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

Second Year Report

Period covered: from 01/10/2014 to 30/09/2015

Project coordinator: Dr. Xavier Luri

University of Barcelona

Tel: (+34) 934039834

Fax: (+34) 934021133

E-mail: xluri@am.ub.es

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 is designed to boost the impact of the next European breakthrough in astrophysics, the Gaia astrometric mission. Gaia is an ESA Cornerstone mission launched in December 2013 and aims at producing the most accurate and complete 3D map of the Milky Way to date. A pan-European consortium named DPAC is working on the implementation of the Gaia data processing, of which the final result will be a catalogue and data archive containing more than one billion objects. The archive system containing the data products will be located at the European Space Astronomy Centre (ESAC) and will serve as the basis for the scientific exploitation of the Gaia data. The design, implementation, and operation of this archive are a task that ESA has opened up to participation from the European scientific community. GENIUS is aimed at significantly contributing 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, NanoJASMINE and JASMINE (Japan); and last but not least, the archive will facilitate outreach and academic activities to foster the public interest in science in general and astronomy in particular. GENIUS fits seamlessly into existing Gaia activities, exploiting the synergies with ongoing developments. Its members actively participate in these ongoing tasks and provide an in-depth knowledge of the mission as well as expertise in key development areas. Furthermore, GENIUS has the support of DPAC, several Gaia national communities in the EU member states, and will establish cooperation with the Japanese astrometric missions already mentioned.

3.1.2 Work performed during the second year of the project

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

User community: the analysis of the archive requirements has been extended and refined, from the general user needs to more specific analysis of several areas specifically targeted by GENIUS, discussed in section 3.2.2.2.2. Thanks to the GENIUS contribution this analysis has been extended beyond the professional astronomy to include the inputs of the amateur astronomers community; their inputs are being gathered and will be included in the reporting for the archive and the actions being agreed with the archive Systems Engineering Group.

The cooperation with the nano Jasmine and Jasmine team has continued, with representatives of these missions participating in several Gaia/GENIUS meetings. The installation of a mirror of the Gaia archive in Japan has been now formally requested by the NAOJ representatives.

Archive system design: following the analysis and design work carried out in GENIUS during its first year, the efforts are now concentrated in the implementation of features and tools. In some cases these implementations have been completed (e.g. Main Database Dictionary Tool enhancements) while in other cases either prototypes are available or

implementations in the main Gaia archive are being developed. Details can be found in section 3.2.2.3.2.

Tools for data exploitation: the task of developing VO tools has continued during this second year, concentrating on the VOSA adaptation to Gaia once the one for TopCat was completed in the previous period. The implementation of the visualization tools has progressed with the deployment of the visualization server at ESAC and the availability of the complete 3D client (in addition to the basic client); specifically, the visualization capabilities will be used for the first time during the data release rehearsal planned for the end of 2015 in view of making them fully available for the first Gaia data release in mid 2015. On the other hand, the development of data mining framework and tools has continued, with the first small scale experiments carried out at CSUC; for this purpose the testbed installed in this site has been upgraded (using GENIUS funding, as agreed in the first year review) and will shorly allow larger scale experiments. Details can be found in 3.2.2.4.2.

Validation: this period has been devoted to the implementation of the tests defined in the previous year, in addition to the definition of new tests arising from an improved knowledge of the data and validation needs. Furthermore, the implementation is now partly operational, deployed at the ESAC main archive site, with some tests already exercised using preliminary versions of the Gaia data, in particular of the TGAS (Tycho-Gaia Astrometric Solution) subset. The validation framework will be more tested during the data release rehearsal planned for the end of 2015 and shall be fully operational for the first Gaia data release in mid 2015. More details 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 which are being 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 which are other Gaia products that will also be used in other areas far beyond specialised astrophysics. Needless to say, GENIUS is a pan-European project that will foster enhanced working relationships and collaboration between European research and higher education establishments, and as such impacts 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 work 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 the potential to realize a 3D 'journey through our Galaxy' introducing many astrophysics to a new generation of students, inspiring the next generation of researchers to enter the physical sciences. Tools for teaching purposes are being developed within WP4 - there are established precedents for Hipparcos which pave the way for Gaia. For postgraduate training, actions are already being 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 origins 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 very limited) number of stars. Hipparcos had a significant impact on astrophysics, as assessed by the number of refereed publications derived from it, 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 will operate 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. Even more, because Gaia will measure the velocity and the physical properties of the observed sources, increasing the dimensionality of the observables to more than 6.

Only time will truly tell, but it is already clear that Gaia will represent 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 up 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 visualize the results to the community, GENIUS will represent the concrete and visible part of the huge work being undertaken by the 430+ European DPAC scientists and engineers, not mentioning the work done by European industry. For this simple reason the impacts of the results from GENIUS are not simply expected but nearly secured.

Indeed, it is through the work being 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:

- gathering the different fields of expertise in the community to provide advanced re-

quirements going much beyond usual queries to data archives;

- distributing the data to the whole astronomical community and enhancing the visibility and impact of Gaia;
- developing visualization and data mining tools to allow the most effective archive analysis;
- combining Gaia with ground-based data, thus extending the interpretation capabilities across archives and wavelength domains.

Although the GENIUS project is focused on the Gaia data archive, the research and development within this project will also benefit other data archives, be they from space or ground based experiments. Part of this benefit will arise naturally through the push for interoperability with other archives, while the public dissemination of the GENIUS results can be 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 necessitate this European approach. Likewise the science community is currently getting together on a European scale in order to prepare for the exploitation of 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 is cementing 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 second year the deployment of the GENIUS portal has been completed, 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 has been 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@am.ub.es

Other coordinator contacts:

* Lola Balaguer lola.balaguer@am.ub.es

* recerca.europea@ub.edu

* Xavier Gutierrez xgutierrez@fbg.ub.edu

3.1.6 Project partners

- Universitat de Barcelona (UB) - COORDINATOR
- Centre National de la Recherche Scientifique (CNRS)
- The University of Edinburgh (UEDIN)
- Universiteit Leiden (UL)
- Consorci de Serveis Universitaris de Catalunya (CSUC, former CESCA)
- Istituto Nazionale di Astrofisica (INAF)
- Agencia Estatal Consejo Superior de Investigaciones Cientificas (CSIC)
- Université de Geneve (UNIGE)
- Université Libre de Bruxelles (ULB)
- Fundacao da Faculdade de Ciencias da Universidade de Lisboa (FFCUL)
- University of Bristol (UBR)
- The Chancellor, Masters and Scholars of the University of Cambridge (UCAM)
- National University Corporation, Kyoto University (KU)

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

3.2.1 Project objectives for the period

The objective of the GENIUS project is 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.

- **Tailor to user needs**

  During the first year of GENIUS the work concentrated on the definition of (a priori) reuquirements for the archive. Now, with the mission well into its second year of observations and the implementation of the Gaia archive getting ready for the first data release, the goal was to become much more specific. We are comparing the actual implementation of the archive at ESAC with the previously defined requirements, identifying new ones and involving the community at large in the task (through the beta-testing of the archive).

  Related milestones for the second year: MS5, MS6, MS8, MS9, MS10, MS11, MS12

- **Optimum archive system**

  Like in the above case, the goal in this area was to make the design work much more specific. The archive implementation at ESAC is being tested and advanced features have been implemented in prototypes. The tasks are now focusing in more detailed and advanced issues related to the archive architecture.

  Related milestones for the second year: MS6, MS8, MS9, MS12

- **Tools for exploitation**

  The goal in this area was also to move from design to prototyping: the visualization prototype has been installed in the main archive at ESAC and the data mining prototype is being tested at CSUC. In the former case it will become an operational system during 2016 and in the later more advanced prototypes will be releases. Also, a new VO tool (VOSA) has been tailored for the Gaia archive in addition to the one in previous year (TopCat).

  Related milestones for the second year: MS6, MS8, MS10, MS12

- **Validation**

  The goal for this period was to make the validation software to become almost operational. The tests developed have been implemented in the main archive and have been run with real Gaia data – the TGAS solution, a relatively small subset of the Gaia catalogue (2 million objects) – and will run shortly in a full dataset representative of the first Gaia data release.

  Related milestones for the second year: MS6, MS8, MS11, MS12

- **Dissemination**

  After a first year of preparatory work we aimed to have a full dissemination system integrated in the wide Gaia outreach activities. The core of this area is the Gaiaverse

portal, that has become fully operational during the second year and now is a hub for the dissemination of GENIUS and Gaia information.

Related milestones for the second year: MS7

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 provides the overall administrative management of GENIUS.

3.2.2.1.2 Summary of progress made

This work package includes 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 are carried out by the GENIUS coordinator assisted by a hired project manager.

Task 1.1 - Global administrative tasks

Task leader: UB

As planned, the second year is moving from analysis and design to the actual implementations into the Gaia archive. The coordination work with the full DPAC consortium in general and the archive team at ESAC in particular has intensified. In particular the dual role of several of the GENIUS members, also acting as coordinators and task managers in DPAC, has 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 have implemented 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.

Specifically, monthly teleconferences have been held with the representatives of each node in order to track progress and, like in the previous period, a joint meeting with the Gaia CU9 was organized; this time it took place in Barcelona with a participation of 95 GENIUS and CU9 members. This meeting served both for the coordination of the GENIUS-CU9 activities and to carry out scheduled internal GENIUS reviews.

Finally, the external advisory board selected at the end of the previous period has been provided with project reports and documentation in order to carry out the mid-term review (milestone MS8); the purpose of this review is to get an independent assessment of the project advancement from external experts and will take place in November 2015.

3.2.2.1.3 Highlights

- Completion of GENIUS first year review (see report at `https://gaia.am.ub.es/Twiki/pub/GENIUS/ReferenceDocuments/GENIUS_Firstyear.pdf`)
- Gaia Archive Review 1: joint meeting with ESAC Science Archives Team (see `https://gaia.am.ub.es/Twiki/bin/view/GENIUS/GaiaarchiveReview`)
- Monthly progress tracking teleconferences (see `https://gaia.am.ub.es/Twiki/bin/view/GENIUS/MeetingsTeleconfs`)
- Joint CU9-GENIUS plenary meeting in Barcelona ((see `https://gaia.am.ub.es/Twiki/bin/view/GENIUS/PlenaryMeeting2015`)
- Set up of the external advisory group and organisation of the mid-term review (see `https://gaia.am.ub.es/Twiki/bin/view/GENIUS/MidTermReviewMeeting2`

3.2.2.1.4 Deviations and impact on tasks and resources

Due to administrative reasons the status of the CSIC partner has created problems with personnel participation and the GENIUS contracting. This problem and the solutions adopted are discussed in detail in section 3.2.3.

Otherwise there have been no deviations in the administrative management tasks. Note that the deviations in the technical development and general project schedule are discussed in section 3.2.3.5.

3.2.2.1.5 Use of resources

The following table lists the person-month per participant in the first 12 months in WP1 There is a difference on the dedication of L. Balaguer with that in the UB Form C due to the payment structure based in 6-months bills.

	UB
WP1	5

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 is to ensure that the corresponding technical developments are driven by and focused on the scientific needs of the astronomical community that will use the Gaia catalogue. That is, the Gaia catalogue and data archive should be tailored to the needs of the scientific end user, but also the interested amateur or curious member of the general public. Tailoring should be done by capturing the end user's scientific requirements and turning those into specifications on the basis of which the Gaia data archive, catalogue and data access methods can be built. This issue has been recognized 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. This process is 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 will this be done in practice and what requirements does that set on the way the Gaia data is published and made accessible? In this work package these top level requirements will be analysed with the goal of turning them into detailed requirements. These requirements should be cast in a language that both the scientists and the archive developers understand. The GAP requirements gathering process has revealed a number of advanced requirements (the Grand Challenges) that go much beyond the normal queries to data archives, and which require research in order to work them out in detail. Implementing these requirements will 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 the following Grand Challenges will be researched in this work package.

3.2.2.2.2 Summary of progress made

Task 2.1 - Technical coordination

Task leader: UL

Brown has 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 telecons and also the GENIUS kick-off and plenary meetings.

Two GENIUS postdocs were hired in Leiden:

- Gráinne Costigan, who started on May 5 2014
- Arkadiusz Hypki, who started on September 1 2014

Costigan concentrates her efforts mainly on T2.2 and T2.5, while Hypki works mainly on T2.3 and

T2.6.

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

Task leader: UL

Contributing partners: FFCUL, UCAM, KU

UL Contribution

Costigan is working alongside, Nic Walton (UCAM) and Gonzalo Garcia (Gaia Project Office) to setup the beta testing of the Gaia archive. This has involved a detailed analysis of the use case scenarios as submitted by the community, and their implementation in the Gaia archive (see GAIA-CU9-TN-LEI-GCO-002). The missing components in the archive, and weak points that cannot be tested at this time, have been identified. These have also been follow-uped with the work package managers and CU leaders responsible. This work was presented at the CU9 plenary in Barcelona during the Architecture splinter session, where useful feedback was recieved. The beta testing will soon begin in the coming months, concentrating on those features needed for the first and second data releases.

Currently she is working on gathering user requirements from amateur astronomers in different countries, currently Germany, Spain, Portugal, Ireland and the UK. By in large this group have very similar needs to those of the professional astronomer, but this activity can also be used to promote Gaia and upcoming data releases. This feedback will soon be written up as use cases in a technical note so that they may become useful for the greater Gaia community.

Costigan has also worked in supporting the GAVIP team in their development of the python access tool. This involved face to face meetings and telecons and drawing connections between different parts of the DPAC team and GAVIP.

KU contribution

The GENIUS activity in Japan is mainly focused to Nano-JASMINE / Gaia synergy. At the April 2015 AGIS meeting, the KU team reported the status of the development of the Nano-JASMINE analysis software, mainly attitude modeling and centroiding. Within CU9 the hosting of Gaia data from a server in Japan is being discussed.

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 may wish to use the GENIUS archive, and outline 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-PM8`) at the European Astronomical Society's 'European Week of Astronomy and Space Science' conference, held in La Laguna, Spain, June 2015.

The scenario collection period remains open, 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 summer 2016. It was therefore decided to delay the delivery of the document 'Conclusion of requirements update gathering exercise' (D2.5) until after that time.

FFCUL contribution

The list of visualisation use cases, initialised in the first year of the project, was revised. These use cases were the starting point for defining the requirements on the visualisation software and services. The use cases were mainly extracted from the following sources:

(i) the general scientific community interested in exploring the Gaia archive. Gathered in the Gaia Data Access Scenarios document (Brown, A., Arenou, F., Hambly, N., et al., 2012, Gaia data access scenarios summary. ref: GAIA-C9-TN-LEI-AB-026)

(ii) the Gaia archive builders, represented by the management of the Coordination Unit 9 (CU9 - Catalogue Access) of the Gaia Data Processing and Analysis Consortium (DPAC); These requirements are laid out in O'Mullane, W., 2009, Gaia Catalogue and Archive Software Requirements and Specification. ref: GAIA-C9-SD-ESAC-WOM-033.

(iii) A previous extensive study for ESA performed by the FFCUL team. The study was named VA-4D — Visual Analysis of 4-Dimensional Fields, Processes & Dynamics (AO 1-6740/11/F/MOS) and the relevant reference is the Framework Requirements Definition and Gap Assessment (ref: VA-4D-100-05)

The Visualisation requirements have led to the definition of an architecture (server, plug-ins and clients) which also puts requirements on the Gaia archive infrastructure. These requirements on the archive are the subject of on-going work and interaction with ESAC for the integration of the visualisation software at ESAC.

In preparation of Gaia-DR1 and of the possible TGAS release, the visualisation use cases made possible by these releases have been assesed. This assesment has resulted in the definition of the Visualisation services to be available by Gaia-DR1.

Task 2.3 - Confronting complex models with complex catalogues

Task leader: UL

Hypki has developed a proposal for an API for the Gaia archive that would enable users to accomplish the following tasks in support of testing large and complex models (e.g., numerical simulations of the Milky Way) against the Gaia data:

- Upload their simulations or models to their user space within the Gaia archive.
- Upload their code to carry out the data analysis or model tests.

- Query the Gaia catalogue from with their code.
- Save and download results for later publication.
- Share their data and code with other users of the archive.

The API could be implemented following the concept of thick serves coupled to thin clients, where the communication between the services above is through HTTP request/response calls following the REST approach. The proposal is documented in two technical notes (HYP-001 and HYP-002). These notes cover the deliverables 'Requirements specification for generic projection module' (D2.3) and 'Requirements specification for model comparison and optimization tools' (D2.9)

Hypki will also explore, through the implementation of a concrete example, the alternative approach to such an API as developed in the context of the GAVIP (Gaia Added Value Interface Portal) project. Costigan has worked on identifying further tools that could be incorporated into GAVIP.

Task 2.4 - Seamless data retrieval across archives and wavelength domains

Task leader: INAF

The activities for this task are closely connected to the DPAC CU9 cross-match activities in the near-infrared/optical/UV range and benefited from the CU9 developments. The INAF team started the multi-wavelength cross-match work 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. He also made a census of the cross-match techniques used in the radio domain and the INAF team are in the process of implementing, testing and comparing different algorithms. The work will continue with infrared, X-ray and gamma-ray domains.

The above efforts resulted in two technical notes which correspond to the deliverable D2.4 (Requirements specification for data retrieval across archives). The first note (PM-011) provides the motivation, and links to, the requirements for data retrieval across wavelength domains, while the second one (MFA-001) provides an overview of available radio catalogues (which could be matched with Gaia) and their characteristics.

Task 2.5 - The living archive

Task leader: UL

Costigan worked on developing ideas behind the Living archive aspects of the Gaia archive. This was written up in the document GAIA-CU9-TN-LEI-GCO-003. Based on this work we recommend that the functionality of the archive be expanded to allow the inclusion of derived parameters by the community — (parameters that may be derived in specific ways for specific sources or based on focused observing campaigns may lead to more accurate results than those provided by the Gaia data processing). This document was presented at the CU9 plenary in Barcelona during the Architecture splinter session.

Task 2.6 - Re-processing of archived (raw) data

Task leader: UL

This activity has not started yet.

3.2.2.2.3 Highlights

3.2.2.2.4 Deviations and impact on tasks and resources

Deliverables D2.3 and D2.4 were delayed and were submitted at the end of September 2015. Deliverable D2.5 is delayed until after the next round of requirements gathering which will take place in the summer of 2016 in anticipation of the second Gaia data release.

3.2.2.2.5 Use of resources

The following table lists the person-month per participant in the second year in WP2.

	UL	**INAF**	**FFCUL**	**UCAM**	**KU**
WP2	25.4	3.0	0.0	0.5	1.0

3.2.2.3 Work Package 3: Aspects of archive system design

Lead Partner: UEDIN

Contributing partners: INAF, CNRS, CSIC

3.2.2.3.1 Overview of WP objective

The objective of this workpackage is to design, prototype and develop aspects of the archive infrastructure needed for the scientific exploitation of Gaia data. The design and technology choices are being 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 are being made with full recognition of the constraints imposed by the ESAC archive system, with which it must interface effectively. Prototypes are being prepared and tested in cooperation with the end user community and with the ESAC science archive team through the DPAC CU9. A core principle is the adoption of Virtual Observatory standards and the development of VO infrastructure 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 continues 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 reflects the complete integration of the GENIUS activities with the wider DPAC CU9 developments. Our agile approach makes use of a planning and reporting worksheet[1] designed by DPAC CU1 whereby work is split into units of half a day and a work plan is assembled and reported on every two months. Again, this is done for the wider CU9 WP930 developments, of which GENIUS WP3 is a major component, and ensures a well integrated program of work, efficient use of resources, and avoidance of duplication of effort.

Task 3.2 - Aspects of archive interface design

Task leader: UEDIN

In addition to definition of a WP system requirements specification[2] and subsystem Interface Control Documents for coordination with the ESAC Science Archives Team[3], interface design has been tackled via enhancements to DPAC's Main Database Dictionary Tool specifically for publication of catalogue data, particularly via Virtual Observatory standards. The latest release of the DPAC tool earlier this year and during the reporting period has made those enhancements available to

[1] for a recent example see `https://docs.google.com/spreadsheets/d/1Y_wysiyS6i8sw62eZ_zXe0u4A-c8Vfsm9pi7zd37UCY`

[2] `https://gaia.am.ub.es/Twiki/pub/GENIUS/MilestonesGenius/GAIA-C9-SP-IFA-NCH-031.pdf`

[3] see for example `https://gaia.am.ub.es/Twiki/pub/GENIUS/DeliverablesGenius/GAIA-C9-TN-IFA-DM-001.pdf`

the development teams and work is now in progress in annotating the catalogue data models with human–readable descriptions and VO standard content descriptors. Note that this task is concerned with archive infrastructural interface design as opposed to end–user interface design (for which see Task 3.4 below).

Task 3.3 - VO infrastructure

Task leader: INAF

Contributing partners: CNRS, UEDIN, CSIC

To generalize the VO infrastructure including even small general researchers datasets a set of tools has been developed at INAF to deploy TAP services out of a DB resource and a few configuration steps. The first year of the project focused on identifying solutions based upon existing software packages available in house, this second year saw the preliminary builds of the three proposed solutions for an easy–to–deploy TAP service for general users. Improvement of the IA2TAP application has continued with respect to standard revision at IVOA level. Moreover, a BasicHTTP Authentication thin layer has been added on top of it enabling data policy capabilities. A KVM Virtual Machine has been built around the IA2TAP web application, providing all the software stack needed to run it (MySQL server, Java environment, Glassfish container, basic configuration). A first package with the IA2TAP application embedding the Glassfish container as been built, even if currently lacking the authentication capability.

UEDIN contributions in this area have continued to focus on working with the International Virtual Observatory Alliance[4], at both face–to–face meetings and via the Standards committee, for the adoption of several new features of the ADQL standard that will be of great importance to users of Gaia data. UEDIN is also directly involved in the IVOA as a member of Data Access Working Group developing the next version of the ADQL standard which will help to improve the interoperability of the Gaia data published by ESAC–SAT with data published by other archives within the IVOA and with data access clients developed using the IVOA standards. UEDIN has also continued to contribute to the ADQL parser implementation that is common to ESAC–SAT, providing enhancements and bug fixes that will be of great benefit to the users of the Gaia Archive Core Systems. Our approach with testing the ADQL parser has been to take a large set of commonly–executed user queries harvested from many years of operations of large–scale catalogue databases at our own Data Centre in order to shake out bugs and limitations when exercised by real–world astronomical queries. The parser is also under test by our colleagues at INAF–OATs as part of the development of their client–side applications (see above). Our modifications to the ADQL parser, forked from the original GitHub trunk, are now in the process of being merged back in to trunk in collaboration with the original author (Grégory Mantelet at MPIA Heidelberg).

The CNRS and CSIC components to this task will be undertaken later in the project.

[4] http://www.ivoa.net

Task 3.4 - Data Centre Collaboration

Task leader: UEDIN

UEDIN has continued to develop Data Centre infrastructure for enhanced User Interface functions and generalised Distributed Query Processing. Our end–user interface design includes several features not presently available in the Gaia Archive Core Systems (GACS) User Interface, e.g. ADQL auto–completion and other context–driven features along with embedded plotting and browsing. In order to demonstrate these developments we have set up a GENIUS "Web2.0" end–user interface demonstrator[5] (Deliverable D3.2). The demonstrator acts as a wrapper for the Distributed Query Prototype, being connected to the GACS TAP service presently serving GUMS data as well as local large survey catalogue holdings like 2MASS. We are now in contact with ESAC–SAT and discussing how to integrate these developments into the GACS. Note that our DQP system is built around OGSA–DAI[6] which is no longer supported by the original implementation team. A significant amount of effort has been expended in bug fixing and enhancing this key infrastructural component.

Task 3.5 - Cloud–based research and data mining environments

Task leader: UEDIN

Preliminary work for this task has been undertaken as part of investigations of containerisation of components for ease of deployment within GACS. ESAC–SAT has understandable concerns about security and availability of their systems and one way of secure deployment of server–side components, at the same time confining processes to reasonable CPU resource, is to use Virtual Machines. The emerging technology here is Docker[7] and to investigate this we have implemented containerised versions of the subsystems described in Task 3.4 and will use these for initial deployment within GACS. In collaboration with Wil O'Mullane of ESAC a paper surveying[8] use of Docker is in preparation (Proceedings of the XXVth conference on Astronomical Data Analysis Software and Systems).

3.2.2.3.3 Highlight

In Figure 3.1 we show a screen shot from the GENIUS "Web2.0" / Distributed Query Processing demonstrator which connects to the GACS TAP service and provides cross–querying capability amongst several large–scale survey catalogue datasets.

3.2.2.3.4 Deviations and impact on tasks and resources

Milestone MS9 ("User prototype archive review"; 1/10/2015) is being delayed by 3 to 6 months in order to synchronise with the DPAC CU9 Gaia DR1 rehearsal and beta–testing of GACS.

As described in section 3.2.3, due to internal organization and administrative problems the CSIC

[5] see `https://gaia.am.ub.es/Twiki/bin/view/GENIUS/Web2p0Demonstrator` for a description and visit `http://genius.roe.ac.uk` to try out the service features.

[6] see `http://www.ogsadai.org.uk`

[7] see `https://www.docker.com`

[8] see `https://docs.google.com/forms/d/1h4j5xJABut5tICfKbskd_wDYjcPGS5lGVdqJNhD821M/viewform`

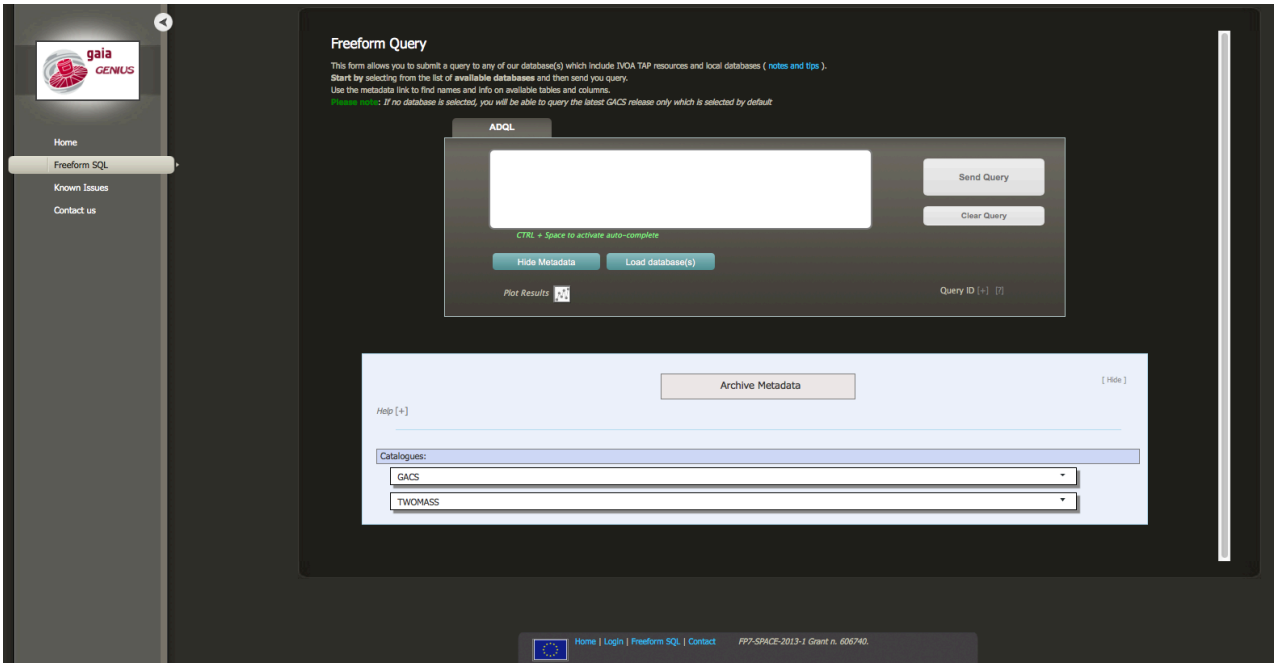


FIGURE 3.1: Screen shot from our GENIUS "Web2.0" and DQP demonstration prototype: see `https://gaia.am.ub.es/Twiki/bin/view/GENIUS/Web2p0Demonstrator` for a description and visit `http://genius.roe.ac.uk` to try out the service features.

partner has been unable to complete the planned hiring of GENIUS personnel for the development of the VO tools. This task has been continued at a slower pace by permanent personnel in CSIC and mainly in the associated institution INTA, but the inclusion of the personnel from this second instituion also created administrative problems (see also section 3.2.3). To solve this situation and specially to allow for the needed hiring of GENIUS personnel without more delays, the associated funding will be reallocated to the UB which will directly manage the contracts (to be agreed by the consortium during the second year review). This will allow the timely completion of this part of GENIUS before the end of the project.

3.2.2.3.5 Use of resources

The following table lists the person-month per participant in the second 12 months in WP3

	UEDIN	**CNRS**	**INAF**	**CSIC**
WP3	11.25	0	9.12	0

3.2.2.4 Work Package 4 Tools for data exploitation

Lead Partner: UB

Contributing partners: CSIC, FFCUL, UBR

3.2.2.4.1 Overview of WP objective

A use of the Gaia archive based on simple queries (i.e. sky region queries) would only allow a basic use of its potential. To fully exploit a billion object data set, containing a wide variety of data (astrometric, photometric, spectrophotometric, spectroscopic, . . .) more advanced and powerful data exploration tools are needed. This work package is devoted to the development of such tools, in close coordination with WP2 to ensure that they are tailored to the actual needs of the scientific user community. It includes:

- Development of visualization tools, adapted both to the potential large size and complexity of the available data of the results of the archive queries.
- Development of data mining tools and infrastructure adapted to the characteristics of the archive (both to its contents and the archive system), allowing the users to perform data mining tasks and extract new knowledge.
- Development or adaptation of VO tools and services to the Gaia archive. In particular, the possibility of cross-matching the contents of the Gaia archive with other archives (specially with large surveys ongoing or in preparation, like LSST) should be easily available.
- Development of tools for the Grand Challenges outlined in WP2, that involve complex and massive exploration of the data.
- Furthermore, this work package also includes the development of some tools for outreach and academic activities. Although not explicitly included in the call, we consider the task of presenting astronomy to the general public and the provision of resources for teaching astronomy based on actual Gaia data as worthy contributions to the dissemination of space mission data on a global scale.

3.2.2.4.2 Summary of progress made

Task 4.1 - Technical coordination

Task leader: UB

The tasks in this work package (visualization, data mining and VO tools) are mostly independent from each other. Therefore, the coordination by the UB has focused on tracking 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 - Visualization tools

Task leader: André Moitinho (FFCUL)

Contributing partner: UB

The visualisation services for the Gaia archive are based on a client-server architecture defined during the first period of the GENIUS project. The Server (aka Object/Visualisation Server) runs close to the data. The Clients provide visual display and user interface, usually at a location away from the archive.

The architecture is designed to support plugins that can be used for extending the server-side capabilities in several ways (such as data transformations, data simplifications, volume calculation, indexing, etc.).

During the reported period the main activities for T4.2 were:

Visualisation server

- Continued working on improving the performance and stability. This included the design and implemented a new baseline approach computing visual levels of detail. Optimisations of data structures, indexing were performed. Together, the changes resulted in an important increase in fluidity, lower storage and optimized data transfer
- Creation and testing of a full 3D interactive visualization of the GUMS Milky Way (GUMS-MW)
- Development of an API and wrappers in different languages (currently C#, Java and Python) for communication of clients with the server
- Confirmed that pre-computations for each full GUMS-MW visualization will take at least 270GB of space to be persisted
- Confirmed that disk I/O is a bottleneck (many random reads and writes needed) for the pre-computations

3D Client

- New approach to detail on demand
- Improved SAMP support
- Implemented several functionalities for 3D interactive data selection
- Added mini-map feature and added axes support

- Tested with visualizations of up to 100M points

Integration with Gaia archive at ESAC

- Defined requirements for computing infrastructure at ESAC for running visualisation services
- Reported and validated capabilities of Virtual Machines provided by ESAC for testing the services
- Created TAP vs JDBC test case, which simulates a some of the workflow (includes, sequential and random access tests)
- Preliminary definition of the deployment mechanisms for the full server-client visualisation services
- Continuous communication (mail exchanges and telecons) with ESAC

Meetings and telecons

- Several telecons and email exchanges with ESAC for integration and deployment of the visualisation services with the Gaia archive
- Physical meeting at ESAC for work on integration and tests
- CU9/ GENIUS telecons
- CU9/GENIUS Plenary meeting, Barcelona September 21-23, 2015.
- Splinter session on Visualisation at the CU9/GENIUS Plenary meeting. Discussions and plans for integration visualisation developed by other groups tools with the CU9 Visualisation architecture.

Presentations

- A. Moitinho, "3D visualization of stars and galaxies from Hipparcos to Gaia. From Hipparcos to Gaia", Light, from the earth to the stars, a IYL2015 Multidisciplinary Conference organized by CiÃªncia Viva and the Academia Europaea, Lisbon July 2-3, 2015
- A. Moitinho, "WP980 Plans for GDR1", CU9/GENIUS Plenary meeting, Barcelona September 21-23, 2015

The following figures illustrates a visualisation of a subset of 100.000.000 stars in the GUMS simulation. It shows a view of the Galactic centre as seen from the Earth. The image is a snapshot of the display panel of the 3D interactive client. In this example, the (detail on demand) data was being dinamically served by the Visualisation Server running in a remote computer at FFCUL.

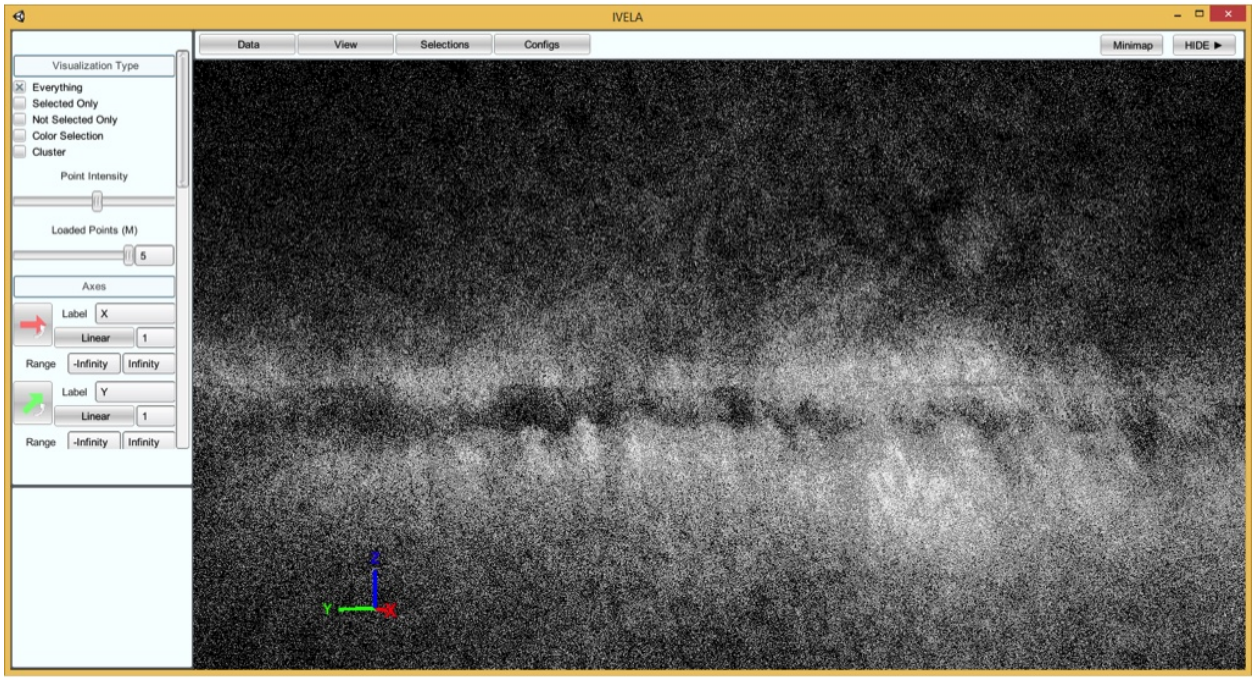


FIGURE 4.2: Snapshot of a 3D interactive visualisation of the GUMS simulation of the Milky Way.

Task 4.3 - Data mining

Task leader: UB

Contributing partner: CSIC

Overview

GAIA will implement an advanced data access framework that will allow performing complex queries to the archive. The complexity of the queries and the size of the archive are the main drivers to approach those advanced queries using Big Data technologies instead of a more traditional archive querying mechanism that would require too many resources and computing time. These technologies are powerful tools that allow the user to extract advanced information from a big archive with a priori hidden correlations in its contents.

This framework has been called GDAF (Gaia Data Analytics Framework)

Participants

The following table lists the DPAC members involved in the Data Mining tasks in the coordination unit 9 (CU9). The members of GENIUS are indicated with an asterisk.

Name	Institute
Bello, Antonio	Univ. Oviedo
Berihuete, Angel	Univ. Cadiz
Borrachero, Raul	Univ. Barcelona
Dafonte, Carlos	Univ. Coruña
di Cecco, Alessandra	INAF-ASDC *
Gil, Alfred	CSUC *
GarcÃa Torres, Miguel	UPO - univ. Pablo de Olavide
Giuffrida, Giuliano	INAF-ASDC *
GonzÃ¡lez-Marcos, Ana	Univ. de la Rioja
Julbe, Francesc	Univ. Barcelona, DPAC-CU9 Work package coordinator *
Luri, Xavier	Univ. Barcelona *
Manteiga, Minia	Univ. Coruña
Marinoni, Silvia	INAF-ASDC *
Marrese, Paola	INAF-ASDC *
Ordieres, Joaquin	Univ. PolitÃ©cnica de Madrid
Sarro, Luis	UNED, CU9 Data Mining Work package deputy *
Tapiador, Daniel	Univ. Comp. Madrid

GENIUS data mining testbed implementation

During second year of activity in this work package we have mainly used the Cadi cluster, an infrastructure composed of 16 nodes with 4 cores each. However, the storage capacity end computing efficiency of this cluster is not good enough to explore the scientific use cases that we want to work within the context of this work package. Following the agreements of the GENIUS first year review, some of the GENIUS funding initially budgeted to the buying of software licenses was instead devoted to the acquisition of a new cluster at CSUC. This partner selected and acquired a brand new cluster infrastructure which can provide the resources to efficiently work with the scientific use cases planned (see USE CASES section).

The new Cluster was funded partly by GENIUS funds and it was finally installed and deployed at CSUC facilities. This deployment took some months of delay which slowed down the development progress. Currently the setup of the system is being carried out by the UB with technical support of CSUC.

This cluster (named Adenya) consists of 6 nodes with 16 cores each, about 70TB of storage capacity and 392GB of RAM memory.

From the Big Data software side, a complete system deployment and setup has been put in place, which consist on a Cloudera Hadoop distribution alongside Spark framework for the Big Data analytics. We have defined a deployment and setup procedure in order to replicate it in ESAC facilities once the infrastructure setup and operation was consolidated, meaning that it is ready and tested with real use cases and real data, being ready for larger tests in final facilities.

Framework features

In opposition to the activity during the first year, which was much more focused on learning new Big Data technologies (introduced in “Preliminary studies” section in first year report), a clearer

feature set to be provided by the platform has been defined during this second year of activity, helping us to design a work plan for the short/mid-term.

The GDAF features can be divided in two parts:

- Basic data analytics tools, which are those commonly and widely used analytical tools such s clustering techniques, regressions and dimensionality reduction methodologies. Spark implements many of them and the list is growing.
- Advanced features: These are those features that offer specific problem solving, more focused on astronomical use cases that can't be addressed using the basic features.
  - Diffusion Maps
  - LLE (Local Linear Embedding)
  - Kernel PCA
  - HMAC (Hierarchical Mode Association Clustering)
  - Self-Organised Feature Maps - Implemented by Univ. Coruña colleagues for Spark and also running on GPUs.
  - Bayesian hierarchical models

Introduced recently in the Spark Framework, there is a brand new binding between Spark and R language (SparkR). R is one of the most widely used language to perform data analytics. This new binding means a significant boost for the usability and potential scientific return of the platform. Initial use cases have been analysed and tested providing interesting results. However, the feature was first introduce in Spark in version 1.4 and currently more newest Spark release is 1.5, so the functionality of SparkR is still very limited. Data visualization features will also be provided. Some initial steps have been taken in order to provide such a features. There is also a work package only to provide visualization capabilities to the archive.

Use cases

The value of the GDAF platform will be based on the astronomical use cases that can be addressed using it.

A library of astronomical use cases is being recollected in order to build a sort of recipes on how to solve basic problems. These use cases library will be the building blocks of more complex use cases, increasing its complexity and scientific value. Besides, these recipes will be a manual on how to use the framework from user perspective.

The initial use cases collected are:

- Clustering on TGAS (Tycho-Gaia Astrometric Solution): Clustering in the space of spatial positions and proper motions. We should rediscover proper motion groups and maybe even find new groups.

- TGAS cross-match with other surveys: Cross match between TGAS and other photometric surveys and do a search for exotic objects. For example, sub dwarfs or white dwarfs. This might be through a classifier, or again with a mixture of Gaussians

Grand challenges

Beyond the basic use cases, there are the so-called "grand challenges", which are those astronomical problems that require high level of computation, data processing and that will provide great scientific value thanks to the quality of Gaia data and the features provided by GDAF.

One of these Grand Challenges consists on a global comparison of existing galactic models with real Gaia Data. Given the accuracy and Gaia observations this problem can now be revisited at a level that couldn't be done before. In this field, the tasks performed so far are:

- Scientific problem identification

  We have proposed a proof of concept to evaluate the scalability of a computationally intensive application through advanced statistical techniques. To do this we set out the following question: what is the shape for the initial mass function (IMF) and star formation rate (SFR) taking into account millions of observations in the Gaia catalogue?

- Methodology

  The question raised can be addressed by Bayesian inference via hierarchical models. The fact considering such models means that:

  - we establish a joint probability distribution for IMF and SFR parameters, which expalin the dependence between the variables (at least the most relevant) involved in the problem.
  - the uncertainties in the observations would transfer appropriately to the different model parameters, i.e., the observations are conditionally modeled to certain parameters, which in turn are defined in probabilistic terms via other parameters known as hyperparameters.

  Due to the complexity of the models it will be necessary to use Markov chain Monte Carlo (MCMC) methods in order to obtain an i.d.d. sample from the joint posterior probability distribution of the parameters to be estimated. We have considered several algorithms for this task:

  - NUTs (http://mc-stan.org/)
    This algorithm was rejected because its high dependency on C++ libraries. In order to do the sampling, it was necessary to compile the model on C++ and this task it was not compatible with the Spark framework.
  - * Nested Samling (http://arxiv.org/abs/0704.3704) This algorithm was rejected because there was not a parallelizable version. But recently a new version has been announced which can run on openMPI. It would be interesting to test this version if we have time for it. See `http://arxiv.org/abs/1506.00171v1` and `http://ccpforge.cse.rl.ac.uk/gf/project/polychord/`

  - emcee (`http://dan.iel.fm/emcee/current/`). This algorithm was accepted. It is based on python (which works well with Spark) and it is highly adaptable to other cluster managers (CONDOR, SLURM). The code of different models are written in python, which is widely used in the community of astronomers.

- Grand Challenge Progress

  - The theoretical justification of the problem has been written on the overleaf portal. This will be the seed of the methodology section in a future paper. See: `https://www.overleaf.com/read/ptgxxtpmwdjf`
  - The estimation of IMF parameters is done. For the first approximation it was considered a classical parametric function of the IMF (three power-law). We considered the marginalization of the likelihood to tackle its intractability due to the huge number of observations. The masses considered were simulated from theoretical cumulative distribution function. No real Gaia data, nor GOG simulations, were used.
  - The estimation of the SFR is in progress. In this case we have increased the complexity of the model considering no parametric function for the SFR. In order to obtain the shape of this function, we have considered the use of a Gaussian Process (GP) which combines the hierarchical Bayesian framework in a regression problem. Again, we used simulated ages derived from theoretical distributions.
  - Joint estimation of IMF and SFR is in progress. This will be the last step of this Grand Challenge where we are trying to model the surface of the joint probability distribution of IMF and SFR by using a GP.

Data lake

In order to develop astronomical use cases on this test platform, a good data archive for distributed computing has to be in place. This has been one of the main priorities during this second year of GENIUS project. A procedure on how to process data for the Gaia relational archive and transform it to a suitable format that can be consumed by GDAF has been designed and prototyped.

Validated data is transformed into ASCII files that are stored in the HDFS distributed file system. A transformation process starts, parsing this data and transforms it to Parquet, a columnar file format compatible with all Big Data Technologies being used in GDAF. This format is very handy and allows easy querying of large amounts of data in a SQL fashion.

This procedure has been successfully tested with TGAS data and also with GACS simulated data.

Task coordination

In this second year there has been a good contribution to the tasks to be performed from the different partners.

Meetings:

- Meeting at ESAC with Luis Sarro and Daniel Tapiador in December 2014
- Work package meeting took place at ESAC facilities in May 2015, with people attending from INAF, UB, CSUC, UNED, University of Coruña and ESAC (SAT team).
- Splinter session in CU9 plenary meeting at Barcelona in September 2015.
- Several follow-up telecons.

The GENIUS wiki platform has been used for documentation exchange and share.

Task organization

The project and work plan, which will be introduced in sections ahead, has been a key aspect during this period. The following working domains have been defined, to be implemented in DPAC-CU9/GENIUS-WP4:

- 000 COORDINATION, coordinated by UB
- 100 INFRASTRUCTURE, coordinated by UB, participation of CSUC and CSG/SAT
- 200 USER PROFILE DEFINITION, coordinated by Luis Sarro
- 300 EXECUTION POLICIES, coordinated by UB, validated/agreed with CSG and SAT
- 400 VISUALIZATION INTERFACE, coordinated by Univ. Coruña
- 500 GRAND CHALLENGES, coordinated by Luis Sarro
- 600 FEATURES, coordinated by Luis Sarro
- 700 SCHOOLS and WORKSHOPS, coordinated by UB

Some of them have good progress, others are envisaged for longer term

Next year goals

Data mining framework is a long term project, of which GENIUS aims to ensure the first implementation. In fact, GDAF platform is not expected to be deployed at ESAC until 2018. However, for next year we expect to have a basic fully operational GDAF platform, with astronomical use cases already defined and evaluated. Also, some visualization capabilities should be available. The Grand Challenge has been tested locally in laptops and desktop computers. For the coming year a large scale execution of the use case will be performed in the Adenya cluster. All these features available will materialize in a GDAF release (pre 1.0 as it is only a development and test platform given the Gaia archive project plan).

Task 4.4 - VO tools and services

Task leader: CSIC

Contributing partners: UBR

CSIC contribution

Introduction

Our efforts in this second reporting period have focused on the VOSA adaptation to Gaia need. VOSA (VO Sed Analyzer, `http://svo2.cab.inta-csic.es/theory/vosa/index.php`) is a public, web application developed by the Spanish Virtual Observatory designed to help users to build SEDs gathering phtometric information from VO archives and to derive physical parameters from the comparison of the observed SED with different collections of theoretical models.

VOSA is in operation since 2008 and counts with more than 500 active users (¿7.000 files uploaded by users and ¿600.000 objects analysed) who have published more than 70 refereed papers making use of this tool.

During the second reporting period we have upgraded VOSA to provide access to Gaia photometry and give a reliable estimation of the physical parameters of thousands of objects at a time. This upgrade has required the implementation of a new computation paradigm (including a distributed environment, the capability of submitting and processing jobs in an asynchronous way, the use of parallelized computing to speed up processes and a new design of the web interface). A detailed description of these developments can be found below.

Access to Gaia photometry

VOSA has implemented TAP to access Gaia data from the Gaia Archive Core System, offering the possibility of incorporating G magnitudes in the SEDs.

Distributed environment

VOSA has been upgraded to perform most of the computational work in a dedicated server, dramatically reducing the load of the VOSA main server, which is no longer affected by the number of jobs or the size of the user files. A future increase in the number of computational servers was taken into account in the system design.

Parallelized computing

We have upgraded VOSA so as to allow computation in a parallelized way, making not necessary to finish an object before starting with the next one. The computation server organizes the jobs so that several of them are carried out simultaneously, and collects the results once all the jobs are finished.

Asynchronous jobs

VOSA communicates with the computation server in an asynchronous way. That is, VOSA submits

a process and does not wait for it to finish. The main advantage of this capability is that the user does not need to wait, with the browser open, to the end of the process. Moreover, long queries are not affected by potential connectivity problems either. Users can start a process, close the computer and come back later to see how it is going. If it is finished, VOSA will show the results. If not, VOSA will give information on the status of the process and will provide an estimation of the remaining time. Moreover, processes can be canceled at any time from the VOSA web interface.

New Web interface

Typical Gaia queries to VOSA will involve files containing thousands of objects. In this scenario, the plain visualization in a web page is not applicable due to the highly demanding memory requirements. In order to avoid this problem, we have redesigned the presentation of large tables implementing a customizable pagination form. This new form shows only a subset of the results and offers the possibility of selecting a particular object, setting the number of object to be shown per page and going to a certain page.

Presentations

Meeting title: The Milky Way Unravelled by Gaia: GREAT Science from the Gaia Data Releases Place & Date: Barcelona. 1-5 December 2014 Title of the presentation: VOSA: SED building and analysis of thousands of stars in the framework of Gaia Authors: C. Rodrigo, E. Solano, A. Bayo Link to presentation: http://dx.doi.org/10.1051/eas/1567088

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; remaining allocated GENIUS travel funds will assist in communicating with the GENIUS project over the rest of its lifetime.

3.2.2.4.3 Highlights

- Completion of tests with prototype data mining frameworks at CSUC, allowing the selection of technologies for GDAF.
- Deployment of new data mining testbed at CSUC, allowing to scale-up tests.
- Formation of a large and representative data mining users group inside DPAC and identification of real use cases for the implementation of the data mining framework.
- Deployment of the visualization server at ESAC main archive completed. First tests successful with live connections to the archive using simulated data.
- Preparation of the visualization infrastructure for real operations; to be tested during rehearsal in Nov. 2015 and applied to the actual Gaia data release in mid-2016.

3.2.2.4.4 Deviations and impact on tasks and resources

There are not significant deviations in this work package. Just to indicate that the VO tasks initially assigned to CSIC have been actually undertaken by INTA personnel (affiliated institution); the proper accounting of these tasks at INTA requires the completion of the amendment of the DOW, which is ongoing with the PO agreement. Due to this the corresponding Person-Months are not yet accounted and will be included next year once the changes in the consortium compositon are completed (see section 3.2.3.2).

3.2.2.4.5 Use of resources

The following table lists the person-month per participant in the second year in WP4

	UB	**CSIC**	**FFCUL**	**UBR**
WP4	13	0.0	9.96	0.0

3.2.2.5 Work Package 5 Tools for data validation and analysis

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

3.2.2.5.1 Overview of WP objective

The preparation of the Gaia archive before its publication requires a careful, detailed and in-depth validation of its contents. The scientific and statistical challenge of this task on a one billion data set containing a wide variety of data (astrometric, photometric, spectrophotometric, spectroscopic, etc.) is daunting, and would be impossible without tools adapted to work on such a massive and data-diverse archive. This work package is producing 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. Therefore, this work package, in connection with WP4, produces tools for the use of the scientific community in its analysis of the Gaia data.

3.2.2.5.2 Summary of progress made

Task 5.1 - Technical coordination

Task leader: CNRS

Documents issued: (Arenou, FA-063), (Arenou, FA-064), (Arenou & Michalik, FA-066), Shih (IS-001), Shih (IS-004), Shih (IS-005)

Benefiting from an organisationnal 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 are being organised, one every about 1.5 months, the minutes of which can be found in the the CU9 WP940 Wiki page[9] [10] The telecon meetings with validation sub-WP managers were organised on:

- Telecon meeting 7 (2014-11-05)
- Telecon meeting 8 (2014-12-19)
- Telecon meeting 9 (2015-04-07)
- Telecon meeting 10 (2015-05-19)
- Telecon meeting 11 (2015-06-18)
- Telecon meeting 12 (2015-07-10)
- Telecon meeting 13 (2015-09-02)

[9] `http://wiki.cosmos.esa.int/gaia-dpac/index.php/WP940`
[10] and can also be found in Livelink documents (Arenou, FA-063) and (Arenou, FA-064)

- Telecon meeting 14 (2015-10-15)

Beside, progress meetings are 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 telecons, and thus, plenary meetings are also organised once per cycle (every six months). These plenary meetings were:

- Validation plenary I during CU9 plenary meeting in Vienna (2014-07-09)[11]
- Validation plenary II: workshop in Meudon (2015-03-02/04)[12]
- Validation plenary III during CU9 plenary meeting in Barcelona (2014-07-09)[13]

Some fundamental activities of the technical coordination are:

- to organise the input data, to run and deliver the tests done by the Work Packages. One key issue is ensuring the integration of the validation software at ESAC, which is needed as the software will not run on the Gaia data in distributed places (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 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, then successive data releases should run smoothly, and allowing as much as possible to optimise the computing resources.
- to develop as much as possible the common tools which may be needed by the various WPs, mainly statistical (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[14];
- The DPAC svn repository[15] for both code and documents;
- Hudson [16], which allows to check, after each software modification, the status of the JUnit tests of the validation.

[11] http://wiki.cosmos.esa.int/gaia-dpac/index.php/CU9:940:Plenary_2014
[12] http://wwwhip.obspm.fr/GaiaValidation2015/
[13] http://wiki.cosmos.esa.int/gaia-dpac/index.php/CU9:Plenary_2015_Validation
[14] http://wiki.cosmos.esa.int/gaia-dpac/index.php/CU9:940
[15] http://gaia.esac.esa.int/dpacsvn/DPAC/CU9/
[16] http://gaiahud.esac.esa.int/view/CU9/

- Mantis, now replaced by Jira, for software bug tracking,
- Nexus library repository, all directories being defined in an 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[17], the role of this Genius work packages is to develop the software implementing these tests. The current status of the test implementation (described in the following Tasks) is shown Figure 5.3.

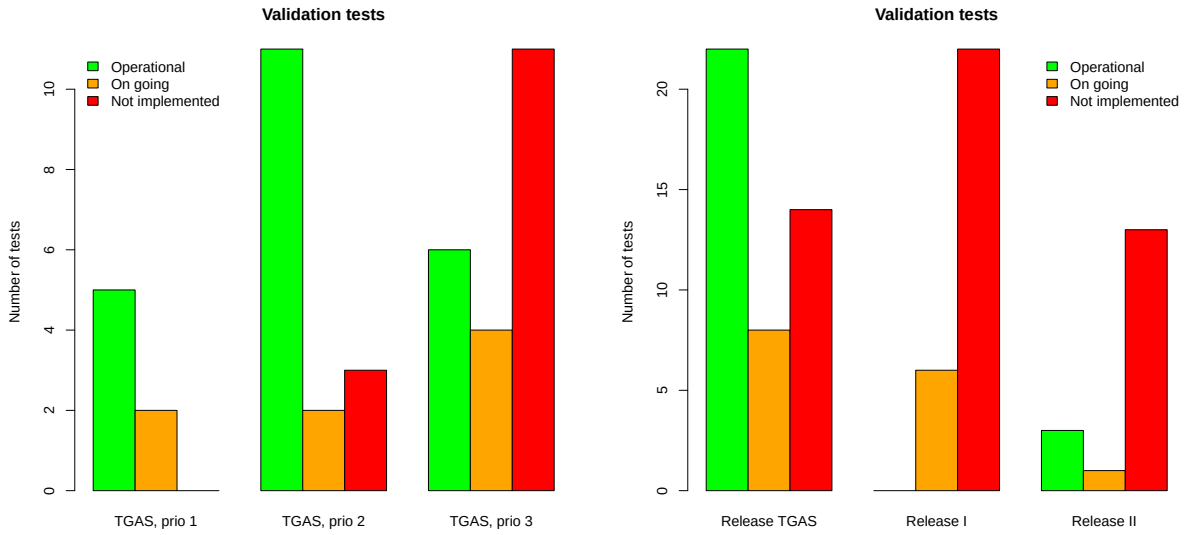


FIGURE 5.3: Software implementation of the validation tests, current situation for TGAS (with various priorities) and release I. Each test in a given release is also applied to the next ones, so that the tests indicated for release TGAS also apply e.g. to GRD2.

As indicated above, a common S/W environment has been developed, the ValidationTools project[18], which allows a consistent interface with the database or gbin files for all the validation users. Beside, 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 requested the development of common software, and the responsibility of this Task was extended to the supervision of the overall framework allowing to run the tests in a homogeneous way. A consequence of this is in terms of FTE is that a fraction of the CNRS involvement in each Task 1 to 6 has been devoted to putting in place this common environment which is (or will be) used by the various Tasks.

For example, within this coordination work, it was needed to generate simulated data in order to test the software[19].

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

[17] `http://gaia.esac.esa.int/dpacsvn/DPAC/CU9/docs/WP940_Validation/ECSS/VTS/`
[18] it is described in several documents, see Shih (IS-001), Shih (IS-004), Shih (IS-005)
[19] described in Arenou & Michalik (FA-066)

on this solution illustrates the progress which has been obtained thanks to GENIUS/CU9 activity, through the development of the needed S/W.

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

Task leader: CNRS

Documents issued: Antiche, E. et al (ELA-011), Antiche, E. et al (ELA-012), Julbe (FJL-002), Arenou & Fabricius (FA-062), Findeisen (KF-002), Antiche & Gutierrez (ELA-017)

This work package intends to define validation scenarios, and to implement the corresponding tests, after basic verifications of the Catalogue content have been performed to ensure that the field contents are as expected, 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. Other tests are also being developed in interface with GAT, a UB software for the statistical analysis of Gaia Catalogues (Antiche, E. et al , ELA-011), (Antiche, E. et al , ELA-012). All the planned tests for TGAS are now operational as planned:

- WP942-VAL-110-010: No fields equal NaN.
- WP942-VAL-110-020: No entries are null.
- WP942-VAL-110-030: All values fall within the specified range.
- WP942-VAL-110-050: Check that the astrometric correlation matrix is a valid correlation matrix
- WP942-VAL-120-020: Verify that the errors are consistent between each other.
- WP942-VAL-210-010: Verify that there are no duplicate sourceIds.
- WP942-VAL-210-020: Verify that there are no duplicates based on the object location.
- WP942-VAL-310-040: Check of parameter continuity at gate transitions
- WP942-VAL-310-080: Detection of groups of proper motion (PM) outliers
- WP942-VAL-310-090: Detection of isolated proper motion (PM) outliers
- WP942-VAL-310-100: Check consistency in proper motions (PM) and parallax
- WP942-VAL-310-120: Check consistency of astrometric error distributions
- WP942-VAL-340-010: Accurate formal parallax errors

The deliverable D5.3, Delivery of internal consistency checking tools, about to be issued, will summarize the work done, see also the other documents indicated above.

Task 5.3 - Simulation versus reality: from models to observables

Task leader: CNRS

Documents issued: Houri Ziaeepour et al. (HZ-001), Mor et al. (RMC-001)

The aim of these validation tests is 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, on stellar models and combined with empirical constraints and dynamical considerations.

The Besançon Galactic Model (BGM) simulates physical and dynamical properties of stars in the Milky May in a self-consistent manner. In addition, it includes phenomenological models for extinction. The model is expected 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.

WP943 (Task 5.3) validation tests share the same methodology for calculation of statistical properties and their comparison. The specification of tests can be found in Arenou et al. (FA-061) and their implementation in Houri Ziaeepour et al. (HZ-001).

The main tasks performed during the last year are:

- Realization of tests for proper motion and parallaxes
- Determination of reliability of the model by comparison with several catalogues, such as, SDSS, 2MASS, TYCHO, and RAVE
- Performing new simulations specifically for comparison with TGAS candidate release, with improved kinematical models.
- Establishing the level of thresholds for test decision (TRUE or FALSE) for different parameters (proper motions, parallaxes, star density, color distributions)
- Running tests on AGISLab simulations and on the preliminary TGAS solution.
- Report on the validation of TGAS for proper motions and parallaxes

The conclusions of the preliminary TGAS solution were the following:

Proper motion distribution validation The test on TGAS data of the l-proper motion distribution all pass, except in a magnitude bin where the number of bad pixels is a bit too large to be validated for the 4th moment of the b- proper motion. A problem in the kinematics in the present

model used could well explain this discrepancy and there is no warning on TGAS data based on this test. However there are also systematic deviations which occurs in both l-proper motion and b-proper motion which depends on latitude, but there are similar in TGAS and in AGISLab. These systematics should be looked at carefully, but are probably due to the model itself.

Parallax distribution validation The tests on the parallax distribution of TGAS generally pass well. However, the threshold have been set relatively high in order that the tests pass on AGISL, this is a good sign that the agreement between the model and TGAS is better than between the model and AGISL. We conclude that AGISL is not a good reference for setting the acceptation thresholds. However, the comparison of the parallax mean and standard deviation in different magnitude and latitude bins show a good agreement with the model. There is however a systematics in the mean difference of parallaxes between the model and TGAS data which also depends on magnitude. The difference is negligible at bright magnitudes but reaches up to 1 mas at magnitude 11. The difference is also slightly dependent on the latitude, which can be due to a systematic effect due to the scanning law. There is also a difference in the standard deviation of the parallaxes, being larger in TGAS data compared to the model. It can be due to underestimating the errors on parallaxes in the simulations. It should be considered in more detail in the next release candidate.

Task 5.4 - Confronting Gaia to external archives

Task leader: CNRS

Contributing partner: CSIC, KU

Deliverable D5.6 of GENIUS, Delivery of prototype of external validation tools
Documents issued: Ruiz-Dern et al. (LL:LRD-001)

One of the first uses of the Gaia data will be 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 to show the consistency between Gaia data and external data, and perform “external” validations.

The first work done in this Task has been to define the various tests in the VTS document. Then, to develop the code to implement the first tests, which includes using the cross-match feature at GACS from Java, on the fly, and this has been done.

The WP944 (Task 5.4) makes use of the CU9 ValidationTools software (IS-001), which provides global methods and tools for the whole Validation team. It also has its own structure for internal tests management, organization and statistics.

Each VTS returns a global flag (fail/pass/warning/exception/...) as well as a log, a detailed summary of the results and eventually some graphs. Only fail/pass are described for each VTS individually. If the test throws a java exception or ended with an error in R, it leads to an exception flag. If an exception is thrown due to the TAP service, it leads to an error flag. If the test was not conclusive (no fail or pass explicit indication), a warning flag is received.

All the tests developed so far are described in Ruiz-Dern et al. (LL:LRD-001).

We have adapted the already developed tests to the new VOTAP, called ValidationTools. We also have updated the VTS document to include new and adapted tests for TGAS. Then we developed the complete internal structure of WP944: crossmatches, tables and objects formats and management, global tools, global running. And we implemented code to run R scripts from Java.

Discussions with ESAC followed Mantis reported issues, due to crossmatch issues with new GACS version 0.5. Interaction and discussions have also occurred with WP946 (Task 5.6) to make the internal tools developed in WP944 (Task 5.4) available to other WP in Validation.

A large amount of time has been devoted to the preparation and selection of new small catalogues and groups of stars to confront with Gaia data: Cepheids, RR Lyrae, Rave, VLBI, Tycho2 and Simbad High Velocity stars. Then the tests have undergone improvements, in particular adapted to ValidationTools optimizations. Tests for TGAS are now ready and operational, already running at ESAC with real TGAS data:

- Priority 1:
  - Sky homogeneity (only duplications test case, WP944_Val_010_001)
  - Comparison to Hipparcos astrometry (WP944_Val_030_003)
  - Parallax Zero-Point using very distant stars (WP944_Val_030_006)
- Priority 2:
  - Known High Proper Motion stars (WP944_Val_030_004)
- Priority 3:
  - Known issues for specific stars (WP944_Val_010_003)
  - Parallax Zero-Point using external distances (WP944_Val_030_005)

We have filled a draft of Validation Test Report with all implemented tests and the results obtained using TGAS simulated data and Hipparcos data, and we have delivered the deliverable D5.6 of GENIUS (Comparison with external catalogues, prototype, Ruiz-Dern et al. (LL:LRD-001)).

Task 5.5 - Data demining: outlier analysis

Task leader: CNRS

Contributing partner: FFCUL

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

The goal of this Task is thus to develop tools which will allow to find outliers, or at least substructures which could then prove to actually be due to artifacts, not real structures. Two different developments are thus already going on.

It is planned to apply this tool to check across results from different CUs (e.g. parallaxes from astrometry, photometry, etc) for the correlations with errors, the characterization of relevant subspaces from models, and similarly, which subspaces should contain "no information", etc.

A mutual-information analysis of 2D-subspaces of the TGAS_00.01 catalogue has been carried out, and of the AgisLab Nominal and AgisLab Discontinuity simulations. The aim of this first exercise is to establish if the clustering/correlation properties of the TGAS dataset are consistent with those in the simulations.

After computing the Kullback-Leibler Divergence (KLD) statistic, the values obtained in the 3 datasets considered are ranked. If the correlation/clustering is similar, then the ranking should be similar. This has been found to be the case for many subspaces, but we also found important deviations, mostly in comparing the TGAS with the Discontinuity simulation. After examining the diagnostic plots, the Discontinuity simulation seems to overemphasize in specific large areas the results obtained by observing too few times (e.g. parallaxes and errors which are too large). On the other hand, what can be seen is the imprints of the scanning law and stripes-issues in the error distribution, for example for $\mu_{\alpha}*$ for the TGAS dataset.

The next steps of the work may include:

- A more realistic AgisLab simulation, and removing the effect of the scanning law (if possible), such that other effects/systematics may surface.
- TGAS vs TGAS comparisons, by taking subsamples of the data that one expects should behave similarly (e.g. symmetries wrt galactic latitude)
- Define a pass/fail criteria that reflects better what is learned (although it seems like we still need to do a lot of exploration before this can be done).

Task 5.6 - Transversal tools for special objects

Task leader: CNRS

Contributing partners: UNIGE, ULB

Documents issued: Vallenari et al. (AV-012), Vallenari et al. (AV-016), Cantat-Gaudin et al. (TCG-001)

Solar System Objects

Solar system objects are particular objects because they are moving with continuously varying velocity and their brightness is continuously changing because of both geometry and intrinsic properties. Observations may thus be corrupted e.g. because of a close approach to a star, or false alerts may also happen. Subtask 561 accordingly intends to validate what concerns solar system objects.

We performed simulations of Solar system object observed by Gaia over 4 years for statistical tests of completeness. Results of the simulations not yet fully analyzed. Tests (cross-match of Gaia

transit ID with asteroid ID and calculation of O-C) for CU4 SSO-ST output data are operational. We compare Gaia observed transits (as detected by CU4) to different predictions: 1) to the positions of all Astorb database asteroids calculated by the simulator DU460 (simulator developed within DPAC consortium); 2) to the positions of Astorb database asteroids calculated by means of Miriade ephemeris generator (a tool, developed within VO of Paris Observatory); Consistency of different simulators reaches 0.3 s for time transits (due to the slight differences into applied Gaia Scanning Laws) and upto 1" in (RA, Dec) due to the differences into integrators and Gaia orbits used in software). Unfortunately, due to internal problems in CU4 SSO-ST pipeline (difficulties in distinguishing real transits and spurious ones; loss of objects; etc.), we still find a lot of "problematic" runs. Nevertheless, test on individual astrometry shows good quality on the identified transits: uncertainty is dominated by IDT, i.e. about 70-100 mas per transit, as expected, wrt calculated positions based on the SL most adapted to the real one and the most recent Gaia orbits).

Stellar clusters

On another aspect of "special objects", the work on stellar clusters, mostly done at Osservatorio Astronomico di Padova with a CNRS contribution, is now in very good shape, after a definition of the external Catalogues for cluster selection, definition of the criteria for cluster selection and preliminary selection of the target clusters for the first data release, see Vallenari et al. (AV-012), Vallenari et al. (AV-016), Cantat-Gaudin et al. (TCG-001). Tests have been implemented for parallaxes and proper motions, in order to detect any zero-point deviation (not found). An underestimation of the formal errors together with a μ_δ trend with α have been found, and this is consistent with what has been found by Task 5.4.

Multiple Stars

One group of non-single stars might eventually lead to some precise stellar masses: those belonging to spectroscopic (SB2) eclipsing binaries. In CU4, these combined classes are first process independently (i.e. spectroscopic binaries on one side, eclipsing binaries on the other) before they get combined into a unique solution. In order to validate both these initial steps and the combining one, some binaries were selected from the literature. A sample of sixty systems was assembled and the CU4 solutions compared with the literature ones. The ULB team made it sure that the spectroscopic data were available on SB9.

With just the G-band photometry available, the eclipsing binary solutions turned out to be of limited value. Indeed, with one photometric band only, the temperature scale cannot be set for both components. Even though the principle of the validation of these solutions with those 60 systems will remain valid through out the mission, it is not applicable yet due to the limited datasets available.

Variable Stars

Objects with intrinsic or extrinsic variability (such as Cepheids and eclipsing binaries) may affect the Gaia data analysis. For instance, variations in luminosity difficult the cross matching of sources, leading to a wrong determination of physical parameters. In the opposite sense, the instrument and/or the data processing can also introduce false variability that might be interpreted as real. UNIGE has taken this aspects into consideration to implement a set of tests that verify no significant statistical biases are present on the Gaia catalogue.

The tests were developed using the Hipparcos catalogue as main reference for differentiating constant sources from variable ones. These two groups are compared to validate:

- The mean parallax difference in both groups is zero (Student's t-test)
- Both groups have the same variance for parallax difference (F-test)
- Both groups have the same distribution (Kolmogorov & Smirnov test)

These validations are specially designed for the TGAS release, which will not contain variable information. They were developed using the CU9 Validation tools software, and tested using simulated data. Wherever it has been possible, code was shared with other tests in order to reduce duplicity and improve the reliability.

Additionally, a list of variable stars with Tycho identifiers has been built to help the rest of WP/Tasks leaders on the comprehension of outliers.

3.2.2.5.3 Highlights

Test	Result	Reason
WP942-110-010	FAILED	Found 56 sources with NaN fields in Gbin data.
WP942-110-020	FAILED	Found 380 sources with null fields in Gbin data.
WP942-110-030	FAILED	Found 53195 sources with out-of-range fields.
WP942-110-050	FAILED	Found 128 sources with invalid correlation matrices.
WP942-120-020	FAILED	Found 17,700 sources with mismatched errors.
WP942-210-010	PASSED	No duplicate sourceIds in Gbin data.
WP942-210-020	FAILED	Found 11 pairs of sources with nearly identical positions.
WP942-310-040	MIXED	No rapid changes were found in gate transitions
WP942-310-080	FAILED	Regions with proper motions ratio exceeding the thresholds were found.
WP942-310-090	PASSED	None isolated sources were found above 4 times the escape velocity.
WP942-310-100	FAILED	Regions are likely to containg systematic effects in muAlpha, muDelta and parallax
WP942-310-120	FAILED	Mean error larger than expected.
WP942-340-010	FAILED	Very large parallax errors
WP943-030-010 mean l-proper motion	FAILED	in north hemisphere and faint magnitude
WP943-030-010 mean b-proper motion	PASSED	only in the faintest magnitude bin
WP943-030-010 stdev l-proper motion	FAILED	in north hemisphere and faint magnitude
WP943-030-010 stdev b-proper motion	PASSED	only at VT$>$11 and south Galactic pole
WP943-030-010 skewness l-proper motion	PASSED	only 1 magnitude/latitude bin
WP943-030-010 skewness b-proper motion	PASSED	only 1 magnitude/latitude bin
WP943-030-010 kurtosis l-proper motion	PASSED	only at the galactic poles
WP943-030-010 kurtosis b-proper motion	FAILED	only in the faintest magnitude bin
WP943-040-010 mean parallaxe	PASSED	No problem
WP943-040-010 parallaxe stdev	PASSED	No problem
WP944-010_001_Tycho2Duplicates	MIXED	no duplicates but tycId list not complete
WP944-010_003_hip67626	PASSED	no tycho2 source corresponding
WP944-030_001_Parallaxes-VLBI_tgas_CHI2	MIXED	to be checked
WP944-030_001_Parallaxes-VLBI_tgas_PLX	MIXED	plx marginally inconsistant
WP944-030_001_Parallaxes-VLBI_tgas_PM	FAILED	pm inconsistant
WP944-030_003_goodHip_PLX	FAILED	small bias + extra dispersion
WP944-030_003_goodHip_PM	FAILED	bias + correlation with mag
WP944-030_003_badchi2HipTycho2	PASSED	solution closer to the tycho2 one
WP944-030_004_tycho2_HV	MIXED	proper motion agrees with tycho2 ones
WP944-030_004_simbadHV_tgas	FAILED	84% of HPM stars not in TGAS
WP944-030_004_TGASHPM	FAILED	6 TGAS HPM without valid tycId
WP944-030_005_Ngeow_tgas	FAILED	plx errors under-estimated
WP944-030_005_RRlyrae	FAILED	plx errors under-estimated
WP944-030_005_rave_tgas	FAILED	stars with 0 plx errors...
WP944-030_006_LMCSMC_hipconfirmed	FAILED	bias detected + extra dispersion
WP944-030_006_SMC_tgas	FAILED	bias detected + extra dispersion
WP944-030_006_LMC_tgas	FAILED	bias detected + extra dispersion
WP944-060_003_TGAS_AGN	PASSED	1 AGN
WP947-010-010	PASSED	Proper motion zero point consistent with reference values
WP947-010-030	PASSED	Parallax zero point (bias+dispersion) consistent with photometric estimate
WP947-010-032	FAILED	Distant cluster parallaxes have larger dispersion than expected

FIGURE 5.4: Validation tests which have been run against the preliminary TGAS data and main result found.

- Two full validation sessions run on TGAS preliminary data
- Six over the seven Work Packages have participated
- 44 tests run, see Fig. 5.4
- Common validation S/W environment operational

The most important point is the fact that the validation tests have been implemented in the frame of the GENIUS/CU9 project and have now run for the first time on the actual (preliminary) data. The outcome of the tests (see Fig. 5.4) has been transmitted to DPAC/CU3/AGIS (and also to the validation Test Review Board) and is now being studied for improving the next solution. The GENIUS project has given the impetus which allowed this to be possible in a short time scale.

3.2.2.5.4 Deviations and impact on tasks and resources

The deliverable D5.3, Delivery of internal consistency checking tools, (Task 5.2) planned for early October will be one month late. In effect, the software in itself has been delivered, and it run for two rehearsals using preliminary data. Time has to be devoted to these rehearsals which prevented to issue the accompanying documentation in due time.

3.2.2.5.5 Use of resources

The following table lists the person-month per participant in the second year in WP5

	CNRS	**CSIC**	**UNIGE**	**ULB**	**FFCUL**	**KU**
WP5	30.42	0	**??**	1.5	0	0.5

Note: resources information for UNIGE has not been received to the date of this draft. Will be added in the final version.

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 aims 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 will 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 will be installed in it for testing and validation and made accessible to the test users.

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

3.2.2.6.2 Summary of progress made

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

The two tasks in this WP are mostly independent and therefore the coordination has been centered on the tracking of progress through telecons, meetings and reports.

The activity on simulations has been lighter during the second year since the execution of simulations at CSUC has been reduced due to the changes in requirements and schedule from the DPAC developments (see Task 6.2). Accordingly, as agreed during the first year review, resources from the simulation activities were transferred to the data mining task.

On the other hand, the definition of requirements for the science alerts has continued and the testbed is now available. Its integration into the realtime Gaia alerts system is progressing.

Task 6.2 - Simulated catalogue data

Task leader: CSUC

Contributing partner: UB

Essentially, only minor simulations have been executed during this second year, postponing the second large scale simulation to 2016 to match the needs of the DPAC. Furthermore, since Gaia data is becoming a reference for other missions, we have received request for usage of the simulations for PLATO and EUCLID; we are discussing now, in coordination with DPAC-CU2, how to fulfill the overall needs with the next year simulations. Meanwhile the error models have been updated to account for the knowledge of the real behaviour of the instruments.

Task 6.3 - Science alerts testbed

Task leader: UCAM

Overview

The Gaia flux-based science alert stream has been issued to the community through the science alert 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-6.3 is developing the interfaces required to connect the real time science alerts classification processing to the main Gaia data products. 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. Thus 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 will in addition provide linkages to external data resources provided through GENIUS, in particular via interfaces to the archive development through WP3. Finally the alerts testbed will plugin to the portal testbed developed in WP7.2. With the termination of the GENIUS WP6.3 testbed activity, the full functionality will be deployed for community use - providing enhanced access to science alert data from 2015 onwards.

Summary of progress in year 2

Year two activity has involved further requirements analysis. The testbed involves the integration of the realtime alerts from the Gaia Alerts stream (from the Gaia/DPAC/CU5), for longer term curation 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 Jun 2015. See `http://gaia.ac.uk/selected-gaia-science-alerts`. This was the initial Gaia Alerts testbed.

Deployment of first public science alerts prototype[20]

The second testbed (deliverable D6.3 Deployment of first public science alerts prototype) is available now in validation mode. It will be released as part of the full operational alert system towards the end 2015. This pause in operations of the photometric science alerts system (July-November 2015) is 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 alerts prototype will therefore be releasing photometric alerts to the community from end

[20] See alerts release at `http://gaia.ac.uk/selected-gaia-science-alerts`

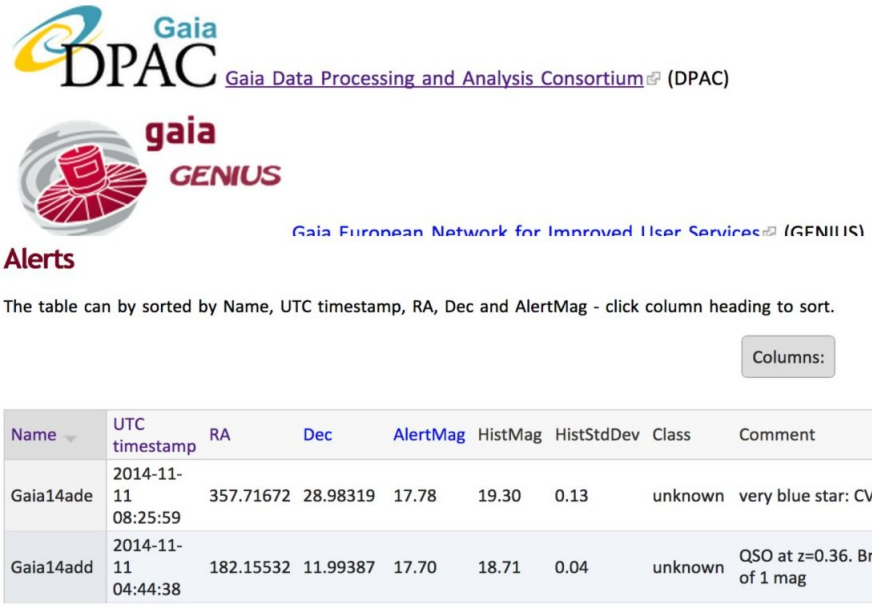
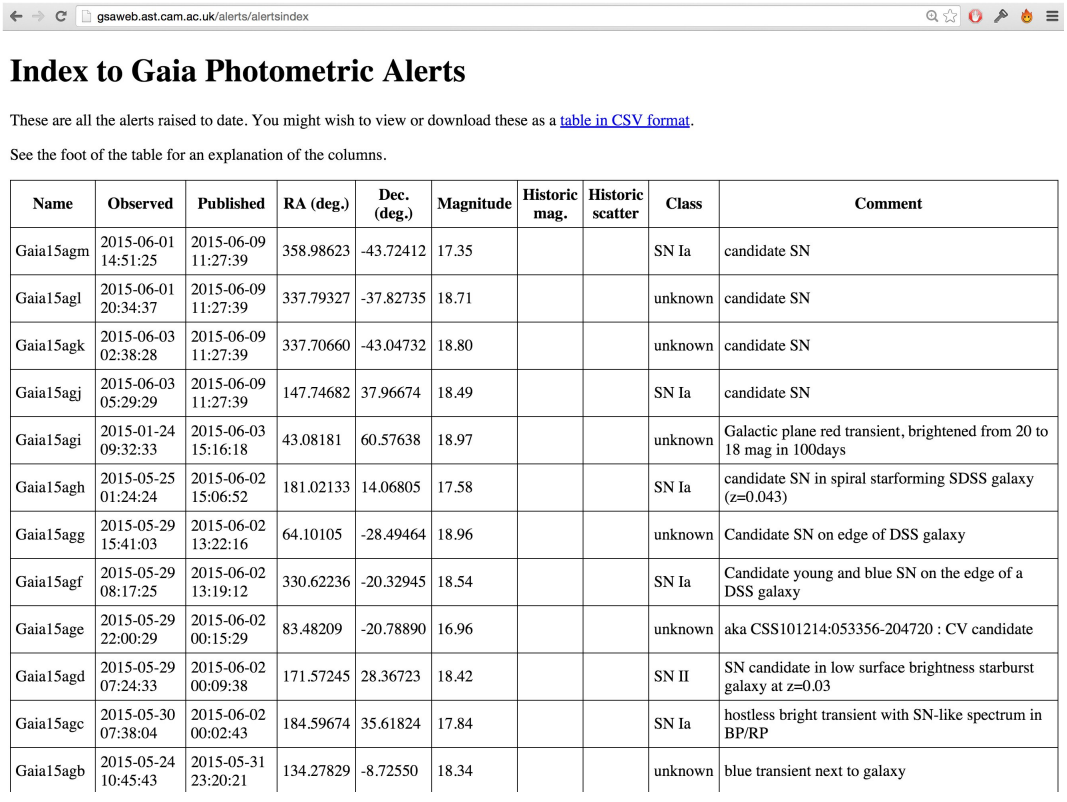


FIGURE 6.5: The Science alerts portal acting as the initial testbed for the CU9/GENIUS alerts archive



Name	Observed	Published	RA (deg.)	Dec. (deg.)	Magnitude	Historic mag.	Historic scatter	Class	Comment
Gaia15agm	2015-06-01 14:51:25	2015-06-09 11:27:39	358.98623	-43.72412	17.35			SN Ia	candidate SN
Gaia15agl	2015-06-01 20:34:37	2015-06-09 11:27:39	337.79327	-37.82735	18.71			unknown	candidate SN
Gaia15agk	2015-06-03 02:38:28	2015-06-09 11:27:39	337.70660	-43.04732	18.80			unknown	candidate SN
Gaia15agj	2015-06-03 05:29:29	2015-06-09 11:27:39	147.74682	37.96674	18.49			SN Ia	candidate SN
Gaia15agi	2015-01-24 09:32:33	2015-06-03 15:16:18	43.08181	60.57638	18.97			unknown	Galactic plane red transient, brightened from 20 to 18 mag in 100days
Gaia15agh	2015-05-25 01:24:24	2015-06-02 15:06:52	181.02133	14.06805	17.58			SN Ia	candidate SN in spiral starforming SDSS galaxy (z=0.043)
Gaia15agg	2015-05-29 15:41:03	2015-06-02 13:22:16	64.10105	-28.49464	18.96			unknown	Candidate SN on edge of DSS galaxy
Gaia15agf	2015-05-29 08:17:25	2015-06-02 13:19:12	330.62236	-20.32945	18.54			SN Ia	Candidate young and blue SN on the edge of a DSS galaxy
Gaia15age	2015-05-29 22:00:29	2015-06-02 00:15:29	83.48209	-20.78890	16.96			unknown	aka CSS101214:053356-204720 : CV candidate
Gaia15agd	2015-05-29 07:24:33	2015-06-02 00:09:38	171.57245	28.36723	18.42			SN II	SN candidate in low surface brightness starburst galaxy at z=0.03
Gaia15agc	2015-05-30 07:38:04	2015-06-02 00:02:43	184.59674	35.61824	17.84			SN Ia	hostless bright transient with SN-like spectrum in BP/RP
Gaia15agb	2015-05-24 10:45:43	2015-05-31 23:20:21	134.27829	-8.72550	18.34			unknown	blue transient next to galaxy

FIGURE 6.6: Deliverable D6.3 Release alert publishing page

2015. This will be coordinated with a significant follow up campaign, obtaining ground based follow up of the alerts in order to characterise the alerts (thus type then as supernova, flare stars, CV's etc). There has been development of a prototype tool to support this follow-up community, developed within the context of Gaia DPAC CU9, thus not reported here.

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.

Use of resources in Year 2

Only 0.5 months of UCAM staff effort has been charged to GENIUS WP6 in Year 2, with this activity being requirements analysis and definition of the data model between the alerts operational system and the alerts CU9 main database. In addition activity related to attendance at GENIUS management telecons and the annual plenary meeting.

In addition, travel costs related to attendance at the GENIUS 2015 Plenary meeting (Barcelona, Sep 2015) have been charged.

All other science alerts activities have been carried out through the DPAC CU9 and CU5.

We note that the bulk of GENIUS resourced development work will occur in the later phase of the project, namely years 3 and 4. This reflects the likely release schedule of packaged periodic releases of the Gaia Photometric Alerts.

3.2.2.6.3 Highlights

- Reception of the first real Gaia alerts.
- Second science alerts testbed is available now in validation mode.

3.2.2.6.4 Deviations and impact on tasks and resources

As discussed above, the large scale simulations scheduled for this year have been postponed to next 2016. This is tied to the actual needs of the Gaia data processing once the satellite is operational. On the one hand, the availability of real data from the satellite has improved our knowledge of the instruments (including several non-nominal effects like stray light); this knowledge has taken time to develop and will now be included into the simulations through the error models. On the other hand, the Gaia data processing schedule itself has evolved with time, with several delays due to various factors. Accordingly the generation of simulations has been rescheduled as mentioned above.

3.2.2.6.5 Use of resources

The following table lists the person-month per participant in the second year in WP6

	UB	**CSUC**	**UCAM**
WP6	1	3.5	0.5

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.).

3.2.2.7.2 Summary of progress made

Task 7.1 - Coordination of dissemination activities

Participants: UB, CNRS

The main activity of the WP7 during 2015 has been the release of a new version of Gaiaverse, the community portal of GENIUS. We have created an Editorial Board in charge of defining and supervising the contents and the contact with collaborators. The current members of the Editorial Board are Frédéric Arenou (Obs. Meudon), Heather Campbell (Cambridge), Gráinne Costigan (Univ. Leiden), Mariateresa Crosta (INAF), Eduard Masana (Univ. Barcelona) and Teresa Via (CSUC). The Editorial Board held two teleconferences during 2015 to start its work in Gaiaverse.

The new version of the gaiaverse.eu portal (see Task 7.2) was introduced to the Gaia community in the CU9 plenary meeting held in Barcelona in September 2015.

Several calls to open the portal to new languages gave as a result the possibility (to be confirmed) to translate the portal to Greek and Portuguese. Also in the context of Gaiaverse, we made an effort to compile all existing outreach material, including videos, presentations, documents and tools. This material was included in the corresponding sections of Gaiaverse.

Task 7.2 - Community portal infrastructure

Task leader: CSUC (formerly CESCA)

After the release of the first public version of the community portal in October 2014, an advanced version of this portal has been released in July 2015. This new version, named gaiaverse.eu, has been mainly prepared in collaboration with the Editorial Board but also with the contributions of members of DPAC (Data Processing and Analysis Consortium) and ESA (European Space Agency).

The first public version of the community portal included those sections agreed using a voting system and it was released in four languages (English, German, Spanish and Catalan), as it is shown in the figure below.

The advanced version of gaiaverse.eu has a new layout (see figure 7.8) to better highlight the contents (available resources, tools on Gaia,...) as well as the news related to the mission. A new

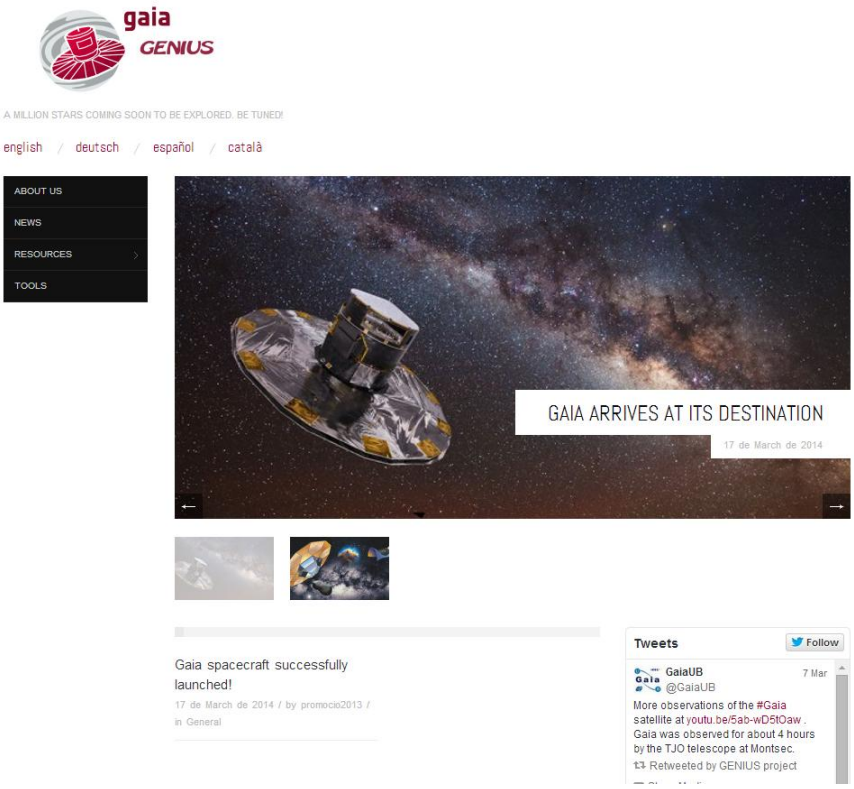


FIGURE 7.7: Community portal: First public version.

system of filters allows now searching within the available resources not only by type of content (video, image, presentation) but also by language. This new functionality facilitates searches on the portal and gives more visibility to available resources. By the end of September, gaiaverse.eu gives visibility to around 70 resources in 6 languages.

The new gaiaverse.eu website gives more flexibility to editors also to have different contents in each language. With regard to available languages, two more have been added to the advanced version of gaiaverse.eu (French and Italian) and new teams are working to have gaiaverse.eu in Greek and Portuguese as well.

In line with the new brand name and the advanced version of the community portal, a new twitter profile has been created also to strengthen dissemination activities. Visibility on Twitter has improved since the new profile was created in July 2015. If we consider number of impressions, let's say, number of times that users saw the tweet on Twitter, @gaiaverse got more than 11,800 impressions in less than 3 months whereas @genius_pr had less than 600 impressions in 4 months.

In addition to improvements on the dissemination portal, gaiaverse.eu, the project's portal (http://genius-euproject.eu), which contains basic general information only about the project itself, has been updated also to include information about the current gaiaverse.eu portal.

Task 7.3 - Community portal, outreach and academic activities

Task leader: UB

FIGURE 7.8: New gaiaverse.eu layout and new functionalities for searches.

Contributing, partner: CNRS

As in the previous year, outreach and academic activities around Gaia during 2015 include several talks and conferences for students and general public.

On the other hand, the outreach activities for the Gaia DR1 (Data Release 1) are being currently defined. After having discussed many ideas during the outreach splinter meeting during the CU9 plenary in September, we plan to concrete the tasks during the DPAC consortium meeting in November.

The ellaboration of a Simple Query Interface, designed for quick access in a clean and easy fashion is one of the projects currently under development for DR1 (G. Costigan). The interface will return a few key parameters of a source entered. The archive should be accessible by as many users as possible, both amateurs and professionals. This simple user interface should be a point where limited, but relevant information on a single target can be retrieved from the archive.

3.2.2.7.3 Highlights

FIGURE 7.9: GENIUS project portal homepage.

- New version of the gaiaverse.eu portal.
- Creation of the Gaiaverse Editorial Board
- Definition of the outreach activities for the Gaia DR1 (Data Release 1).

3.2.2.7.4 Deviations and impact on tasks and resources

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

3.2.2.7.5 Use of resources

The following table lists the person-month per participant in the second year in WP7

	UB	**CSUC**	**CNRS**
WP7	2.5	4.5	0

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 is, by necessity, tightly integrated into the already existing structure of the Gaia Data Processing and Analysis Consortium (DPAC), and specifically into its Coordination Unit 9 (CU9), 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 occurred and applied solutions

During this period there have been recurring administrative problems with the CSIC partner. Its participation was vehiculated through an affiliated institution, the INTA, where most of the participation in GENIUS are concentrated (including Enrique Solano, the main responsible of the contribution). Legal and administrative issues in the relationship between these two institutions have prevented a proper accounting of the contribution of the permanent staff and the formalisation of the contracts allocated for this partner.

These problems have had an impact on WP3, delaying some of the developments to next year (see Section 3.2.2.3) and a minor impact in WP4 (see Section 3.2.2.3). This situation, if continued, would create a significant impact in these areas of GENIUS; therefore, the UB in its role of coordinator has undertaken two mitigation actions:

- Amendmentof the DOW: removing the CSIC as GENIUS partner and directly including INTA
- Reallocation of budget from CSIC/INTA to UB in order to immediately issue the contracts assigned to this task. These contracts will be in place from 1st January onwards, allowing to start in force the required developments.

3.2.3.3 Changes in the consortium

Dr. Solano has been part of the project from its beginning. Dr. Solano works for Centro de Astrobiología (CAB) a combined centre of two institutions CSIC and INTA, but he is formally contracted by INTA: At the signature of the project, CSIC did not show any inconvenience in

signing the project led by Dr. Solano but at the time of the GENIUS first report CSIC informed us that the dedicated time by Dr. Solano that was included in the project, could not be charged to the project as Dr. Solano is not part of their personnel.

After many discussions with all parts involved, and a first proposal of including INTA as a Third Party of CSIC, CSIC decided last September that the most appropriate solution is to be withdrawn from the project. A year has passed and Dr. Solano and collaborators have been travelling and charging their expenses to the project. This results in the CSIC withdrawing from the project with date 01/10/2015, after justifying the expenses incurred in the second year (reduced to travel expenses). Now we are finally compiling the documentation needed to fulfil the necessary amendment, where we expect the inclusion of INTA will be from the first day of the project to be able to include the work developed by Dr. Solano.

We will as well update the change of name of CSUC (formerly CESCA) in this amendment.

We understand that all these changes do not affect 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
Gaia CU3 AGIS meeting	Nice	23-28/11/2014	Y. Yamada, A. Brown, F. Arenou
Data Mining meeting	ESAC (Madrid)	27/10/2014	X. Luri, F. Julbe
Gaia Challenge workshop (User requirements)	Max Plank Institute for Astronomy (Heidelberg)	27-31/10/2014	G. Costigan and A. Hypki
Meeting of the INAF contributors to WP2 GENIUS	ASDC (Rome)	04-05/11/2014	INAF contributors
Data Mining Meeting	OATo (Torino)	6-7/11/2014	L.M. Sarro
Gaia CU7 Meeting	ISDC (Geneva)	12-14/11/2014	L.M. Sarro, L. Eyer
Strata+Hadoop World Meeting	Barcelona	19/11/2014	F. Julbe
The Milky Way Unravelled by Gaia. GENIUS Coordination Meeting	Barcelona	01-05/12/2014	E. Solano, X. Luri, F. Arenou, L. Balaguer, A. Brown, L. Eyer, N. Walton, A. Robin, J. Torra, E. Masana
First Annual Review Report	REA (Brussels)	16/12/2014	X. Luri, N. Benitez, A. Brown, F. Arenou, N. Hambly, L. Balaguer
Data Mining Meeting	ESAC (Madrid)	22/12/2014	F. Julbe, L. Sarro, R. Borrachero
GENIUS Gaia Archive Review	ESAC (Madrid)	20-21/01/2015	X. Luri, G. Costigan
Data Clasification tests workshop	Geneva	22-25/02/2015	L. Eyer, L.M. Sarro, M. Lopez
First Cross Match Meeting	ASDC (Rome)	26-27/02/2015	X. Luri, R. Smart, P. Marese, N. Hambly
Nano-JASMINE Science Workshop	National Astronomical Observatory of Japan (Tokyo)	16/02/2015	Y. Yamada, R. Nishi
Three days visit to discuss the adoption of the Object Server API in Gaia Sandbox	Lisboa	02/2015	T. Sagrista, S. Jordan

Gaia CU9 Validation Workshop	Observatoire de Paris-Meudon	02-04/03/2015	F. Arenou, S. Blanco-Cuaresma, A. Brown, K. Findeisen, M. Fouesneau, H. Ziaeepour, A. Krone-Martins, M. Kudryashova, C. Reylé, A. Robin, L. Ruiz-Dern, I. Shih et al.
Nano-JASMINE data analysis review	National Astronomical Observatory of Japan, Mitaka, Tokyo	10/03/2015	Y. Yamada, S. Hozumi
Nano-JASMINE project review	The University of Tokyo, Bunkyo ward, Tokyo	17/03/2015	Y. Yamada
Meeting for the integration of OCA software (Object Clustering Analysis) in SAGA/CNES	CNES (Toulouse	17-19/03/2015	L.M. Sarro
Visit, talk and meeting with Parameter Space	University College Dublin	07/04/2015	G. Costigan
GAVIP System Concept workshop	ESAC (Madrid)	14-15/04/2015	G. Costigan, X. Luri
Gaia CU3 AGIS meeting	Heidelberg	08-10/04/2015	Y. Yamada, A. Brown
Data Mining ESAC Work Meeting	ESAC (Madrid)	14-15/05/2015	X. Luri, F. Julbe
20th Gaia CU7 Meeting	Lisbon	20-22/05/2015	A. Moitinho, S. Blanco-Cuaresma, L. Eyer, L.M. Sarro, M. López
IVOA Interoperability meeting	Sesto (Italy)	14-19/06/2015	S. Voutsinas
Summer School: New Era of the Cosmic Distance Scale	University of Tokyo	29/06-03/07/2015	Y. Yamada, F. Mignard and G. Gilmore
Workshop on future space astrometry missions	Cambridge	6-8/07/2015	F. Arenou, A. Brown
Dutch Gaia day	Nijmegen, Netherlands	03/09/2015	C. Costigan, A. Brown
37th Polish Astronomical Society Assembly	Poznán, Poland	07–10/09/2015	G. Costigan, A. Hypki
Gaia CU8 Plenary Meeting	Uppsala	15-17/09/2015	L.M. Sarro
Joint GENIUS CU9 Plenary Meeting	University of Barcelona	21-23/09/2015	ALL (see complete list in webpage)
Gaia CU3 AGIS meeting 24	Leiden	13-14/10/2015	F. Arenou, A. Brown

3.2.3.5 Project planning and status

As described in the GENIUS proposal, section 1.3, the execution of the project is based on a cyclic development where several prototypes are produced, each one building on the experience of the previous. During the second year of the project this approach has continued, in close coordination with the develiopments 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 has been established. In the later case this has allowed the execution of the validation software on preliminary Gaia results, producing the first validation of real Gaia astrometric data. [21].

In other areas the developments have been more independent of the Gaia archive, although also tied to it. It is the case of the Visualization, VO tools and Data mining tools for instance. These areas have also progressed as expected, with mature prototypes or fully functional versions available.

On the other hand, the GENIUS work on dissemination has made a significant contribution to the outreach of Gaia. The Gaiaverse GENIUS portal is becoming a hub for multi-lingual dissemination of Gaia and GENIUS is significantly contributing to many outreach activities organized around the mission and its data. For instance, GENIUS will participate in 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 meetings contribute to the strenghtening of the collaboration between scientific teams all around Europe.

Finally, GENIUS is now preparing for two crucial activities that will start in the coming months: a full rehearsal of the archive system (November-December 2015) and the first Gaia data release (summer 2016). During these events the GENIUS contributions will for the first time materialise into operational systems.

3.2.3.6 Impact of deviations from the planned milestones and deliverables

Delayed deliverables

D1.5 Plenary meeting to take place in Sept in Barcelona. Mid term review delayed to Nov. 2015 in Leiden (agreed with PO during the 1st year review) in combinatuon with the DPAC plenary meeting. No impact.

D1.6 Wrongly assumed that semestral report could be superseded by 2nd year report. Will be delivered shortly after 2nd year review.

D2.5 The user requirements analysis is still ongoing with an update expected in the summer of 2016 once the first Gaia data release is available and feedback can gathered from users. Can potentially cause delays in project, will be tracked by GENIUS executive board.

D5.3 Task has been completed but report is pending. Document delivery will be tracked by GENIUS coordinator.

[21] These results and the Gaia data itself are subject to limited distribution following the ESA agreed rules and can not be included in the GENIUS reporting

D6.3 The need of Gaia simulations for the reduction consortium has changed from the initial previsions. An updated error model is being developed, taking into account the current knowledge of the instruments and second catalogue data has been delayed until 2016; it will cover not only Gaia needs but also support to other missions like PLATO and EUCLID. No impact on GENIUS developments.

Delayed milestones

MS8 Mid term review delayed to Nov. 2015 in Leiden (agreed with PO during the 1st year review) in combinatuon with the DPAC plenary meeting. No impact.

MS9 Prototype archive not available in time for review. Delayed by 3 to 6 months in order to synchronise with the DPAC CU9 Gaia DR1 rehearsal and beta?testing of GACS. Minor impact (small delay in tailoring activities; can catch up during next year).

MS12 Prototype archive not available, can not open to community. Planned opening to DPAC using simulated data towars end 2015 (TBC by DPAC). Minor impact, development not stopped.

3.2.3.7 Changes to the legal status of beneficiaries

The inclusion of a new partner (INTA), and the withdrawal of another (CSIC) are explained in detail in section 3.2.3.3.

3.2.3.8 Use of resources

The following table lists the person-month per participant in the second year in WP1

	WP1	WP2	WP3	WP4	WP5	WP6	WP7
UB	5	-	-	13	-	1	2.5
CNRS	-	-	0	-	30.42	-	0
UEDIN	-	-	11.25	-	-	-	-
UL	-	25.4	-	-	-	-	-
CSUC	-	-	-	-	-	3.5	4.5
INAF	-	3	9.12	-	-	-	-
CSIC	-	-	0	0	0	-	-
UNIGE	-	-	-	.	**??**	-	-
ULB	-	-	-	-	1.5	-	-
FFCUL	-	0	-	9.96	0	-	-
UBR	-	-	-	0	-	-	-
UCAM	-	0.5	-	-	-	0.5	-
KU	-	1	-	-	0.5	-	-

Note: resources information for UNIGE has not been received to the date of this draft. Will be added in the final version.

The following table lists the actual efforts (**AE**) for the second period against the theoretical effort for the full lenght of the project (**TE**)

	WP1	**WP2**	**WP3**	**WP4**	**WP5**	**WP6**	**WP7**
	AE / TE	AE / TE	AE / TE	AE / TE	AE / TE	AE / TE	AE / TE
UB	5/18	-	-	13/40	-	1/6	2.5/7
CNRS	-	-	0/1.8	-	30.42/77.6	-	0/1.8
UEDIN	-	-	11.25/43	-	-	-	-
UL	-	25.4/49	-	-	-	-	-
CSUC	-	-	-	-	-	3.5/12	4.5/12
INAF	-	3/16	9.12/18	-	-	-	-
CSIC	-	-	0/6	0/12	0/4	-	-
UNIGE	-	-	-	-	**??**/6.5	-	-
ULB	-	-	-	-	1.5/6	-	-
FFCUL	-	0/2	-	9.96/28	0/2	-	-
UBR	-	-	-	0/3	-	-	-
UCAM	-	0.5/4	-	-	-	0.5/12	-
KU	-	1/2	-	-	0.5/2	-	-

Note: resources information for UNIGE has not been received to the date of this draft. Will be added in the final version.

3.2.3.9 Dissemination & Development of the project Website

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.

3.3 Deliverables and Milestones

3.3.1 Defliverables submitted in the second year 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	DELAYED Nov 2015
D1.6	Semestral report 4	WP1	Oct 2015
D2.3	Requirements specification for generic projection module	WP2	Oct 2015
D2.4	Requirements specification for data retrieval across archives	WP2	Oct 2015
D2.5	Conclusion of requirements update gathering exercise	WP2	DELAYED Jul 2016
D4.3	Delivery of second prototype of exploitation tools	WP4	Oct 2015
D5.3	Delivery of internal consistency checking tools (WP 520)	WP5	Oct 2015
D5.6	Delivery of prototype of external validation tools (WP 540)	WP5	Oct 2015
D6.3	Delivery of second simulated catalogue data	WP6	DELAYED
D6.4	Deployment of second public science alerts prototype	WP6	Oct 2015
D7.3	Upgraded public version of the community portal	WP7	Oct 2015

3.3.2 Milestones in the second year 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
MS5	Archive user requirements document	WP2	Nov 2014
MS6	Requirements document for each subsystem	WP2, WP3, WP4, WP5	Nov 2014
MS7	Public version of GENIUS portal	WP6, WP7	Jul 2015
MS8	Mid-term review	WP1, WP2, WP3, WP4, WP5, WP6, WP7	Nov 2015
MS9	User prototype archive review	WP2, WP3	Delayed 2016
MS10	Exploitation tools review	WP2, WP4	Sep 2015
MS11	Validation tools review	WP2, WP5	Sep 2015
MS12	Prototype archive tools open to community	WP1, WP2, WP3, WP4, WP5, WP6, WP7	Delayed 2016

References

[ELA-017], Antiche, E., Gutierrez, R., 2015, *Statistical validation of Gaia Archive*,
`GAIA-C9-TN-UB-ELA-017`,
URL `http://www.rssd.esa.int/cs/livelink/open/3311627`

[ELA-012], Antiche, E.C., Castañeda. J., Fabricius, C., 2014, *GAT Statistics for the new Attitude Star catalogue*,
`GAIA-C9-TN-UB-ELA-012`,
URL `http://www.rssd.esa.int/cs/livelink/open/3261151`

[FA-063], Arenou, F., 2014, *Minutes of CU9 validation telecons 1-2*,
`GAIA-C9-MN-OPM-FA-063`,
URL `http://www.rssd.esa.int/cs/livelink/open/3312224`

[FA-064], Arenou, F., 2014, *Minutes of CU9 validation telecons 3-8*,
`GAIA-C9-MN-OPM-FA-064`,
URL `http://www.rssd.esa.int/cs/livelink/open/3311945`

[FA-062], Arenou, F., Fabricius, C., 2014, *Prototype of internal validation tools*,
`GAIA-CO-TN-OPM-FA-062`,
URL `http://www.rssd.esa.int/cs/livelink/open/3311554`

[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`

[TCG-001], Cantat-Gaudin, T., Arenou, F., Vallenari, A., 2015, *TGAS Test description WP947-010-030, Spatial variation and precision of parallaxes in clusters*,
GAIA-C9-TN-OAPD-TCG-001,
URL http://www.rssd.esa.int/cs/livelink/open/3367488

[ELA-011], Antiche, E., Castañeda, J., Fabricius, C., 2014, *GAT Statistics for the old Attitude Star Catalogue - IGSL*,
GAIA-C9-TN-UB-ELA-011,
URL http://www.rssd.esa.int/cs/livelink/open/3262683

[KF-002], Findeisen, K., 2015, *Calibration of Parallax Deconvolution Parameters for WP942-VAL-340-010*,
GAIA-C9-TN-OPM-KF-002,
URL http://www.rssd.esa.int/cs/livelink/open/3368435

[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

[FJL-002], Julbe, F., 2014, *GAT development plan for CU9*,
GAIA-C9-TN-UB-FJL-002,
URL http://www.rssd.esa.int/cs/livelink/open/3252803

[RMC-001], Mor, R., ; Robin, A.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

[AV-016], Vallenari, A., Sordo, R., Cantat-Gaudin, T., Costigan, G., 2015, *WP-947 Cluster Validation, Auxiliary data description for TGAS validation*,
GAIA-C9-SP-OAPD-AV-016,
URL http://www.rssd.esa.int/cs/livelink/open/3368012

[AV-012], Vallenari, A.B., A; Marrese, et al., 2014, *Interface Control Document for GWP947 Observational Auxiliary data*,
GAIA-C9-SP-OAPD-AV-012,
URL http://www.rssd.esa.int/cs/livelink/open/3263826