FIGURE 6.4: Alerts Marshall. This graphic shows presentation of data available for a particular published alert.

The alerts prototype has been releasing photometric alerts to the community from early 2016. Improved publishing of alerts has been enabled by the publisher, this shown diagrammatically in Figure6.3. Significant improvements have been undertaken in terms of functionality and usability of both the Alerts pages and the Alerts Marshall.

Figure 6.4 shows an example of the information that is presented for a Gaia alert candidate. This shows the Gaia available photometry for the candidate, together with the photometry at and post alert. Other tabs give access to information concerning follow-up observations of the alert, classification and user provide annotation and comments.

Figure 6.5 shows the summary page for the Gaia Marshall. This has been developed to support the ground based follow-up observational campaign. It gives information for all Gaia alerts requiring follow-up and classification. Other pages in the drop-down list give more detail on individual alerts, organised by priority, and also for those previously classified. The Gaia Marshall is now in active use. A significant follow-up up community is actively exploiting Gaia Alerts.

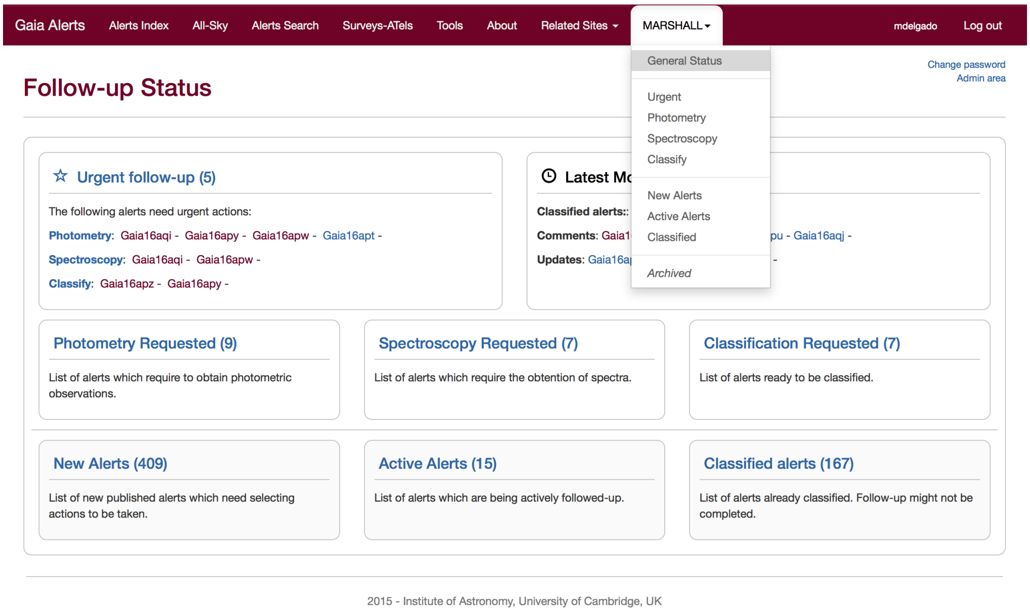

FIGURE 6.5: Gaia Alerts Marshall. This graphic shows the (auto-generated) list of alerts candidates awaiting classification (through the ground-based followup program).

The Alert testbed is described in the CU9 Software Development Plan (GAIA-C9-PL2-ESAC-WOM-086-01). In addition, the DPAC document GAIA-C5-TN-OU-RBG-001 - Proposed Alert Dissemination and Format for the Gaia Science Alerts - Publication System updated to note the interface from the CU5 alerts testbed to the main CU9 archive, and the specific GENIUS-supported elements of this.

3.2.2.6.3 Deviations and impact on tasks and resources

With PO agreement, funds initially budgeted for the acquisition of DataBase licenses were moved to the hiring of personnel at UB. These licenses were finally not needed because the DataBase technology initially envisaged (Intersystems Cache) was not finally selected for the archive. The final choice was PostGres, for which the CSUC already had licenses available. The budget was then used for hiring of a software engineer (M. Farras) that worked for some months in the development of the archive documentation system; the reason for this contribution was the urgent need for the development of this tool to ensure the availability of proper documentation for Gaia DR1. GENIUS that contributed to make this possible and the documentation system is now part of the archive and an additional GENIUS legacy for Gaia.

3.2.2.6.4 Use of resources

The following table lists the person-months per participant in the last period in WP6[26].

	UB	**CSUC**	**UCAM**
WP6	8.4	3.6	19.1

[26] Provisional numbers until final report from financial offices of each node

3.2.2.7 Work Package 7: dissemination

Lead Partner: CSUC (formerly CESCA)

Contributing partners: UB, CNRS

3.2.2.7.1 Overview of WP objective

The development and implementation of the basic infrastructure for the community portal (hardware, content-management system, design, etc.) and outreach activities.

3.2.2.7.2 Summary of progress made

Task 7.1 - Coordination of dissemination activities

Participants: UB, CNRS

The main activities of the WP7 in this period have been related to the preparation of the Gaia Data Release 1 (Gaia DR1) outreach activities, the addition of more languages (from 6 to 13) and the coordination of contents in general on Gaiaverse.

After releasing the Gaiaverse website and Twitter profile in 2015 July, 8th, the Editorial Board has coordinated activities to prepare the portal to make the Gaia catalogue available from the homepage, to provide information on DR1 and to decide on particular actions to cover the DR1 launch on September 14th.

The portal has been presented at the CU9 plenary meeting hosted by the University of Barcelona in 2015 September 21st-23rd, and at the GENIUS Sitges plenary meeting organised in 2017, January 25th.

At the Gaia DPAC Consortium Meeting #1, (Leiden, Netherlands, November 2015), we organised two sessions on science communication addressed to the participants. The sessions were conducted by Malcom Love from the UK. Each session had more than 30 participants. Furthermore a video explaining the Gaia mission and some important parts of the Gaia DR1 was recorded by K. Jäger and S. Jordan during the meeting, with the participation of several DPAC members.

In the framework of the Gaia DPAC Consortium Meeting #2 (Sitges, Spain, January 2017), a meeting was held to review outreach activities done and to discuss future plans for the hand-over of Gaiaverse to the community, beyond GENIUS. During the meeting in Sitges, two sessions were organised by Xavier Luri and conducted by Oriol Marimon and Helena González (Big Vang Science group) to improve scientific communication skills with the participation of 50 researchers.

Also, an oral presentation about Gaiverse portal, *Gaiaverse: the Gaia's outreach portal* was given by Eduard Masana at the XII Scientific Meeting of the Spanish Astronomical Society, Bilbao (Spain) in July 2016.

Not only during these congresses and conferences but also at several meetings, many contacts with

Gaia members have been done by the Gaiaverse Editorial Board to attract new collaborators to Gaiaverse and to include new languages.

Task 7.2 - Community portal infrastructure

Task leader: CSUC (formerly CESCA)

In order to prepare the portal for the first Gaia catalogue release, a new menu and submenu structure was added on Gaiaverse. It is "The Gaia Mission", having the following submenu options:

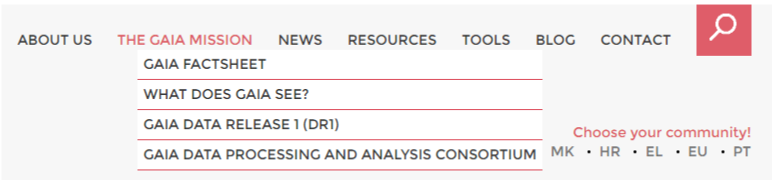

FIGURE 7.1: New section and subsections included for DR1.

Apart from that and coinciding with the release of the Gaia Archive, a direct link to the catalogue has been included also on the homepage. This link has been highlighted using one of the boxes used on Gaiaverse to provide a quick access to some pages such as resources or the Gaia Science Alerts.

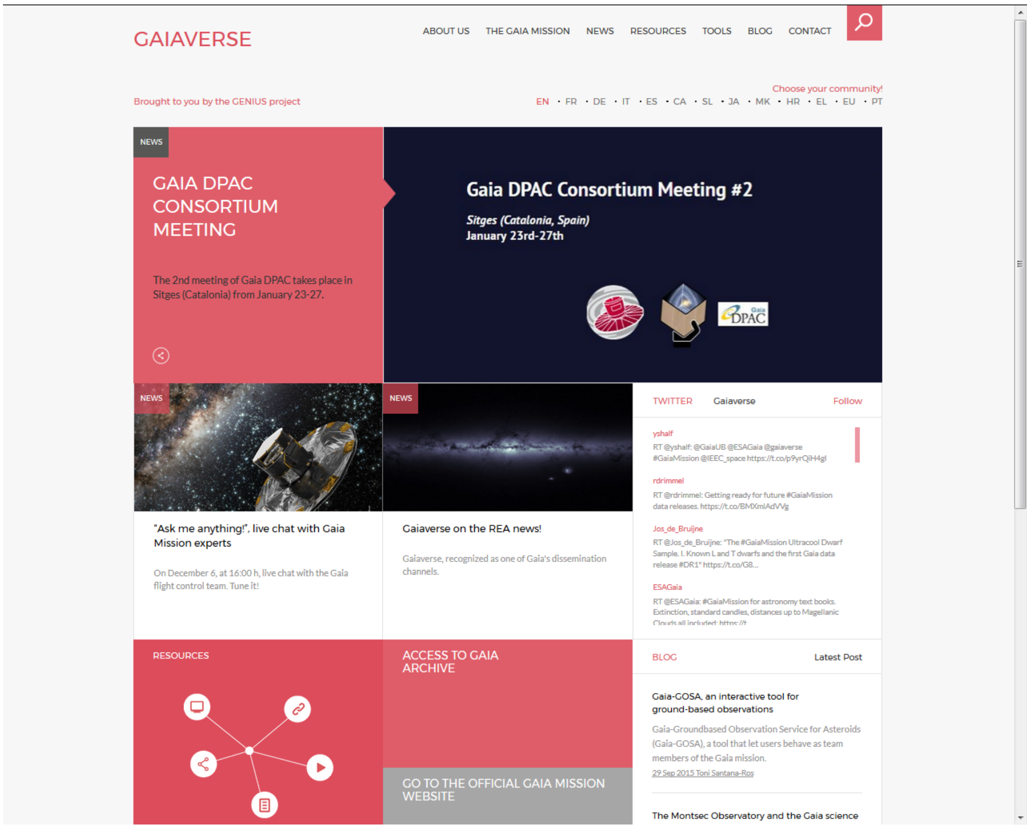

FIGURE 7.2: Access to Gaia Archive is provided from the homepage.

While preparing the portal for the DR1, some other modifications have been introduced on featured news on the homepage: menu adjustment, introduction of improvements on the “Filter” button in Resources, etc.

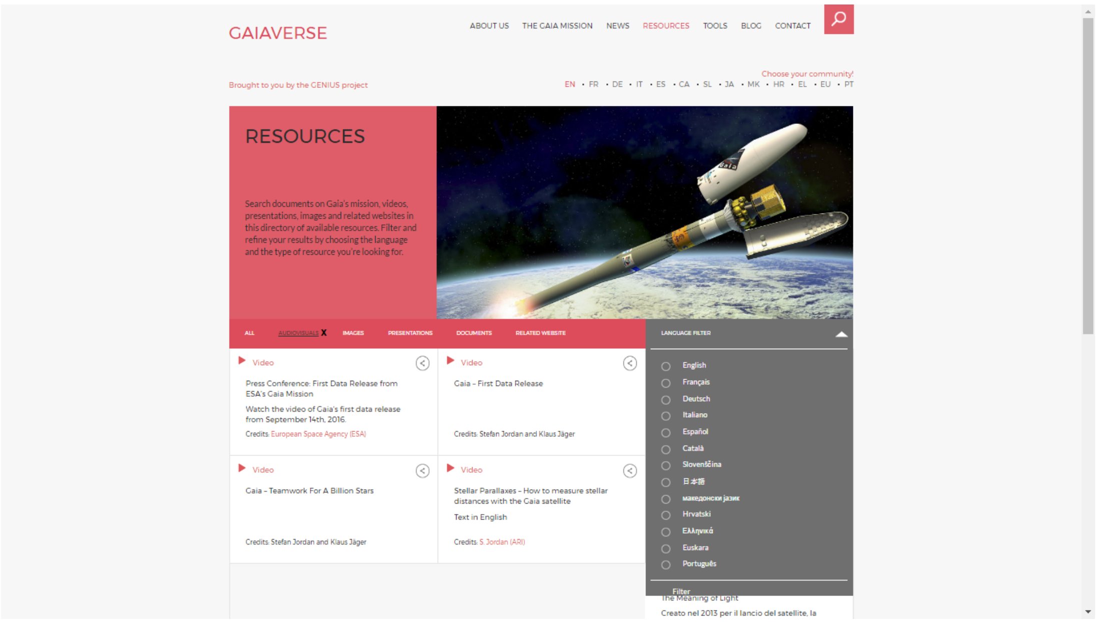

FIGURE 7.3: The filter functionality has been improved to allow more possibilities of search.

There have been some functionality changes inside the homepage and more flexibility has been introduced. Also, minor modifications and general style arrangements have been done.

During this period, more languages have been added: Slovenian, Japanese, Macedonian, Croatian, Greek, Portuguese and Euskera. In total, by the end of September 2016 the portal was available in 13 languages (7.4). In coordination with other project members, who became editors of the Gaiaverse portal as well, the work consisted of setting these languages into the platform by incorporating the translation packages distributed to them.

Task 7.3 - Community portal, outreach and academic activities

Task leader: UB

Contributing, partner: CNRS

One of the main tasks during this period has been the preparation of outreach activities for the Gaia Data Release 1. It is worth mentioning that, in terms of dissemination, September 2016 was a great month with regard to visibility thanks to the release of the first Gaia data catalogue. A video was used to promote the data release (even by ESA itself, notice the GENIUS credits all around and the proper credits to GENIUS/FP7 at the end). The all-sky image that became the icon of the data release was created and conceived by André Moitinho (GENIUS Portugal partner) and Màrcia Barros (GENIUS fellow). The GENIUS Gaiaverse site has been a node for multilingual, multicountry dissemination of the data release (available in 13 languages at the time of the DR1).

After launching the website and the Twitter profile, these platforms have helped to increase visibility of all the information, tools and resources generated within the GENIUS project and, in

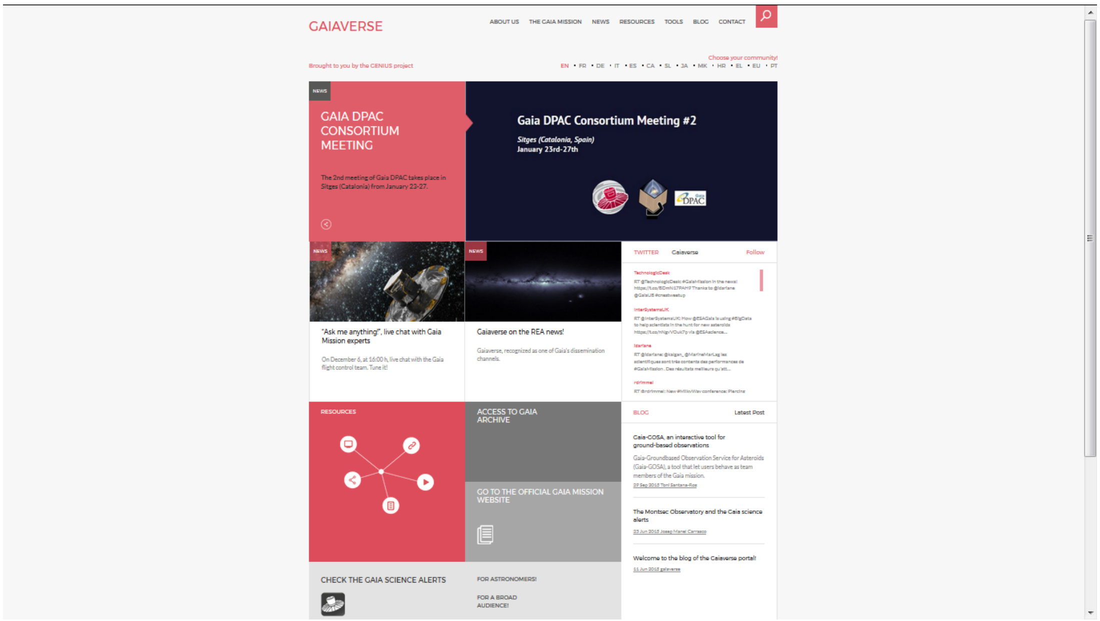

FIGURE 7.4: Seven languages have been included on Gaiaverse as well as a direct access to the Gaia archive.

general, the Gaia mission.

The Gaiaverse portal has been visited by more than 5,000 users during the period that goes from October 1st, 2015, to March 31st, 2017, of which 67.9% have been new visitors and 32.1% have been returning visitors. In general, users are aged between 25 and 44 years old, as well as 63.9% of these users are male and 36.1% are female.

Gaiaverse users have initiated more than 7,500 sessions, and have viewed more than 18,600 pages, which represents an average of 2.47 pages per session with an average duration of 2 minutes and 41 seconds.

Apart from the homepage, those pages that users have visited the most on Gaiaverse have been the resources page, the one dedicated to the Gaia Data Release (DR1), the news and the tools section.

Regarding traffic channels, 37% has been referral (links in other pages), 28% direct (bookmarks or typing the URL), 25% organic search (on search engines like Google) and 10% social (from Twitter, Facebook...). Indeed, from this 10% of social traffic, 85% has come from Twitter, 9% from Facebook, and the rest (6%) comes from other social networks such as Reddit.

Regarding social networks, the Twitter profile got more than 12,000 visits, of which 79% where male and 21% female (organic audience), mainly aged between 25 and 54 years old.

The number of followers has been rising progressively since October 2015, having at present more than 430 followers, of which 74% are male and 26% are female. The latter represents an increase of 5 points with respect to the number of women being targeted as Gaiaverse'public.

The engagement rate has been fluctuating between 0.60% and 1.70%, with the highest interaction of users in periods where the Gaia mission got important achievements, such as the first Gaia Data Release (DR1) in September 2016, or coinciding with follow-up meetings, workshops and so on. On average, the engagement rate has been of 1.06%.

During the first 24 hours (14-15 Sept.) of the DR1 release, the Twitter account had 10.4K impressions, described as the delivery of a post or tweet to an account's Twitter stream. The engagement rate was close to 2%, while the average number is 1%. On average, we earned 41 link clicks, 69 retweet and 29 likes per day. The same two days, from Sept. 14 to Sept. 15, the website www.gaiaverse.eu got 420 sessions initiated by more than 350 users, of which 22.6% were returning visitors and 77.4% were new visitors. They visited more than 1,000 pages (1,071 to be exact), that is an average of 2.55 pages per session.

3.2.2.7.3 Highlights

- Preparation of the final version of the Gaiaverse website.
- Coordination of outreach materials and activities for the Gaia DR1.
- Dissemination of news, tools and resources through the portal and social networks.
- Inclusion of additional languages (from 6 to 13 in this period).
- Sessions to improve the science communication skills addressed to DPAC members.
- Several outreach material, including a video about Gaia and DR1.

3.2.2.7.4 Deviations and impact on tasks and resources

There are no deviations to report and the tasks were developed according to the DOW.

3.2.2.7.5 Use of resources

The following table lists the person-months per participant in the last period in WP7[27].

	UB	**CSUC**	**CNRS**
WP7	3	8.9	0

[27] Provisional numbers until final report from financial offices of each node

3.2.3 Project management in the reporting period

3.2.3.1 Consortium management tasks and achievements

As described in the GENIUS proposal the project was, by necessity, tightly integrated into the previously existing structure of the Gaia Data Processing and Analysis Consortium (DPAC), and specifically into its Coordination Unit 9 (CU9) which is in charge of the development of the Gaia archive. The GENIUS coordinator Xavier Luri is also manager of the CU9, a combination that has facilitated and enhanced the developments in GENIUS for the archive.

Regarding the deliverables and milestones of the project, they have been included in the Participant Portal. A summary of status is provided in the GENIUS Twiki and in Sections 3.2.3.6 and 3.3:

- Deliverables:
  `https://gaia.am.ub.es/Twiki/bin/view/GENIUS/DeliverablesGenius`
- Milestones:
  `https://gaia.am.ub.es/Twiki/bin/view/GENIUS/MilestonesGenius`

3.2.3.2 Problems that occurred and solutions applied

During this period there was the signature of the amendment (Feb 2016) to correct the legal and administrative issues with one of the partners that led to a change of node from CSIC to INTA (but staffed by the same research group led by Enrique Solano). Since then we have been able to properly account for the work done by this group. The amendment also contained the change of name of CSUC (previously CESCA).

From January 2016 UB issued the contracts assigned to some of the CSIC tasks that have been successfully developed since then.

3.2.3.3 Changes in the consortium

As explained above, the CSIC withdrew from the project with date 01/10/2015, after justifying the expenses incurred in the second year (reduced to travel expenses). The amendment with the inclusion of INTA from the first day of the project to be able to include the work developed by Dr. Solano was finally signed in Feb. 2016. We also updated the change of name of CSUC (formerly CESCA) in this amendment.

We emphasise that all these changes have not affected the scientific content or results of the project as they are purely administrative issues.

3.2.3.4 List of project meetings, dates, venues and participants

Kind of meeting	Location	Date	Participants
Joint GENIUS DPAC Plenary Meeting + Mid Term Review Meeting	University of Leiden	16-20/11/2015	ALL (see complete list in webpage)+ EAB
Gaia CU9 Documentation meeting	ESAC (Madrid)	6-8/10/2015	M. Farras
Gaia Science Alerts Workshop	Liverpool (UK)	10-13/11/2015	A. Delgado, N. Walton
GENIUS/Modest meeting	Japan	7-11/12/2015	A. Hypki
Gaia CU9 coordination meeting	ESAC (Madrid)	26/01/2016	X. Luri
Data Mining meeting	ESAC (Madrid)	03/02/2016	F. Julbe
Gaia CU9 coordination meeting	ESAC (Madrid)	19/04/2016	X. Luri
WP4 New Tool: Clusterix	University of Barcelona	21-22/04/2016	D. Galadi-Enriquez, E. Solano, M. Lopez, C. Jordi, L. Balaguer-Nunez
XMM Newton: The Next Decade	Madrid	09-11/05/2016	X. Luri
IVOA in Stellenbosch	Stellenbosch (South Africa)	08-13/05/2016	S. Voutsinas, D. Morris
AGIS meeting (collaboration JASMINE - Gaia)	Lund (Sweden)	31/05-04/06/2016	Y. Yamada
IV Scientific Meeting of the Spanish Gaia Network	Barcelona	23-25/05/2016	E. Solano, F. Jimenez-Esteban, M. Lopez, D. Galadi-Enriquez, L. Balaguer, C. Jordi
Gaia CU9 Validation Meeting	Paris	26-27/05/2016	N. Walton
Genius Meeting to discuss integration of TAP autocomplete Javascript library	ESAC (Madrid)	5-10/07/2016	S. Voutsinas
XII Scientific Meeting of the Spanish Astronomical Society	Bilbao	18-22/07/2016	E. Solano
Cooperation Japan Europe (Gaia) with Michael Biermann and Wolfgang Loeffler	National Astronomical Observatory of Japan, Mitaka, Japan	01-05/08/2016	Y. Yamada
Gaia First Data Release DR1 presentation	ESAC (Madrid)	14/09/2016	X. Luri

Astronomical Data Analysis Software Systems (ADASS) XXVI meeting	Trieste	16-20/10/2016	M. Taylor, A. Delgado
ICCUB School - Machine Learning and Data Mining in Physics	Barcelona	16-22/10/2016	M. Barros
Gaia Data Release 1 Workshop	ESAC (Madrid)	2-4/11/2016	E. Solano, F. Jimenez-Esteban, M. Barros, N. Walton
Gaia CU9 meeting	ESAC (Madrid)	28/11/2016	X. Luri
Gaia DR2 Archive Requirements meeting	ESAC (Madrid)	30/11/2016	N. Walton
Gaia JASMINE Joint Meeting	Tokyo (Japan)	4-10/12/2016	A. Brown, X. Luri, F. Arenou, Y. Yamada, A. Mora, D. Hobbs, T.Prusti, W. Loeffler, U. Lammers, S. Hozumi, A. Nakagawa
Gaia CU5 plenary. Requirements and beta testing presentation.	Barcelona	13-14/12/2016	N. Walton
Cooperation Japan Europe (Gaia). ISAS, Sagamihara,	National Astronomical Observatory of Japan, Mitaka, Japan	15-17/12/2016	Y. Yamada
Cooperation NanoJssmine - Gaia Meeting	ISAS, Sagamihara, Japan	05-06/01/2017	Y. Yamada
Cooperation NanoJssmine - Gaia Meeting	The University of Tokyo, Bunkyo-ku, Japan	19-20/01/2017	Y. Yamada
Joint GENIUS DPAC Plenary Meeting	Sitges (Barcelona)	23-27/01/2017	ALL (see complete list in webpage)
Gaia CU9 WP980 - Visualisation Coordination Meeting,	Austria (Vienna)	30/01 - 02/02/2017	A. Moitinho
Final Review External Advisory Board meeting	Webex	27/03/2017	All+EAB
Final Review meeting	Brussels	19/04/2017	All

3.2.3.5 Project planning and status

As described in the GENIUS proposal (Section 1.3) the execution of the project was based on cyclic development where several prototypes were produced, each one building on the experience of the previous. In this third year this approach has converged into final versions of the GENIUS products that have been integrated into the developments of the Gaia data processing system and the main Gaia archive at ESAC.

In some areas, the developments have been closely tied to (and driven by) the main archive itself, and the real Gaia data. This is the case of the Architecture and Validation activities, where a close (and fruitful) collaboration with the Science Archives Team at ESAC was established, leading to the current archive open to the public with the Gaia Data Release 1 contents. Furthermore these data contents were submitted to an intensive validation with significant contributions from GENIUS.

The GENIUS visualisation system, although not directly tied to the archive itself, has been integrated in the archive infrastructure with its own server installed side–by–side with the archive database, and is part of the archive interface. Other tools have been developed in a more stand–alone way but are also serving the scientific exploitation of the archive.

Moreover the GENIUS work on dissemination has made a significant contribution to the outreach of Gaia. The Gaiaverse GENIUS portal is a hub for multi–lingual dissemination of Gaia and GENIUS has significantly contributed to many outreach activities organised around the mission and its data. In particular, GENIUS has funded the making of an outreach video during the GENIUS-DPAC plenary meeting 2015 in Leiden, aiming to explain to the general public how this type of meeting contributes to the strengthening of the collaboration between scientific teams all around Europe. GENIUS has also contributed to the formation in public communication of the Gaia community with the organisation of two courses during the GENIUS-DPAC plenary meetings in 2015 and 2017. Finally, the iconic Gaia image that has become the symbol of the first Data Release, the Sky panoramic view from Gaia counts, was generated using the visualisation services by the GENIUS team in Portugal.

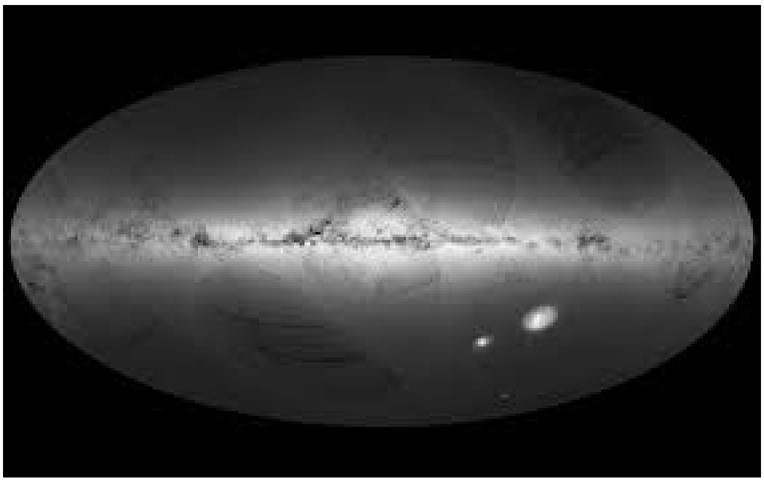

FIGURE 7.5: The Gaia Sky: a rendering based on the star counts of Gaia DR1. A. Moitinho and the Portugal GENIUS team.

3.2.3.6 Impact of deviations from the planned milestones and deliverables

Delayed deliverables

None

Delayed milestones

None

3.2.3.7 Changes to the legal status of beneficiaries

None since the last periodic report.

3.2.3.8 Use of resources

The following table lists the person–months per participant in the last period[28].

	WP1	WP2	WP3	WP4	WP5	WP6	WP7
UB	10.4	-	-	41.3	-	8.4	3
CNRS	-	-	0	-	29.2	-	0
UEDIN	-	-	21.4	-	-	-	-
UL	-	27.3	-	-	-	-	-
CSUC	-	-	-	-	-	3.6	8.9
INAF	-	9	8	-	-	-	-
INTA	-	-	0	2	0	-	-
UNIGE	-	-	-	.	0	-	-
ULB	-	-	-	-	0	-	-
FFCUL	-	2	-	18.2	2	-	-
UBR	-	-	-	0	-	-	-
UCAM	-	4	-	-	-	19.1	-
KU	-	0	-	-	1	-	-

[28]Provisional numbers until final report from financial offices of each node

The following table lists the actual effort (**AE**) for the last period against the theoretical effort for the full length of the project (**TE**)[29]. The UB actual effort in WP4 covers most of the work done by INTA.

	WP1	**WP2**	**WP3**	**WP4**	**WP5**	**WP6**	**WP7**
	AE / TE	AE / TE	AE / TE	AE / TE	AE / TE	AE / TE	AE / TE
UB	10.4/18	-	-	41.3/40	-	8.4/6	3/7
CNRS	-	-	0/1.8	-	29.2/77.6	-	0/1.8
UEDIN	-	-	21.4/43	-	-	-	-
UL	-	27.3/49	-	-	-	-	-
CSUC	-	-	-	-	-	3.6/12	8.9/12
INAF	-	9/16	8/18	-	-	-	-
INTA	-	-	0/6	2/12	0/4	-	-
UNIGE	-	-	-	-	0/6.5	-	-
ULB	-	-	-	-	0/6	-	-
FFCUL	-	2/2	-	18.2/28	2/2	-	-
UBR	-	-	-	0/3	-	-	-
UCAM	-	4/4	-	-	-	19.1/12	-
KU	-	0/2	-	-	1/2	-	-

3.2.3.9 Dissemination & Development of the project Website

Please refer to Section 3.2.2.7.2 for details on the GENIUS portal for outreach and to Section 3.2.2.1 for the internal GENIUS wiki system.

[29] Provisional numbers until final report from financial offices of each node

3.3 Deliverables and Milestones

3.3.1 Deliverables submitted in the last period of project

Find at `https://gaia.am.ub.es/Twiki/bin/view/GENIUS/DeliverablesGenius` the corresponding documents and links.

Deliverable N.	Deliverable Title	WP number	Delivery date
D1.5	Mid-term Meeting (plenary)	WP1	Nov 2015
D1.7	Semestral report 5	WP1	May 2016
D1.8	Semestral report 6	WP1	Oct 2016
D1.9	Completion meeting (plenary)	WP1	Jan 2017
D1.10	Final Report to the External advisory board	WP1	Mar 2017
D1.11	Report on dynamics of the gender balance	WP1	Mar 2016
D2.5	Conclusion of requirements update gathering exercise.	WP2	Nov 2015
D2.6	Requirements specification for incorporating new information into an existing archive	WP2	Apr 2016
D2.7	Requirements specification for the archiving of raw and intermediate data.	WP2	Nov 2016
D2.8	Requirements specification for the archiving of the original software with which the archive was produced.	WP2	Nov 2016
D2.9	Requirements specification for model comparison and optimisation tools.	WP2	Feb 2017
D3.3	Gaia Data IVOA compliance document	WP3	Mar 2017
D3.4	Deployed web services, code and documentation	WP3	Mar 2017
D3.5	Data centre report and analysis document	WP3	Mar 2017
D3.6	TAP+ code and documentation	WP3	Mar 2017
D3.7	Deployed CANFAR-style VM research environment and produced reports and documentation	WP3	Mar 2017
D4.4	Delivery of third prototype of exploitation tools	WP4	Nov 2016
D4.5	Delivery of exploitation tools user manuals and technical documentation	WP4	Nov 2016
D4.6	Deployment of exploitation tools on the first actual Gaia archive	WP4	Mar 2017
D5.4	Delivery of statistical tools (WP 550)	WP5	Nov 2016
D5.5	Delivery of model-based validation tools (WP 530)	WP5	Nov 2016
D5.7	Delivery of external validation tools (WP 540)	WP5	Nov 2016
D5.8	Delivery of special object tools (WP 560)	WP5	Oct 2015
D5.9	Deployment of validation tools on the Gaia archive	WP5	Mar 2017
D6.3	Delivery of second simulated catalogue data	WP6	Apr 2016

D6.5	Delivery of third simulated catalogue data	WP6	Nov 2016
D6.6	Deployment of third public science alerts prototype	WP6	Nov 2016
D7.4	Final (complete) version of the community portal. Handover to the Gaia community	Mar 2017	

3.3.2 Milestones in the last period of project

Find at `https://gaia.am.ub.es/Twiki/bin/view/GENIUS/MilestonesGenius` the corresponding documents and links.

Milestones N.	**Milestones Title**	**WP number**	**Delivery date**
MS9	User prototype archive review	WP2, WP3	Jan 2017
MS12	Prototype archive tools open to community	WP1, WP2, WP3, WP4, WP5, WP6, WP7	Jan 2017
MS13	Stress test	WP3	Sep 2016
MS14	Load of actual Gaia data	WP3, WP5	Jan 2017
MS15	Completion meeting & final external review	WP1	Mar 2017
MS16	GENIUS products availability	WP1, WP2, WP3, WP4, WP5, WP6, WP7	Mar 2017
MS17	Handover of GENIUS portal to Gaia community	WP7	Mar 2017

References

[ELA-026], Antiche, E., 2016, *StatisticalTools Software User Manual*,
GAIA-C9-UG-UB-ELA-026,
URL http://www.rssd.esa.int/cs/livelink/open/3396747

[GAIA-CU9-SP-UB-ELA-025], Antiche, E.F., K.; Leclerc, N., 2016, *StatisticalTools 18.2 Software Release Note*,
GAIA-CU9-SP-UB-ELA-025,
URL http://www.rssd.esa.int/cs/livelink/open/3390340

[FA-067], Arenou, A., Babusiaux, C., Fabricius, C., et al., 2016, *CU9 Validation Tests Report, preliminary TGAS 00.01 data*,
GAIA-C9-SP-OPM-FA-067,
URL http://www.rssd.esa.int/cs/livelink/open/3472775

[FA-066], Arenou, F., Michalik, D., 2015, *Simulated TGAS data used for validation Interface Control Document*,
GAIA-C9-SP-OPM-FA-066,
URL http://www.rssd.esa.int/cs/livelink/open/3332814

[FA-061], Arenou, F., Babusiaux, C., Boudreault, S., et al., 2015, *Validation Tests Specification (WP940)*,
GAIA-C9-SP-OPM-FA-061,
URL http://www.rssd.esa.int/cs/livelink/open/3311561

[FA-070], Arenou, F., Babusiaux, C., Fabricius, C., et al., 2016, *Validation Tests Report, preliminary TGAS 01.00 data*,
GAIA-C9-SP-OPM-FA-070,
URL http://www.rssd.esa.int/cs/livelink/open/3473456

[FA-073], Arenou, F., Babusiaux, C., Fabricius, C., et al., 2016, *Validation Tests Report, GDR1 (pre- and post-filtering)*,
GAIA-C9-TN-OPM-FA-073,
URL http://www.rssd.esa.int/cs/livelink/open/3473659

[DP-055], Arenou F., D.Y., Gavras P., et al., 2015, *NSS OR5 Stage 3 scientific lessons*,
GAIA-C4-TN-ULB-DP-055,
URL http://www.rssd.esa.int/cs/livelink/open/3382243

Eyer, L., Grenon, M., 2000, In: Breger, M., Montgomery, M. (eds.) Delta Scuti and Related Stars, vol. 210 of Astronomical Society of the Pacific Conference Series, 482, ADS Link

[KF-004], Findeisen, K., 2016, *StatisticalTools 18.4 Software Release Note*,
GAIA-C9-SP-OPM-KF-004,
URL http://www.rssd.esa.int/cs/livelink/open/3397614

[GGA-038], Gracia, G., Luri, X., Arenou, F., et al., 2015, *CU9 roadmap for publication of TGAS results*,
GAIA-C9-PL-ESAC-GGA-038,
URL http://www.rssd.esa.int/cs/livelink/open/3315314

[HZ-001], Houri Ziaeepour, H., Robin, A., Reyle, C., et al., 2015, *WP943 validation code, Software Release Notes*,
GAIA-C9-TN-IUOB-HZ-001,
URL http://www.rssd.esa.int/cs/livelink/open/3339079

[RMC-001], Mor, R., Robin, A., Figueras, F., et al., 2015, *Besancon Galaxy Model Simulation for CU9-WP943 (BGMBTG 2.0) I. First steps comparing Tycho-2 and Model.*,
GAIA-C9-TN-UB-RMC-001,
URL http://www.rssd.esa.int/cs/livelink/open/3325145

[LL:LRD-001], Ruiz-Dern, L., Babusiaux, C., Turon, C., Costigan, G., Arenou, F., 2015, *WP944 v18.1, Validation with external catalogues: Detailed Description and Validation Tests Report*, Tech. rep., Observatoire de Paris

[IS-001], Shih, I., 2015, *Software User Manual for CU9 ValidationTools*,
GAIA-C9-UG-OPM-IS-001,
URL http://www.rssd.esa.int/cs/livelink/open/3367978

[IS-004], Shih, I., 2015, *CU9 Validation Procedure Specification*,
GAIA-C9-SP-OPM-IS-004,
URL http://www.rssd.esa.int/cs/livelink/open/3368913

[IS-005], Shih, I., 2015, *CU9 ValidationTools Software Release Note*,
GAIA-C9-SP-OPM-IS-005,
URL http://www.rssd.esa.int/cs/livelink/open/3368582

[IS-006], Shih, I.C., 2017, *Cookbook of CU9 Scientific Validation for Data Release 1*,
GAIA-C9-TN-OPM-IS-006,
URL http://www.rssd.esa.int/cs/livelink/open/3475119

[FSP-001], Spoto, F., Tanga, P., Mignard, F., et al., 2016, *SSO Short Term processing, validation of the short arc*,
GAIA-C9-TN-OCA-FSP-001,
URL http://www.rssd.esa.int/cs/livelink/open/3422483