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

First Year Report

Period covered: from 01/10/2013 to 30/09/2014

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://www.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 first year of the project

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

User community: we have conducted an analysis of the existing archive requirements documents and compared those to the use cases defined by the user's community. The goal was to find out if there are any gaps in the current CU9 requirements with respect to the use cases. The conclusions have been summarized in a document and iterated with the archive developers so they can incorporate the findings in the implementation of their systems. Actions are going to be agreed in the archive Systems Engineering Group.

A solid coordination has been established with the nano Jasmine and Jasmine team in Japan. The incorporation of their astrometric results into the Gaia archive is being defined and also the possibility of installing a mirror of the Gaia archive in a Japanese data centre is being discussed.

Archive system design: GENIUS has made a significant contribution to the design of the archive in several areas. First, in the definition of the technical requirements on the archive DB system, based on an analysis of the user needs. Second, in the definition of the archive's data model, and the improvement of the tool used at DPAC for its implementation and

update (the so called Main Database Dictionary tool). Third, by providing a tool to allow easy deployment of a TAP service on top of their data resources, incorporationg the Distributed Query Processing tools from UEDIN.

Tools for data exploitation: this task has contributed with several tools to the archive. Main and foremost, the adaptation of the popular tool for handling of catalogue and table data, TopCat, for the Gaia archive. TopCat is now 100% compatible with the TAP interface at the main Gaia archive at ESAC and can also directly handle the native data format used in DPAC (gbin files). Also, a very complete visualization package has been developed and is going to be integrated into the archive by means of a visualization server installed at the ESAC premises, directly connected to the archive database. Furthermore, two developments have been started in this frame, to be completed in future phases of the project: a testbed data mining prototype has been implemented and is being evaluated and the framework for the integration of the photometric Science Alerts in the archive has been defined.

Validation: GENIUS is also making a significant contribution in the preparation of the tasks for validation of the Gaia data, a crucial issue for the success of the mission. The activity has focused on the full definition of the validation tests, compiled in a Validation Test Specification document; the implementation of these tests has already been started in GENIUS and in a few cases it has even been tested in the main archive at ESAC.

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 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 requirements 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

Two different web sites have been created with different purposes. The GENIUS project itself is advertised through the `http://www.genius-euproject.eu` portal, which contains only basic general information in English about the project.

Apart from that, to fulfil one of the GENIUS project objectives, which is dissemination on Gaia, a different web site, `http://www.gaiaverse.eu`, has been created with a multilingual approach and focused on diffusion of tools, resources, news...

The reason behind setting up two different portals is to decrease difficulties for users and facilitate access to dissemination resources. Thus, `gaiaverse` is exclusively dedicated to dissemination in several languages while genius-euproject is just focused on paperwork and general information on the project.

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)
- 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.

Although the initial Gaia planning on which the GENIUS schedule was based has been changed by DPAC (see Section 3.2.3.2 for a complete discussion) with the first two releases now foresee for 2016 and 2017. Therefore, the GENIUS contributions to the Gaia archive will first be tested on real data through an internal data release exercise which will take place in mid-2015.

- **Tailor to user needs** A first objective of GENIUS is to ensure that the requirements that drive the design of the Gaia archive and the tools provided for its use are fully in line with the foreseen scientific usage of the Gaia data. To achieve this the user community is being involved in all stages of the development of the project, ensuring that the community's needs are translated into requirements, design features, user interfaces and tools available. Essentially, we intend to avoid the kind of situation where a data archive is an elegant exercise of engineering skills but may not fulfil the needs of its users. Furthermore, we intend to provide special attention to what we have called the Grand Challenges (see WP2 for more details), relevant scientific cases of use of the Gaia data that require an intensive and/or complex access to the archive. (see Section 3.2.2.2).

  Related milestones for the first year: 3, 5

- **Optimum archive system** Deriving from the above goal, a second objective is that the design of the archive itself and its interfaces are tailored and optimised for the needs defined by and with the users. The European Space Agency is assuming responsibility for developing and hosting the Gaia archive at ESAC, where a team is already working on the hardware infrastructure and database design for the purposes of serving all current and future ESA missions. However, contributions from other development groups are required in the areas of server-side infrastructure to support richly functioned interactive and Virtual Observatory compliant user interfaces. We are providing just such a contribution for archive systems development, as described in detail in Section 3.2.2.3).

  Related milestones for the first year: 2, 3, 6

- **Tools for exploitation** The next logical step once the archive system is available is the development of tools allowing its effective exploitation. Our third objective is the definition of such tools, based on user needs, and their implementation in the Gaia archive. It is important to remark here that we do not claim we will make all the possible tools or the basic query interfaces (although we expect to cooperate in this tasks with ESAC) but that our goal is to concentrate on tools that can significantly enhance the scientific exploitation of the catalogue beyond what is currently possible.

This is specially relevant in the case of the above described Grand Challenges, where we intend to specifically work on enabling tools. The developments, described in more detail in Sections WP4 and WP7, 3.2.2.4 and 3.2.2.7, cover three areas: visualization tools, data mining tools, Virtual Observatory tools and outreach tools.

Related milestones for the first year: 3, 6, 7

- **Validation** As defined by the Gaia Archive Preparation Group, a key task for the release of the Gaia catalogue, which should be assumed by the community working on the Gaia archive, is the validation of its contents. Thus, our fourth objective is to contribute to the validation of the catalogue in close cooperation with DPAC. Ensuring a high quality for a one-billion object catalogue containing a wide variety of data (astrometric, photometric, spectrophotometric, spectroscopic, ...) will be a major scientific and statistical challenge. We intend to adapt the exploration tools mentioned above to this task as well as to provide specific tools enabling an effective validation before publication, making sure that the validation methods are based on solid scientific grounds thanks to the involvement of the scientific community around WP2. In particular, this validation requires an intensive cross-matching and interoperation of Gaia data with existing and future astronomical archives as well as with specific ground-based observations (which are already being organised). An additional objective is to ensure that these cross-archive activities are possible. More details are provided in Section 3.2.2.5.

  Related milestones for the first year: 3, 6

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 first six months have been devoted to preparatory activities in anticipation of the actual development stage, including position advertising, coordination and integration with DPAC/CU9, kick-off meeting and hiring of personnel. Specifically for this work package, a part-time project manager (Lola Balaguer) was hired from 1st of January, to help in the management tasks of GENIUS. She has participated in the supervision of this set-up period, the reporting tasks and the general coordination and supervision activities.

First of all we developed a Twiki system for internal information management `http://gaia.am.ub.edu/twikigenius` and we held the kick-off Meeting on 4-5 December 2013 at the University of Barcelona premises, with the 21 participants from the 13 nodes (five of them via videoconference) plus the Project Officer. We set the bases of the management and organization of the network as well as the planning for the first year. Furthermore, in order to fully integrate the GENIUS activities in the overall Gaia/DPAC effort, we organised a joint meeting in Vienna in coordination with the Gaia CU9.

We have developed a reporting system where we report every three months to the project officer from the information collected by each WP on our Twiki. We have had periodic telecons with all the nodes to be able to track all the developments.

Most of the personnel was hired on the first months of this first year, always keeping in mind the principles of the European Charter for Researchers and the Code of Conduct for their recruitment. It has taken place in a globally coordinated way and placing an emphasis on individual excellence and capacity for team working and taking special care to ensure equal opportunity and gender balance. All the positions were advertised in our webpages as well as following the local procedures

of each node.

To be able to provide the best work-life balance in the development of this project we have proposed to allow for more but shorter trips and "virtual international contacts". With that aim we have purchased the required equipment to make it happen: a Webex licence for teleconferences and a conference microphone. We have organized around 15 teleconferences in this year.

As also specified in the proposal, an External Advisory Board has been formed; several candidates were contacted and we finally selected four members: Mark Wilkinson (University of Leicester), William O'Mullane (ESAC), Tadafumi Takata (National Astronomical Observatory of Japan) and Françoise Genova (Centre de Donees de Strasbourg). Their work will begin early 2015 and their feedback will be received in the mid-term meeting.

3.2.2.1.3 Highlights

- Integration of GENIUS tasks and contributions into the DPAC-CU9 structure
- Set-up of Twiki system for internal information management: `https://gaia.am.ub.es/Twiki/bin/view/GENIUS/`
- Set-up of internal repository for file management (SVN repository at UB server)
- Acquisition and management of WebEx system for teleconferences
- Kick-off meeting organization in Barcelona, 4-5 December 2013
- Organization of the joint CU9-GENIUS meeting, Vienna 7-9 July 2014
- Organization of monthly progress review teleconferences `https://gaia.am.ub.es/Twiki/bin/view/GENIUS/MeetingsTeleconfs`

3.2.2.1.4 Deviations and impact on tasks and resources

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 of 0.9 PM of the dedication of L. Balaguer (2.7 PM Jan - Sep) with that in the UB Form C due to the payment structure based in 6-months bills.

	UB
WP1	3.8

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 will concentrate 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 has conducted an analysis of the existing CU9 requirements documents and compared those to the use cases presented in Brown, A., Arenou, F., Hambly, N., et al., 2012 (AB-026), Gaia data access scenarios summary, GAIA-C9-TN-LEI-AB-026 `http://www.rssd.esa.int/cs/livelink/open/3125400`. The goal was to find out if there are any gaps in the current CU9 requirements with respect to the use cases. The results were written up in Costigan, G., 2014 (GCO-001), Gaia Archive Requirement Analysis, GAIA-C9-TN-LEI-GCO-001, `http://www.rssd.esa.int/cs/livelink/open/3282045`. The document has been circulated among the CU9 work package managers so that they can respond to the findings in GCO-001 (incorporating any missing requirements).

Costigan also worked with Walton to further advertise the ongoing process of gathering user requirements through presentations at the Gaia/CU9 meetings she attended.

KU contribution

The KU team participated in various GENIUS and CU9 meetings and out of the discussion the following arose:

- The first Gaia catalogue release and the first NanoJasmine releases are planned for the Summer of 2016. It was agreed to integrate the first NanoJasmine catalogue data into the second Gaia Catalogue release.
- An institution will be identified in Japan which may become a DPAC affiliated data centre and receive the Gaia data when the first catalogue is released.
- Nano Jasmine data could also be used for validation after the first release and cross match.
- Another possible area for collaboration might be outreach; material produced at both sides might be shared and used by both groups.

The NanoJasmine team summarized the requirements on the Gaia archive from the viewpoint of scientific users in Japanese meetings, and also reported them in the joint GENIUS/CU9 plenary meeting. At that meeting it was agreed to intensify the Gaia-NanoJasmine contacts in the areas of data publication and data validation.

UCAM contribution

As part of the GENIUS science requirements update (see `http://great.ast.cam.ac.uk/Greatwiki/GaiaDataAccess`) the community has been asked to provide additional usage scenarios detailing how they may wish to use the GENIUS archive, and outline functionality required. The call for new requirements was announced (in a presentation by Walton) at the GREAT plenary (see `http://great.ast.cam.ac.uk/Greatwiki/GreatMeet-PM7`) at the European Astronomical Society's 'European Week of Astronomy and Space Science' conference, held in Geneva, July 2014. The scenario collection period is currently open. The main activity in terms of analysis of the new scenarios will occur in Year 2 and 3 of GENIUS.

No UCAM staff effort was charged to GENIUS WP2 in the year 1 reporting period.

FFCUL contribution

The team at FFCUL will define the list of requirements and feasible use cases to be covered by visualization.

The list of requirements and feasible use cases to be covered by visualization have been compiled in the CU9 Visualisation Software Requirement Specification (Moitinho et al. 2014, WP980 Visualisation - Software Requirement Specifications GAIA-C9-SP-SIM-AM-002). The main drivers of the requirements come from the visualisation use cases, according to the following sources:

1. the general scientific community interested in exploring the Gaia archive. Community feedback has been 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)
2. 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.
3. 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 Framework Requirements Definition and Gap Assessment (ref: VA-4D_100-05).

The Visualisation SRS is mostly done. The document will be released on Gaia Livelink, and is already in the CU9 SVN repository.

The list in the Visualisation SRS addresses requirements on the visualisation software. These requirements have led to the definition of an architecture (server, plug-ins and clients) which will put requirements on the Gaia archive infrastructure. These requirements on the archive are being identified and are the subject of on-going work and interaction with ESAC for the integration of the visualisation software at ESAC.

Task 2.3 - Confronting complex models with complex catalogues

Task leader: UL

Hypki has been getting acquainted with the Gaia project, CU9, and the work required for Ts 2.3 (and 2.6). He has attended the GaiaChallenge meeting in Heidelberg (together with Costigan) in order to directly talk to the community of astronomers interested in making comparisons of complex models (of the Galaxy, globular clusters, etc) to the Gaia catalogue data.

Task 2.4 - Seamless data retrieval across archives and wavelength domains

Task leader: INAF

The ASDC group have begun planning for the required work for user enabling activities. The first work has been to build robust quick cross matching programs for large ($\sim 10^9$) catalogues which is currently at a very good stage. Tests using the 2MASS catalog have resulted in matching times of 2–3 hours for matching 5×10^8 to 5×10^8 objects. The work to include these large catalogues and smaller catalogues from different wavelengths, with figures of merit and user friendly access, is planned for mid-2015 when we expect the GENIUS contracted person to start.

Task 2.5 - The living archive

Task leader: UL

Brown and Costigan have brain-stormed about the living archive concept. Costigan is in the process of identifying already existing archives that offer the option to incorporate updates in order to see what we can learn from others.

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

Task leader: UL

This activity has not started yet.

3.2.2.2.3 Highlights

The analysis of the requirements by Costigan showed that, with the exception of a few missing items, the DPAC CU9 requirements cover most of what the science community has specified through AB-026. This does not preclude that more advanced requirements — which come up as the Gaia project develops — may have to be included in addition.

3.2.2.2.4 Deviations and impact on tasks and resources

No deviations encountered up to this point.

3.2.2.2.5 Use of resources

The following table lists the person-month per participant in the first 12 months in WP2.

	UL	INAF	FFCUL	UCAM	KU
WP2	7.07	0.59	0.0	0.0	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.

3.2.2.3.1 Summary of progress made

Task 3.1 - Technical coordination

Task leader: UEDIN

At the highest level, WP coordination has been achieved via the following collaboration tools:

- The DPAC Wiki[1];
- The DPAC SVN[2] for both code and documents;
- The DPAC Main Database Dictionary Tool[3] for detailed technical definition of data model interfaces;
- Regular teleconferences in addition to periodic 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 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.

[1] See the WP930 pages linked from the CU9 top–level at `http://wiki.cosmos.esa.int/gaia-dpac/index.php/CU9:930`

[2] See `http://gaia.esac.esa.int/dpacsvn/DPAC/CU9/`

[3] See `http://gaia.esac.esa.int/maindb/mdbtools/`

In the reporting period the WP members have attended 11 meetings (either by teleconference or face–to–face, including meetings organised with other WPs). For full details including dates, agendas and minutes see the CU9 WP930 meetings Wiki page[4].

Task 3.2 - Aspects of archive interface design

Task leader: UEDIN

Interface design has been tackled in three ways:

- Definition of a WP system requirements specification[5]
- Definition of subsystem Interface Control Documents for coordination with the ESAC Science Archives Team[6]
- Enhancement of the DPAC Main Database Dictionary Tool for use not only more generally amongst CU9 in partnership with GENIUS.

The last item is in particular very important. Early on in the WP definition phase it became clear that there were several important enhancements needed to optimise the DPAC database dictionary tool for use generally within CU9 and in particular with GENIUS WP3. Our enhancements include provision of methods to specify human–readable metadata (default values, detailed descriptions) and Virtual Observatory protocol metadata (XML unified content descriptors etc.) through the archive systems to the end–user. The provision of detailed descriptions incorporating LaTeX markup, heavily used within DPAC for mathematical descriptions, translated into XML for inclusion in end–user archive documentation has been an important contribution. Elsewhere, UEDIN has played a key role in the specification of the archive data model in the areas of general design features and a specification of the relational design for crossmatching large catalogue datasets (the so–called 'crossneighbour' approach[7])

Task 3.3 - VO infrastructure

Task leader: INAF

Contributing partners: CNRS, UEDIN, CSIC

INAF contribution have focused on providing the scientific user community with a tool to allow easy deployment of a TAP service on top of their data resources. One goal is to allow the Distributed Query Processing from UEDIN (see Task 3.4) to work on top of the GAIA archive and

[4] http://wiki.cosmos.esa.int/gaia-dpac/index.php/WP930-meetings

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

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

[7] http://wiki.cosmos.esa.int/gaia-dpac/index.php/CU9-WP930-WBS-WP934:Database_Collaboration#The_cross-neighbour_table_approach

local user resources. The development takes its start from the existing VO-Dance application[8] and its TAP implementation companion IA2TAP. Work has progressed here to modify the current web application behaviour from the base one, where the application required a web container (Glassfish) to run, to more 'out–of–the–box' solutions. One features the web container embedded into the application, another provides a full stack service system to be distributed as a standalone Virtual Machine disk image. Both of them will require only service configuration, while the base web application solution required also setup steps for the web container. Widening the DBMS connectivity of the web application (and other solutions consequently) is also underway.

UEDIN contributions in this area have focussed on lobbying the International Virtual Observatory Alliance[9], 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. This includes making the case for adoption of commonly implemented, but as yet not adopted, SQL features for inclusion in ADQL to facilitate greater server–side processing in scale–out situations. UEDIN has also continued to develop 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.

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

Task 3.4 - Data Centre Collaboration

Task leader: UEDIN

Substantive contributions have been made in this area. UEDIN has designed, documented and prototyped a complete subsystem that implements Distributed Query Processing. This has long been an ambition in the VO world as it is seen as fundamental to the ubiquitous usage scenario of cross–querying multi–terabyte, multi–wavelength survey datasets of which the Gaia catalogue is a prime example. The interface control document[10] (i.e. the formal technical specification and coordination agreement with ESAC–SAT) has been delivered to ESAC. Proof–of–concepts has included deploying the prototype system against GUMS data provided by the GACS TAP interface. The prototype code presently resides in UEDIN's code repository[11], and is fundamentally designed around the relational 'cross–neighbour' table approach (see above).

Task 3.5 - Cloud–based research and data mining environments

Task leader: UEDIN

This task is to be activated later in the project.

[8] http://ia2.oats.inaf.it/vo-services/vo-dance
[9] http://www.ivoa.net
[10] https://gaia.am.ub.es/Twiki/pub/GENIUS/DeliverablesGenius/GAIA-C9-TN-IFA-DM-001.pdf
[11] http://redmine.roe.ac.uk/projects/wva/repository

3.2.2.3.3 Highlights

In Figure 3.1 we illustrate how key components integrate via WP3 developments. User access to the Gaia archive (or indeed any publicly available observational data set) can be via our Distributed Query Processing layer as part of a usage scenario involving several data sets, be they large–scale mission data published by a Data Centre or small–scale as a result of an individual's own measurements. These developments hence have wider significance in enabling the cross–exploitation of the contents of the burgeoning Virtual Observatory, a set of space science data holdings distributed over the wide–area network.

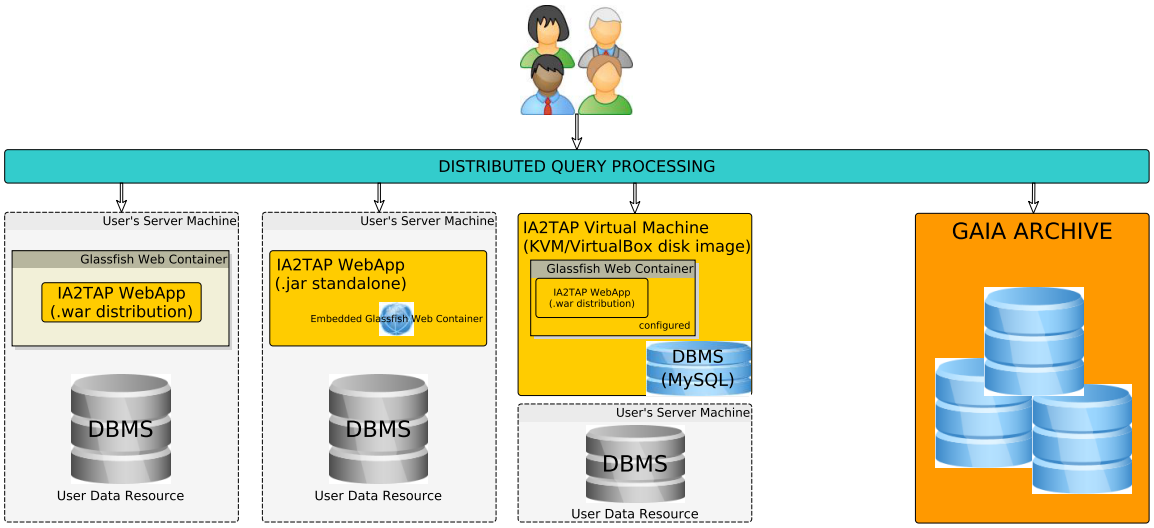


FIGURE 3.1: Interoperability between client–side tools (in this case, INAF's VODance), the Gaia archive, and the Distributed Query Processing layer illustrating one way in which these components interact and how, in a more general scenario, these infrastructural developments will lead to cross–exploitation of observational resources distributed over the wide–area network.

3.2.2.3.4 Deviations and impact on tasks and resources

Task 3.2: the original UEDIN work description included 'Web 2.0' contributions (i.e. richly functioned, browser–embedded end–user interface features) by UEDIN for the Gaia Archive Core Systems interface. In fact, ESAC–SAT implemented this kind of functionality as part of it's wider archive systems development programme and without the need for external contributions. Hence we concentrated more on the integration of the DPAC Main Database Dictionary Tool (see above) into CU9 developments. Work is still planned in this sub–work package in the area of browser–embedded data interaction and visualisation, in collaboration with the UBR developer responsible for the high–performance on–the–fly plotting functionality embedded within the TOPCAT data exploration tool[12]. There has been no impact on resources resulting from this small deviation from the work plan originally specified.

Task 3.3: UEDIN originally envisaged contributing to server–side user database facilities (so–called 'VOSpace' or 'MyDB' functionality). Again, much of this has now been implemented within the GACS without the need for external contributions. Hence UEDIN has concentrated

[12] `http://www.star.bris.ac.uk/~mbt/topcat/`

on the ADQL side of things (see above). Again, this has had no impact on resources, rather it is simply a small change in emphasis of the existing effort in this area.

3.2.2.3.5 Use of resources

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

	UEDIN	**CNRS**	**INAF**	**CSIC**
WP3	7.9	0	5.51	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, complemented with global updates

during the general meetings and teleconferences listed in section 3.2.3. 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

This Work Package addresses the development of visualization tools and solutions, adapted to the large size and complexity of the Gaia archive. Its main objectives are:

- Define the architecture to support the visualization requirements.
- Identify the existing open-source visualization tools to be used or extended to support the graphical view of the Gaia archive.
- Define the proper data models for the visualization of the requirements. In particular:
  - Define in collaboration with T4.3 the requirements for data mining visualization.
  - Define in collaboration with T4.4 the infrastructure technology compatibility and extensions to use VO standards and services.
- Implement, test and monitor the visualisation and interaction tools (widgets and algorithms).

We start by spanning the progress on the main objectives. Additional activities are reported in the end of the section.

- Definition of the architecture to support the visualization requirements. The client-server architecture to support the visualization requirements was defined during the first period of the GENIUS project. The server-side component of the Visualization Infrastructure is responsible for interfacing with the Gaia archive, for pre-processing the archive contents, for streaming the (pre-processed) archive to connected Visualization Clients. The aim of the Server is to hide the volume and complexity of the Gaia Archive from the Visualization clients, in order to allow the visualization of billions of individual objects using off-the-shelf hardware. A top level diagram of the server architecture may be seen in Figure 4.2.

  This 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.). The plugins are automatically discovered, and are implemented via a Service Provider Interface (SPI) enabling dynamic loading. Moreover, the defined server-side plugins currently support:

  - Simple Application Messaging Protocol (SAMP) masquerading for clients that do not support this VO standard;

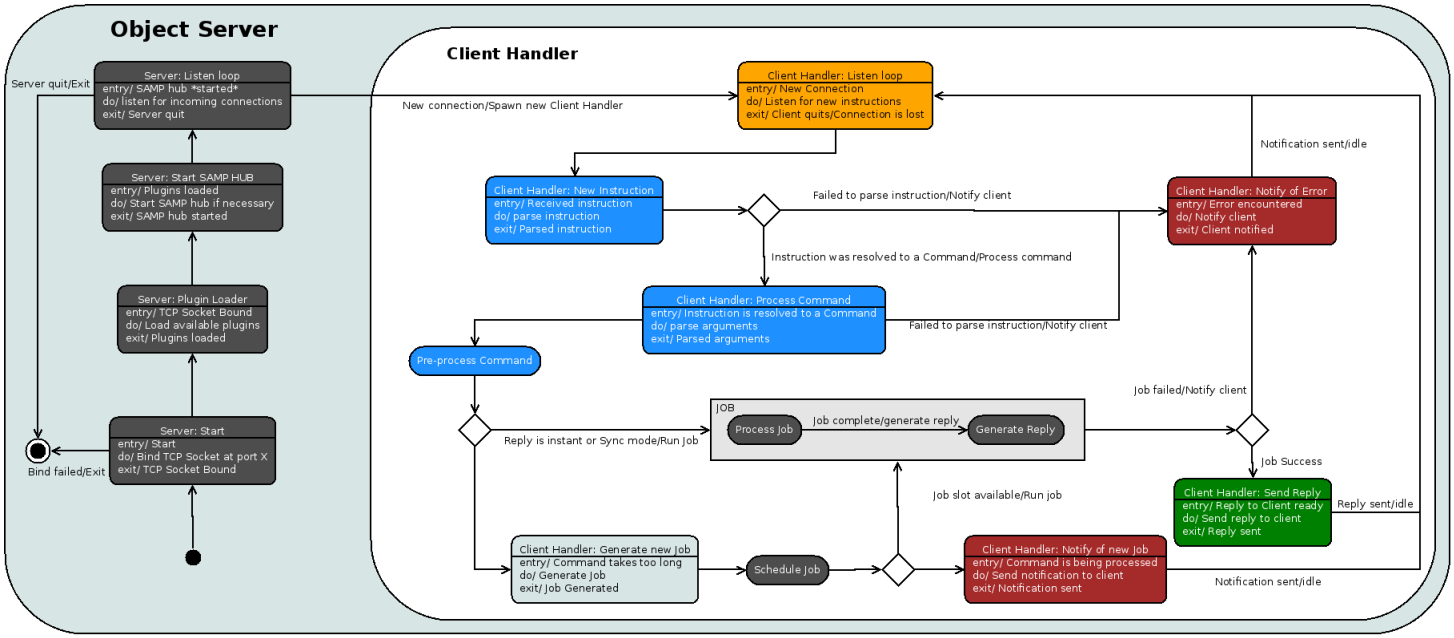


FIGURE 4.2: Top level diagram of the server architecture

- Table Access Protocol (TAP) Service;
- Universal Work Service (UWS) compliant Server (for uses not covered with TAP);
- SAMP Hub;
- 2D and 3D Visualizations;
- Visualisation Levels of Detail (LoD);
- Indexing.

A preliminary architecture for the client-side application is being created. The architecture of both components (server and client) can be found at the Gaia Wikipage, in the CU9 area: `http://wiki.cosmos.esa.int/gaia-dpac/index.php/Visualisation_infrastructure`

A formal document, to be released on Livelink, will be written during the next period of the GENIUS project.

- Identification of existing open-source visualization tools to be used or extended for supporting the graphical view of the Gaia archive. This is partially done. Still needs to be completed and consolidated in a document
- The general data model for visualization has been defined and included in the CU9 data model (DPC/CU9/Visualization). Still to be done:
  - Define in collaboration with T4.3 the requirements for data mining visualization.
  - Define in collaboration with T4.4 the infrastructure technology compatibility and extensions to use VO standards and services.

We remark that the implemented VO standards are supported as defined in the IVOA documents.

- Implement, test and monitor the visualisation and interaction tools (widgets and algorithms). This is the core activity of T4.2. As foreseen in the proposal, it is underway and will occupy the whole duration of the project. An alpha version of the server-side component, together with API wrappers for Java, C# and Python, is ready. An API wrapper for R is under way. Deployment of the alpha version at the Gaia archive in ESAC is expected before the end of 2014.

  The next figures show two examples of functionalities delivered by the current server-side component.

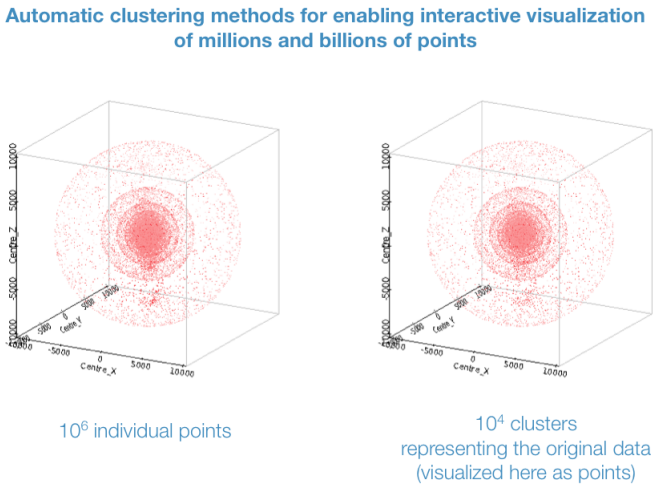


FIGURE 4.3: Two levels of detail for the spatial view of tycho-2 data set. Highlights a case in which overall visual features are be preserved with a reduction of two orders of magnitude in the number of points.

Besides the research and development within the first year of GENIUS, the following activities were also conducted:

- Organisation of a Visualisation infrastructure workshop in Lisbon, January 2014: `http://wiki.cosmos.esa.int/gaia-dpac/index.php/Visualisation_infrastru workshop`
- Organisation of a Gaia Visualisation (GaiaViz) workshop, Vienna, `https://gaiaviz.univie.ac.at/`
- Presentations in the GaiaViz workshop.
  - Alberto Krone-Martins: It was "overview of existing tools", but... What is not already out there?
  - Andre Moitinho: Gaia CU9 Visualisation Plans

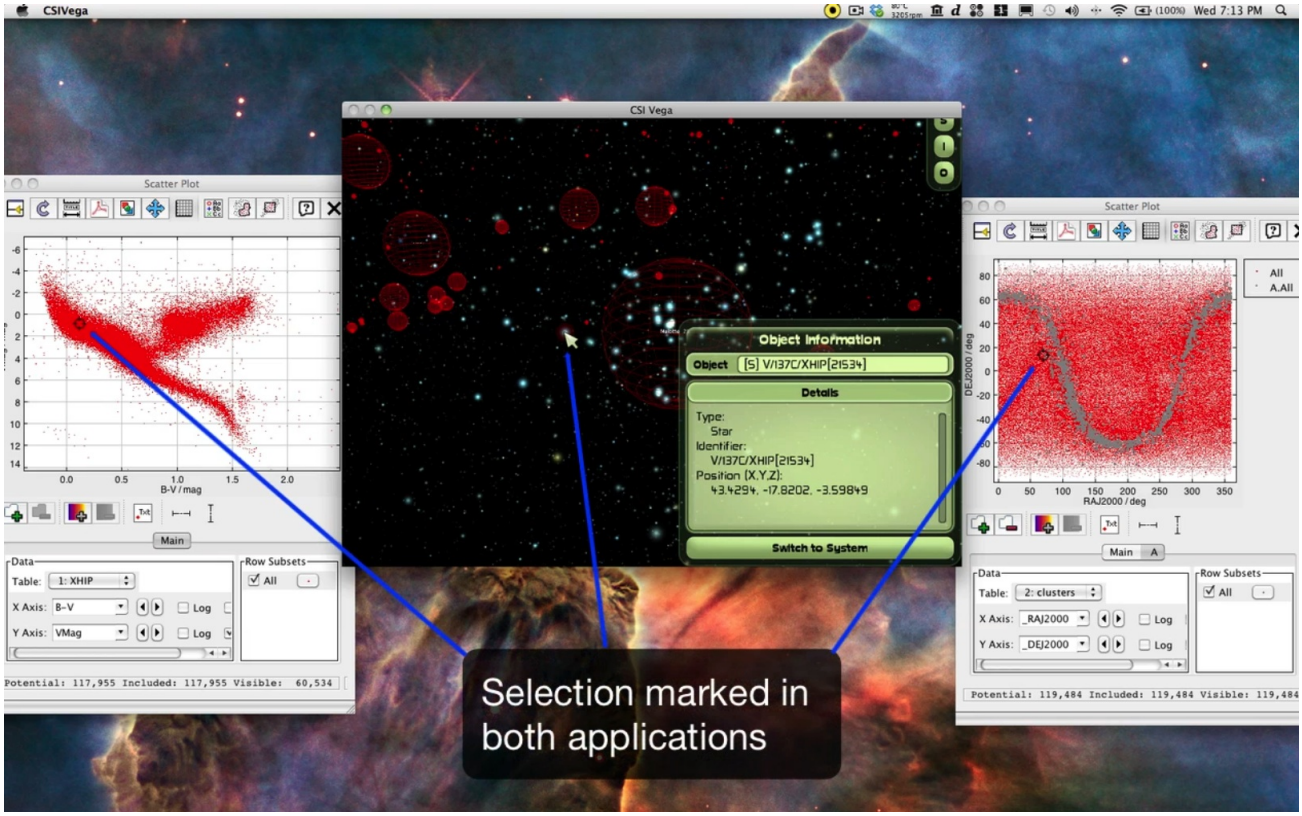


FIGURE 4.4: Tracking points in 7 different dimensions: linked views between our testbench 3D client and two Topcat 2D panels. Linked views are enabled by the SAMP support provided by the visualisation server. Data from the hipparcos catalogue is provided to both applications by the visualisation server.

- Presentation in the CU9 plenary meeting, Vienna: Andre Moitinho, WP980 - Visualisation Status
- Regular CU9 telecons

Task 4.3 - Data mining

Task leader: UB

Contributing partner: CSIC

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.

People involved in this work Package:

- Francesc Julbe, Fundació Bosh i Gimpera, Genius FTE and work package coordinator.
- Luis Sarro, UNED (Universidad Nacional de Educación a Distancia). Data Mining Work package deputy and astronomical data analysis expert (CSIC partner)
- Benjamin Montesinos, CSIC.
- Daniel Tapiador, external engineer providing support tasks and big data expertise.

Preliminary studies

Currently there is a set of brand new Big Data technologies that are evolving quickly, some of them were adopted by the community and the market only a few years ago are already being discarded in benefit of other platforms. Our first task in this work package was to analyze, understand and test these technologies in order to decide which would be our technological choice. We initially explored brand new technologies that can allow us to perform the complex operations that an astronomical archive such as the Gaia archive requires. In the context of "Data Science" (study of the generalizable extraction of knowledge from data), Hadoop is the current leader on distributed data processing.

Hadoop http://hadoop.apache.org/ is an open-source software framework for distributed storage and distributed processing of Big Data on clusters of commodity hardware. We started our research analyzing the HADOOP based technologies. Hadoop provides a set of Big Data services (MapReduce, IMPALA, HIVE, HDFS, YARN -Hadoop 2- mainly) to be used by other frameworks such as:

- **Mahout** (https://mahout.apache.org/): a project of the Apache Software Foundation to produce free implementations of distributed or otherwise scalable machine learning algorithms. Based on Hadoop and MapReduce, it provides a set of out-of-the-box features to perform mathematical analysis, clustering, recommendation and some other features
- **Spark** (https://spark.apache.org/): An open-source data analytics cluster computing framework. It provides great performance loading datasets in memory and it is fully integrated with Hadoop.

Test bed

CSUC (Consorci de Serveis Universitaris de Catalunya),former CESCA(Centre de Supercomputació de Catalunya), our partner in the GENIUS project, has deployed a test bed for the study, design and test of our work in this package. Initially this test bed was a 4 virtual nodes cluster, deployed with OpenNebula (http://opennebula.org/). This small Cluster allowed us to familiarize with the way to use these tools, but is way too small to perform any significant use case getting any clear conclusion. We invested about 3 months in learning about these technologies with the first cluster. An upgrade to a bigger cluster was necessary in order to implement more complex use cases. CSUC provided a bigger cluster, with 16 nodes, 16GB of RAM memory per node and 4TB storage capacity for the entire cluster. Although it is not enough for production-like use cases, this cluster is powerful enough to perform the initial use cases we want to run.

In the test bed we deployed Clouders 5, a Hadoop distribution with all the services needed, **Spark 1.1.0** (although it is quickly evolving so version upgrades are being deployed regularly) and **Mahout 0.8**.

Use cases

Dimensionality reduction using Principal Component Analysis -PCA-: Using the Gaia simulator (GOG -Gaia Object Simulator-), we generated 212.193 spectra for simulated sources, characterized by 480 wavelengths points (arrays of 480 values). So the use case consists on obtaining the 480 PCA for these arrays. The size of the package was about 2GB and 8192 files with these spectra. These files were uploaded to the HDFS system of the cluster which are automatically distributed among the 16 nodes. To perform the PCA analysis we have used the two big data technologies presented in 4.3.3: Mahout and Spark.

Spark setup for such an exercise was far much easier both in usability and performance. Spark processing time was about **170 seconds**, while Mahout processing took about **9240 seconds**, about 54 times more. We extended the test to a bigger spectra sample generating up to 2088535 spectra with 480 flux values, which is about 10 times the size of the previous test. With Spark, we generated the PCA components in **293 sec**. The hardware provider partner estimated in a factor 30 the gain we could get from the current cluster to the more advanced one (per node), so with 32 nodes instead of the current 16, we can get a x60 gain in rough numbers, so a full-sky PCA over the BP end-of-mission spectra (2e9 spectra) could take about 4800 sec. with Spark (no memory considerations are taken, still to be evaluated).

This exercise was performed only with a small subset of BP spectra and a using a not too powerful cluster (16 'old' nodes). A linear projection of these results show that a much more powerful cluster is needed to perform a PCA analysis over a full sky (1.2E9 sources). Spark requires a lot of memory, but offers an excellent performance, especially compared to the more classical Mahout's Map/Reduce approach. And according to some benchmarks, even when lacking some memory, its performance is better than the Mahout one.

Design

After our analysis, we concluded that our first technology choice was Spark as a basic framework for Big Data activities inside Gaia given its usability and great performance.

In fact, this decision was reinforced by the fact that the other framework tested (Mahout) is abandoning Map/Reduce to switch to run over Spark in a near future, giving us the confidence that the platform was becoming the principal player in the Big Data scenario. This was fulfilling another of our criterion to choose a technology which is confidence in the long term durability on the market. In parallel to the technological study, we have designed a first version of the high level framework.

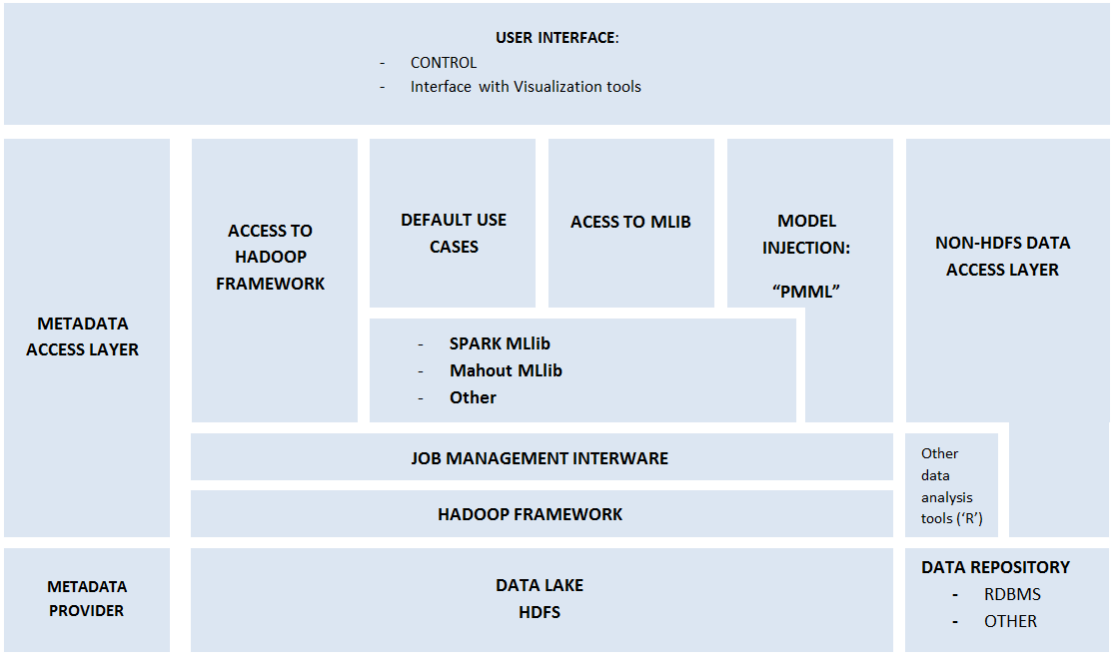


FIGURE 4.5: Data Mining framework high level design

The proposed design consists on a layered structure:

1. Graphical user interface that will allow the users to perform advances queries. It will be based on web technologies.
2. Spark will provide the MlLib implementations that are the building blocks to buld complex use cases ("default use cases", which will implement the most common astronomical use cases and "Access to MlLib", which will allow the users to perform MlLib operations to the platform).

3. PMML model injection: PMML is an XML standard for the interchange of predictive analytic models developed by the Data Mining Group.

4. Advanced users will be able to submit tasks directly to the Spark and/or Hadoop ("Access to Hadoop Framework" in the diagram).

5. The Data Mining framework should have access to the metadata from the Archive that contains descriptive information of the archive, needed to option and data selection.

6. We don't discard to incorporate more traditional technologies such as R language and relational data repositories for more simple data access tasks.

7. Distributed data will be stored in a HDFS (Hadoop Distributed File System) data repository (also known as "Data Lake").

Prototype

We have started the implementation of a basic prototype of the Gaia Big Data framework as a proof of concept which will allow users to perform supervised classification using **Lasso-Ridge** regression, already available in Spark MlLib implementation. We propose a graphical user interface accessible via web that can submit jobs to our Spark cluster, displaying the final results to the client.

We successfully implemented a skeleton of this prototype, with a Web application implemented using Google Web Toolkit (http://www.gwtproject.org/) that submits tasks to the cluster and can handle the results, being displayed on the screen. Tasks are executed synchronously (that means, the client submits and waits for the results). We expect to add an intermediate player that can handle the job execution. After the completion of this prototype, expected before the end of 2014, we will expand its capabilities, both in the client side, offering more and more options to the user with a more complete interface and on the server side, with more diverse operations available to build more advanced use cases.

Organization and monitoring resources

Francesc Julbe, Luis Sarro and Daniel Tapiador have had several telecons and meetings at UNED to define lower level requirements, follow-up the progress of the work package, define new tasks and establishing a work plan for the short/mid term. We have had regular meeting and telecons with CSUC, our infrastructure provider, to follow up the capabilities of the cluster and its possible improvements. Our activity has been reported in DPAC's twiki http://wiki.cosmos.esa.int/gaia-dpac/index.php/CU9:970:973. For development notes, we have used the Evernote service

(https://evernote.com/intl/es/)

Relevant meetings:

- We presented our work to the GENIUS community in Vienna, July 2014.

- Meeting with SAT (Science Archive Team) at ESA - ESAC, European Space Astronomy Centre in October 2014.
- Work package kick-off meeting at UNED with Luis Sarro, Daniel Tapiador, Benjamin Montesinos and Joaquin Ordieres, February 2014.
- Follow-up Meeting at UNED with Luis Sarro, Daniel Tapiador, Benjamin Montesinos. September 2014.
- Meeting at CSUC facilities to define the initial infrastructure, December 2013.
- Meeting at CSUC facilities to set new infrastructure requirements.

Task 4.4 - VO tools and services

Task leader: CSIC

Contributing partners: UBR

CSIC contribution

The main CSIC contribution to GENIUS has recently started, in September 2014. The reason for this late start is mostly by planning, but, internal administrative problems at CSIC have also retarded the availability of GENIUS funding to the team involved in GENIUS and have delayed the hiring of personnel until now (work has been carried out by CSIC permanent staff).

Virtual Observatory tools take advantage of VO standardization and open new lines of research by facilitating the discovery, access and intercomparison of astronomical datasets. This task has started (out of the four VO tools included in GENIUS) by focusing on the improvement of the most widely used VO tool developed by the Spanish Virtual Observatory: VOSA.

VOSA (VO Sed Analyzer, http://svo2.cab.inta-csic.es/theory/vosa/) is a web-based tool designed to combine private photometric measurements with data available in VO services distributed worldwide to build the observational spectral energy distributions (SEDs) of hundreds of objects. VOSA also accesses various collections of models to simulate the theoretical SEDs, allows the user to decide the range of physical parameters to explore, performs the SED comparison, provides the best fitting models to the user following two different approaches (chi square and Bayesian fitting), and, for stellar sources, compares these parameters with isochrones and evolutionary tracks to estimate masses and ages.

VOSA was firstly released in 2008 and its functionalities are described in Bayo et al. (2008, A&A 492, 277). At present there are more than 300 users in VOSA who have published more than 50 refereed papers.

In the framework of the GENIUS project we have identified the following improvements to VOSA for its use with the Gaia archive:

- More collections of observational and theoretical data (including mathematical functions).
- Implementation of extinction maps, different extinction laws and Rv values.
- Proper handling of upper limits in the fitting process.
- Include proper motion treatment
- Provide a batch mode option for very large queries. Nowadays, depending the connection and load of the server, it takes several hours to do the full workflow for a few hundreds objects. These numbers do not fit with what is expected to come from the GENIUS community. We are currently working in the implementation of asynchronous and parallelization capabilities in VOSA so that big queries that cannot be processed in a reasonable web-response time, are treated as batch jobs.

UBR contribution

This is the whole of the effort allocated for UBR within GENIUS. Taylor/UBR will remain 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.

Activity within GENIUS has been as follows:

- TOPCAT STILTS access to Gaia catalogue from ESAC-hosted GACS TAP service

  TOPCAT and STILTS contain Table Access Protocol (TAP) clients, which make them able to execute SQL-like queries, including crossmatches against locally-held tables, against any data exposed via a TAP service. The GACS server at ESAC implements the TAP protocol, and hence TOPCAT/STILTS were in principle already able to provide direct access to Gaia catalogue data from the ESAC-hosted TAP service.

  UBR GENIUS activity in this area has consisted in ensuring that the client and service do in fact interoperate as required. The service initially had a number of compliance issues; I have run extensive tests, and posted 16 Mantis tickets [IDs available on request] detailing points of non-compliance of the service with the relevant VO standards. I have also provided ESAC with the taplint TAP validation tool (part of the STILTS suite) along with support in using it, so that they can use it [Mantis issue 0028011] as part of their own testing processes. This work has included some upgrades and adjustments to taplint, as well as discussions with ESAC staff and within the IVOA about some fine points of TAP compliance.

  ESAC have now resolved nearly all of the relevant Mantis issues, with the result that TOPCAT and STILTS can successfully perform direct queries, including ones involving upload of local data, on the ESAC Gaia catalogue service. Other TAP clients, of which several are implemented and in use within the astronomical community, are expected to benefit in the same way from these improvements to the TAP service implementation.

- TOPCAT STILTS access to Gaia catalogue from CDS X-Match service

  The CDS X-Match service (`http://cdsxmatch.u-strasbg.fr/xmatch`) provides a very high-performance public service for positional crossmatching of uploaded catalogues against catalogue data held in the VizieR database.

  Results of the GUMS simulation are already available from this service (`VI/137/gum_mw` etc), and the Gaia catalogue will similarly be imported when it becomes available.

  UBR GENIUS activity in this area has been to implement an integral client to the CDS X-Match service within TOPCAT (`http://www.starlink.ac.uk/topcat/sun253/CdsUploadMatchWindow.html`, and STILTS `http://www.starlink.ac.uk/stilts/sun256/cdsskymatch.html`). The implementation includes a number of client-side optimisations, including chunking of uploaded tables to enable arbitrarily large crossmatches to be performed. It is now very straightforward for users to take local catalogues, possibly forming part of a custom processing workflow or interactive TOPCAT analysis session, and perform positional crossmatches against Gaia data (or any other large or small catalogues held at CDS), receiving the result back into the same batch or interactive context in the most appropriate form. Typical matching speed is of the order of a million positions a minute. This capability is less flexible than the TAP-based facility described above, but it is easier to use, faster to run, and capable of dealing with larger local tables.

- Software support

  Informal support by email have been also supplied to other GENIUS members concerning use of the STIL table I/O library for use in implementing TAP services, and the STILTS command-line suite for crossmatching and generating visualisations.

3.2.2.4.3 Highlights

- Release of a version of TopCat adapted to the Gaia archive. This version was tested against the current archive prototype and allowed to detect and solve several deficiencies of the archive's TAP interface, that were solved by the ESAC team. Furthermore, access to Gaia server is now integrated into TopCat, that allows even to directly read the internal data format of DPAC (gbin files); although not intended for scientific exploitation this will significantly help the work of the DPAC.

- A prototype for the data mining of Gaia has been installed and operated at CSUC. On a first stage, four virtual machines were deployed on a private OpenNebula cloud environment. That configuration allowed to get expertise in the installation and management of CDH, the popular Apache Hadoop distribution. Some programming and performing analysis were executed, which revealed the need to go further and deploy the system using physical machines.

  Then, a cluster consisting of 16 nodes were set up to provide the service. Processors in this HP CP4000 cluster came from the AMD Opteron 275 family, with a dual core

architecture, working at 2.2 GHz. Although it remains a pilot, this configuration permitted to test CDH in a more realistic environment, with 16 GB of RAM memory and 300 GB of disk per node.

The work with this prototype has allowed to consolidate the concepts for the Gaia data mining, that are now being discussed with the Science Archives Team at ESAC for its future implementation in the main archive.

- The developments on visualization of Gaia data are now converging into an actual implementation on the archive. A visualization server is going to be installed at ESAC before end of 2014, linked to the current archive prototype, in order to allow for testing of the visualization software and concepts. This work is being carried on in collaboration with the Science Archives Team at ESAC.

3.2.2.4.4 Deviations and impact on tasks and resources

A first deviation occurred due to internal administrative problems in one of the partners, CSIC. Although the funding for this institute was available in due time, it did not internally reach the GENIUS team until very recently (August 2014). This prevented the planned hiring and therefore the full start of the development of the VO applications planned. The consequence is that the development was partially delayed and will initially concentrate in the VOSA tool (see previous section) in 2015, for which the requirement analysis has already been carried out. This tool will still be available for the first archive release, although other tools from CSIC will only be available in the second archive release (to be confirmed in early 2015).

A second deviation has occured in the use of the GENIUS budget devoted to database licenses and specific formation for the database system. This budget was allocated at the proposal stage based on the current information at the time. The expectation was that the main archive system at ESAC would use the Intersystems Cache DB or alternatively the Oracle DB system (both in use at the time there). Both these systems require commercial licenses for its usage, and even for academic or scientific use they are quite expensive, and the proposed budget covered the expected costs. Additionally, an extra funding was allocated for formation on the use of these systems, expecting that the GENIUS participants involved in the DB activities would need some courses to use the DB system. However, the final implementation of the main archive at ESAC uses a different DB system, PostGres. This is an open-source system for which free licenses are available, therefore eliminating the licensing cost initially expected. Furthermore, the CSUC partner already has a good experience in this DB system and can provide all needed expertise in it without the need for expensive formation from Intersystems cache or Oracle.

Due to this, the budget devoted to these two areas has not been spent as planned and will likely not require spending in this areas in the next phases (there are currently no plans at ESAC to reconsider the choice of PostGres). We therefore propose a change in budget allocation that will at the same time solve a third deviation, related to the data mining activities. During the first year of development of the data mining prototype described in previous sections we have found out that the storage and computation requirements for this task are higher than expected. The initial planning covered a small test bed at CSUC for limited proof of concept, which has been operating at CSUC in this first year. However, the actual experiments have quickly reached the limits of the testbed, both in terms of disk storage and scalability, and full proof of concept will require in several use

cases a more poweful testbed. Thus, we propose to reallocate the budget initially devoted to DB licensing and formation to the upgrading of the data mining testbed; this will allow to overcome the bottlenecks that the data mining activities would face in the next year if the current testbed is not upgraded. More details about this proposed budget reallocation are given in section 3.2.3.2.

3.2.2.4.5 Use of resources

The following table lists the person-month per participant in the first 12 months in WP4

	UB	**CSIC**	**FFCUL**	**UBR**
WP4	10	0.25	2.9	3

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

Benefiting from an organisational structure parallel to the DPAC CU9 one, this work package has put in place a technical coordination which first relies on the coordination of the CU9 validation sub-work packages themselves. Indeed, the large size of the CU9 validation group preventing a coordination of the whole group, the CU9 WP94x sub-workpackages managers organise instead their group and then report their achievements.

For this purpose, regular teleconferences with sub-WP managers are being organised, one every about 1.5 months, in addition to two plenary meetings planned per year, the minutes of which can be found in the the CU9 WP940 Wiki page[13]. The telecon meetings with validation sub-WP managers were organised on:

- Telecon meeting 1 (2013-09-02)
- Telecon meeting 2 (2013-10-24)
- Telecon meeting 3 (2014-02-05)
- Telecon meeting 4 (2014-03-25)
- Telecon meeting 5 (2014-05-27)
- Validation splinter during CU9 plenary meeting in Vienna (2014-07-09)

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

- Telecon meeting 6 (2014-09-18)
- Telecon meeting 7 (2014-11-05)

Beside, progress meetings also occur within the sub work packages, sometimes very regularly (every two weeks) within Task 5.2 (CU9 WP942).

One of the main activities of the technical coordination has been to put in place a software environment which will allow to run as automatically as possible the validation tests. The rationale for this is obviously that many tests will have to be implemented, that working on a billion star Catalogue can't (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.

This being said, one key issue is ensuring the integration of the validation software at ESAC, which is needed as the software won't run on the Gaia data in distributed places (non-disclosure restrictions on Gaia data). Beside email exchanges, several meetings have been organised for this purpose:

- Validation Team/ESAC interface telecon meeting #1 (2014-01-27)
- Telecon on WP940 integration with ESAC (2014-09-10)
- Face to face meeting WP940 integration at ESAC (2014-10-16/17)

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;
- The full software development environment

The various tools present in the DPAC software environment have thus also been activated for the validation software area, trying to capitalise on the experience developed within DPAC. This includes:

- Hudson [16], which allows to check, after each software modification, the status of the JUnit tests of the validation. The Hudson view in Figure 5.6 shows the validation organisation, with one project per sub-work package, and a common VOTAP project (see below).

[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/

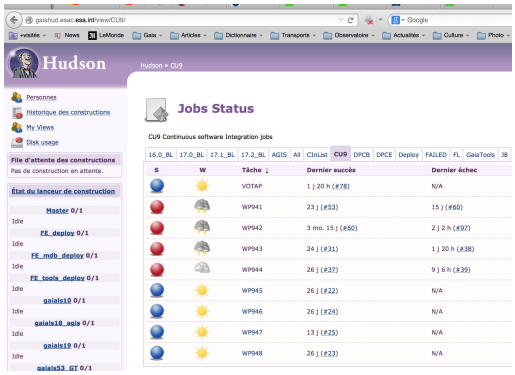


FIGURE 5.6: Continuous software integration Hudson, view of the validation software area.

- Mantis (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 has focused 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 implementation of the tests (described in the following Tasks) is shown Figure 5.7.

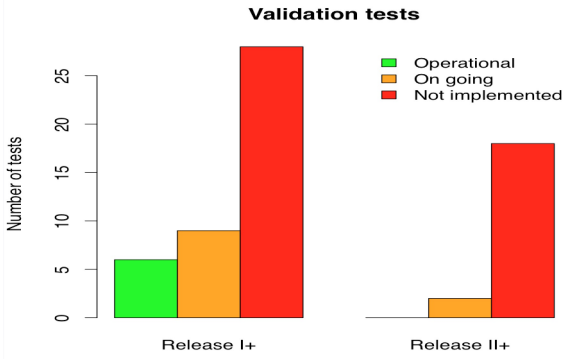


FIGURE 5.7: Software implementation of validation tests, current situation. Each test in a given release is also applied to the next ones, so that the tests indicated for release II are incremental.

One main point favouring the homogeneity of the code development has been the definition of common Java procedures thanks to the development of first common tools for the interface

- with Table Access Protocol (TAP) services
- with VO output

This VOTAP project[18] allows a consistent interface with the database for all the validation users. Beside, a local platform with DB and TAP for tests has been set up in Meudon (CNRS).

[17] `http://gaia.esac.esa.int/dpacsvn/DPAC/CU9/docs/WP940_Validation/ECSS/VTS/`
[18] see Software User Manual GAIA-C9-SUM-OPM-IS-001

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.

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

Task leader: CNRS

This Task has been the one where the main activity has occured. This work package intends to define validation scenarios, and 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 simplest basic tests have thus been developed and fulfill the Deliverable planned after one year and devoted to the development of a prototype of internal checking tools in a close collaboration of CNRS with the University of Barcelona (UB).

Tests on the completeness of the Catalogue are being implemented. Other tests are being developed in interface with GAT, a UB software for the statistical analysis of Gaia Catalogues.

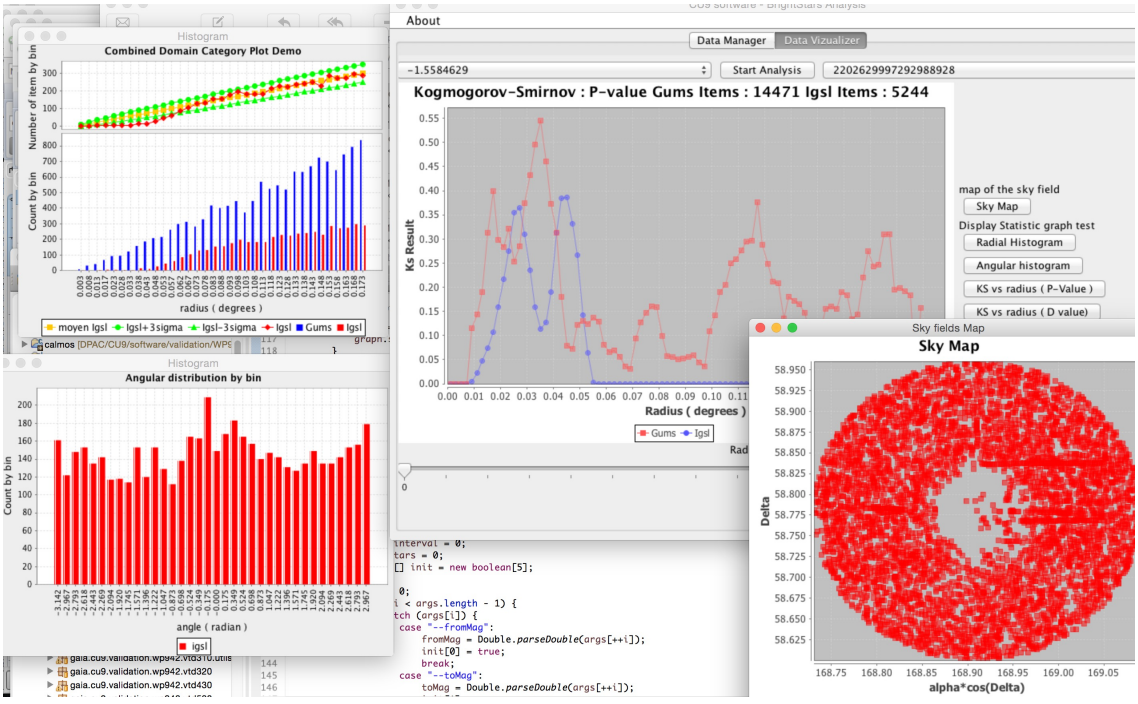


FIGURE 5.8: Software developed for the analysis of surrounding around bright stars to look for artifacts, showing that for this IGSL star, there is both a lack of detection at small radius and at some angle.

Another specific work undertaken has been the analysis of the surroundings of very bright stars, either missing objects or detection of artifacts (false detections). Statistical and visualisation tools have been developed for the purpose of this analysis, cf. Figure 5.8.

Of course, this development has been preceded by the definition of the tests themselves, a large part of the involvement having been devoted to writing the definition for the Validation Test Specification document.

Task 5.3 - Simulation versus reality: from models to observables

Task leader: CNRS

Validation of Gaia data can also be performed by comparing it with predictions of the Besançon model. In turn, predictions of Besançon model for various observed quantities e.g. distribution of stars in magnitude bins, colours, proper motion, etc. must first be compared with previous observations to confirm the consistency of model with present observations.

For this purpose the IGSL catalogue is being used. It is created by combining data from several big astronomical catalogues such as SDSS, 2MASS, GSC2.3, Hipparcos (for bright stars), etc. and compared with GUMS, the catalogue of mock stars generated according to Besançon model of the Milky Way. Despite the presence of artifacts due to calibration errors in IGSL, and possible over-estimation in galactic plane and under-estimation of number of stars in high galactic latitude by GUMS, they are grossly in agreement with each other and can be used to verify for Gaia data, cf. Figure 5.9.

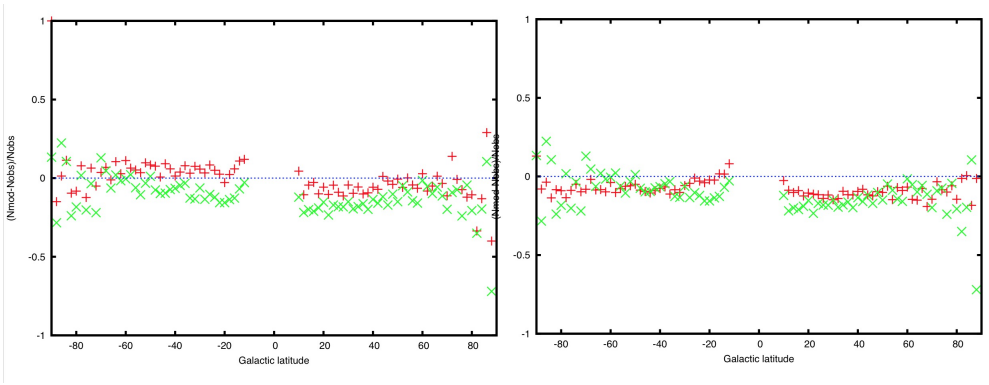


FIGURE 5.9: Comparison with Tycho Catalogue of two Galaxy models for stars with $6 < G < 9$ and $9 < G < 11$: relative errors of the number of stars vs galactic latitude.

Furthermore to discriminate between artifacts and real differences of two catalogues we are performing a decomposition to harmonic oscillator. This method allows to discriminate large and small scale structures in the data. Currently the results and utility of this method is being analysed.

The next step will be to develop software to compare mean proper motions, standard deviations, skewness and kurtosis between these two catalogues to establish the reliability of the model and the level of acceptation of Gaia data by comparison with the model.

Task 5.4 - Confronting Gaia to external archives

Task leader: CNRS

Contributing partner: CSIC, KU

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 main 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. Obviously, for the moment, this concerns very small catalogue cross-matches and this is currently tested against IGSL. What has been implemented are the following tests:

- proper motions versus literature values
- small catalogues of distances (HST Parallaxes for Classical Novae)
- QSO

Task 5.5 - Data demining: outlier analysis

Task leader: CNRS

Contributing partner: FFCUL

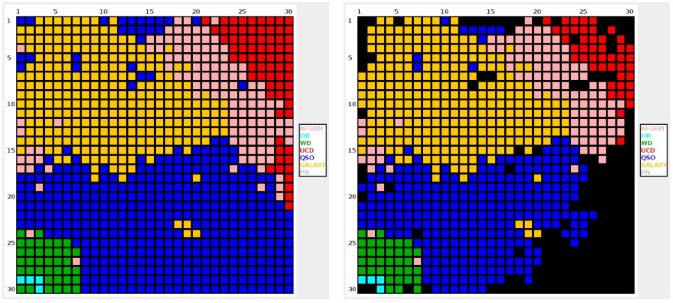
Outliers being by definition objects which deviate from an assumed model, it would be surprising that a mission such as Gaia planned for deciphering the complex structure of the Galaxy would exhibit no outliers departing from our current knowledge.

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

One is the outlier analysis using the Self-Organizing Maps (SOM) clustering tool being done at the Universidad A Coruña. An example of the SOM analysis of SDSS spectroscopic outliers[19] is shown Figure 5.10. The software has initially be developed in Java and integrated in CNES for the CU8 software chain and tested with BP/RP libraries. It is now prepared for use in the validation area, i.e. with "Big Data", in a distributed CPU scheme with Hadoop. It must still be tested with astrometric data.

A second approach is based on a new technique developed in the University of Groningen for infering Galactic parameters from stellar streams and other debris. This technique also takes advantage of the action-space behavior of stellar streams: in the true potential, each stream will cluster in action space[20]. For the purpose of measuring the degree of clustering, the Kullback-Leibler divergence is being used.

[19] see 2013arXiv1309.2418F
[20] see 2014IAUS..298..207S and Figure 5.11



(a) QE<92.0E-4 (b) QE<34.7E-5

(c) QE<14.3E-5 (d) QE<53.3E-6

FIGURE 5.10: Identifications obtained for the SOM of SDSS outliers using Gaia simulations. The clusters receive a black color when the distance between the outlier prototype and the corresponding template is above the established limit (from 2013arXiv1309.2418F).

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. This is on-going work.

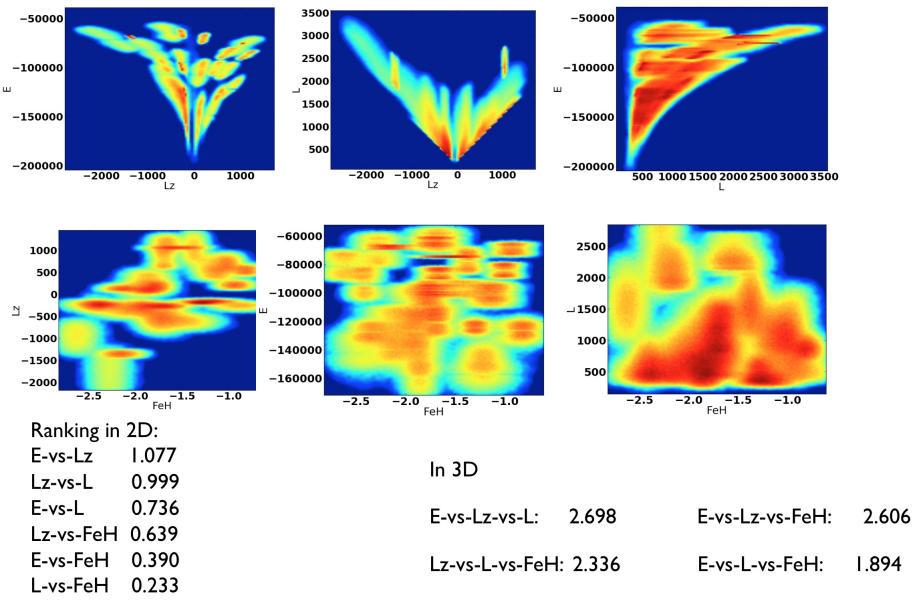


FIGURE 5.11: Example of ranking for the measure of clustering and correlation using the Kullback-Leibler divergence in the space made by energy, angular momentum and metallicity.

Task 5.6 - Transversal tools for special objects

Task leader: CNRS

Contributing partners: UNIGE, ULB

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.

As the publication of these objects will not occur at the first releases, some delay is thus available, and this subtask has just began its work. The definition of an engineer post-doc profile and the successful selection of a candidate is recent. M. Kudryashova for 24 month at CNRS/IMCCE, started in September 2014. First step is to acquire expertise in ephemerides computation, Gaia observations of SSOs, and then define the various specific tests, before their design in a second step.

Work on stellar clusters has also begun, with 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, of the problem cases for the first 3 tests and developing the first codes.

The Task 562 is associated to the validation of Non-Single Stars (NSS), under the responsibility of ULB. The output of DPAC CU4, in charge of the NSS data treatment is not always of the same nature as available from ground based data. Indeed, over, say, 5 years of observations, only a fraction of the orbit was possibly covered, thus leading to a solution which is not a genuine Keplerian one (e.g. acceleration with 7 or 9 terms).

It is therefore impossible to directly compare most of the Gaia solutions with the truth obtained with ground based observations. This situation will improve as new observations accumulate, making a smart validation of early solution mandatory.

The work which has been undertaken by the ULB team has thus mainly been oriented at building the foundations which will ensure a much easier comparison of the Gaia data to the ground-based ones. One way to do this has been done by setting up a framework to validate the NSS solution against simulated data (Gaia CU2 GOG) even when the former is not Keplerian. By filling up a GOG-like file with ground based solutions, the same framework will make the validation against such solutions possible.

Finally, the work concerning the validation of the variability analysis done at UNIGE (Task 563) has also started, although some delay is also available as variability will not be published at the first releases of the Catalogue. The administrative procedure to select candidate on the GENIUS position (partly funded by) has nevertheless been started, for a job position starting from 1st of December 2014 on.

The first data on variables that could be seen, for example in the Gaia “image of the week”, as well as what was shown at the Science Alert workshop in Warsaw on in September are of good enough quality so that it can be however envisioned to have releases of variable groups early, as mentioned in the release scenario and the validation process should now advance quicker.

3.2.2.5.3 Highlights

- Validation Test Specification document written
- Common validation environment installed
- Common interface software developed
- First tests implemented

3.2.2.5.4 Deviations and impact on tasks and resources

The deliverable D5.1, Delivery of prototype of WP5 internal checking tools (WP 5.2), has been delayed by 2 months, due to the unexpected departure of the GENIUS postdoc S. Boudreault mid August and the recruitment of a new postdoc two months later only in WP5.2.

3.2.2.5.5 Use of resources

The following table lists the person-month per participant in the first 12 months in WP5

	CNRS	**CSIC**	**UNIGE**	**ULB**	**FFCUL**	**KU**
WP5	16.8	0	0.7	4.5	0	0.5

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 an 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 coordination of this WP from the UB during the first year has been rather straightforward. On the one hand, the execution of simulations at CSUC has been a rather straightforward process since both the infrastructure and the software have been used for this purpose for several years. The coordination has just involved the tracking of progress and the checking of the formal deliveries of simulated data.

On the other hand, the activities on science alerts have been limited in this first year to the requirement analysis (they will start in full force next year, see next section). The GENIUS coordinator was invited[21] in Feb. 2014 to the PESSTO "GREAT ESF Workshop"[22] where the coordination of activities started and the integration on the main archive was defined. This was then followed by the participation of the Science Alerts task leader, Nic Walton, in the Vienna CU9/GENIUS meeting for preparation of the start of the development in 2015.

Task 6.2 - Simulated catalogue data

Task leader: CSUC

Contributing partner: UB

As described in the GENIUS proposal, the availability of simulations is crucial for the development

[21] although closely related to GENIUS, the trip and meeting is not included in this report as a GENIUS activity because it was funded by the GREAT FP7 project

[22] `http://wiki.pessto.org/pessto-wiki/pessto-meetings/gaia---pessto-great-esf-workshop`

of the systems and tools of the Gaia archive, and therefore for the GENIUS tasks: the real Gaia data will not be available until later in the schedule and the simulations are needed to test the systems.

With this purpose, the GOG simulator (developed for Gaia-DPAC since 2006 at the University of Barcelona) has been deployed at CSUC. The execution tests led to an execution of GOG that combined the computers at CSUC with the MareNostrum supercomputer at the *Barcelona Supercomputing Centre*, also available to the Gaia Barcelona team through the *Red Española de Supercomputación*:

a) The CSUC shared memory systems are used for the simulations of very dense regions of the sky, where the large number of objects require an intensive memory usage (provided by GENIUS).

b) The MareNostrum supercomputer, with large a large number of processors but with a limited amount of memory per processor, is used to generate the simulations for the lower density regions of the sky

This combination has allowed the generation of several simulations of the Gaia catalogue, the latest of which has recently been ingested in the main archive at ESAC and is being used for testing (see Luri et al. 2014A&A...566A.119L).

However, the development of GOG and its optimization in the last year have largely improved the performance of this simulator. The CPU time required for the generation of a full-sky catalogue simulation has been drastically reduced, by a factor of 10, and the memory handling has been improved so that very dense regions can now be simulated withouth requiring large amounts of memory.

Therefore, the next batch of simulations scheduled for 2015 (full sky with an updated error model taking into account the actual performance of the satellite as obtained from its operations) will require much less resources than expected in the original GENIUS planning. Thus, we will move resources from this task to T4.3, where the CSUC is contributing the data mining testbed and they are needed (see section 3.2.2.6.4 below).

Task 6.3 - Science alerts testbed

Task leader: UCAM

Overview

The Gaia flux-based science alert stream will be issued to the community through the science alert processing carried out at the Cambridge Photometric Data Processing Centre (DPCI). The science alerts processing will issue 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 will contain basic characterisation information for each event, including parameters such as estimated alert object type, and more advanced classification for certain objects such as su-pernovae (SNe). For these, inherent Gaia photometric data will be used to provide additional information concerning SNe alerts including class, epoch, redshift, reddening.

The testbed work to be carried out in T6.3 will develop 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 T7.2. With the termination of the GENIUS T6.3 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 1

Year one activity has involved an initial 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 has been activated, currently in a validation phase. See `http://gaia.ac.uk/selected-gaia-science-alerts`. This is the initial Gaia Alerts testbed - D6.2 Deployment of first public science alerts prototype .

The alerts prototype is now releasing photometric alerts to the community - with a significant follow up campaign underway, obtaining ground based follow up of the alerts in order to characterise the alerts (thus type then as supernova, flare stars, CV's etc).

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 - is being 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 1

One week of UCAM staff effort has been charged to GENIUS WP6 in Year 1, representing attendance at and related preparation for the GENIUS Kick Off meeting in Vienna (July 2014).

In addition, travel costs related to attendance at the GENIUS kickoff meeting 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 supported development work will occur in years 2 and 3.

3.2.2.6.3 Highlights

- Delivery of the first full final catalogue simulation to ESAC for ingestion into the archive prototype.
- Definition of the integration of the Science Alerts data from UCAM into the main Gaia

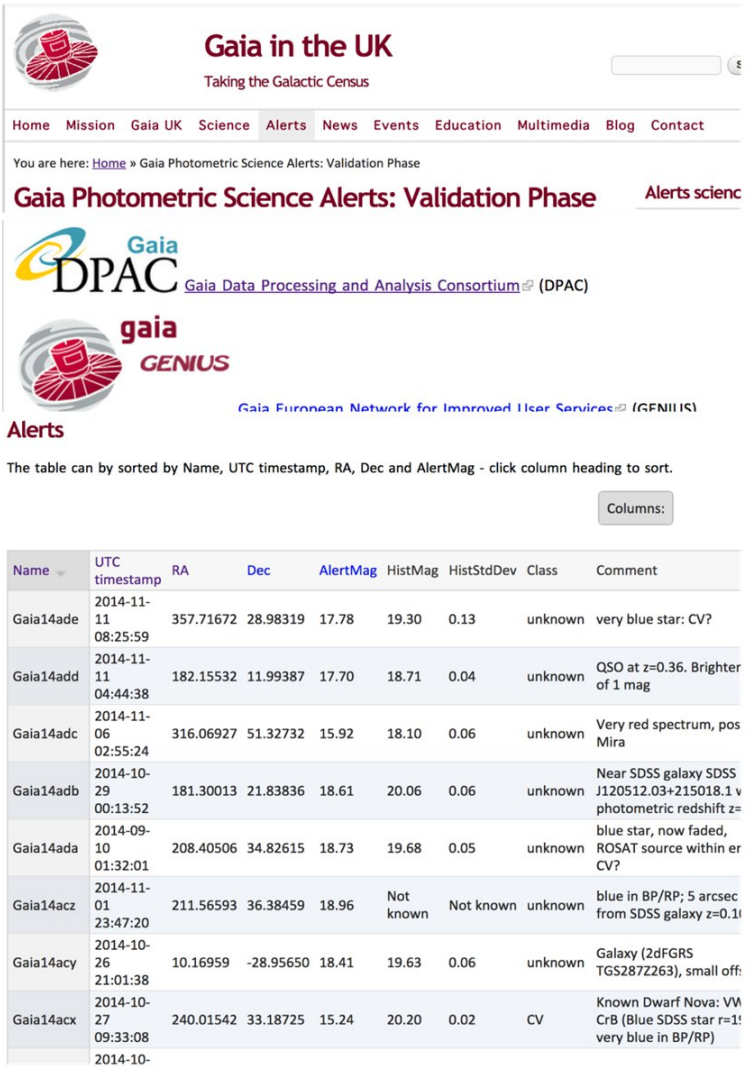


FIGURE 6.12: The Science alerts portal acting as the initial testbed for the CU9/GENIUS alerts archive

archive. Development of a Data Model to include this data in the main archive schema.

3.2.2.6.4 Deviations and impact on tasks and resources

As mentioned in the previous sections, after a strong optimization of the simulator the generation of simulated Gaia catalogues requires much less computing resources that planned. Therefore, in order to take full advantage of the resources available in CSUC, we will transfer the manpower originally allocatad to this work package in that centre to T4.3 to support the upgrading of the data mining testbed that is required for the next GENIUS phases (see section 3.2.2.4.4).

3.2.2.6.5 Use of resources

The following table lists the person-month per participant in the first 12 months in WP6

	UB	CSUC	UCAM
WP6	1.3	2.81	0.25

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

At present, almost all the partners of the Genius project have web portals devoted to the Gaia mission, including in several cases specific sections for outreach and academic activities. We have compiled all this material to included it in the Community portal, together to a contact person. This list of outreach material will help to share and coordinate initiatives from different countries.

On the other hand, from the very first moment we have open the community portal to everybody involved in the project. In this sense, the main sections of the Community portal were chosen with the contributions of different people from GENIUS project, DPAC and ESA, via an VoteIt survey.

Finally we have set up an editorial board for the high-level supervision of the portal contents. It is formed by: Gráinne Costigane (Leiden), Frederic Arenou (Meudon), Mariateresa Crosta (INAF), Heather Campbell (University of Cambridge) and Eduard Masana (UB). A second level of collaborators started in September to coordinate the content provision and translation into several languages, and we have invited members of the German and Portuguese teams to join the network of collaborators. The editorial board will meet periodically.

Task 7.2 - Community portal infrastructure

Task leader: CSUC (formerly CESCA)

CSUC has assigned two virtual machines to host the two available project websites:

`www.genius-euproject.eu` and `www.gaiaverse.eu`. The reason behind setting up two different portals is to decrease difficulties for users by separating formal information on the project itself from dissemination done within the project. On the one hand, the administrative website (www.genius-euproject.eu) provides, in English, information about the general purpose of GENIUS and participant partners, as well as contact information and the twitter timeline. On the other hand, the community portal (`www.gaiaverse.eu`) is aimed to provide enhanced dissemination tools for the scientific community but also to bring astronomy to the general public and to provide resources for teaching astronomy. Both sites are based on WordPress Content

Management System (CMS), although efforts dedicated differ considerably one from the other. The www.genius-euproject.eu website has only four sections with permanent information (Home, About Genius, Partners and Contact) which was published a few weeks after the beginning of the project, whereas the process to launch www.gaiaverse.eu has been more complex with the aim to fulfil researchers' dissemination needs, as it was explained in the DOW.

In September 2013, a first technical note including proposals on the community portal contents was released. After several teleconferences and meetings, including coordination with ESA at ESTEC in December, it was come to the decision of providing a multilingual portal which brings outreach to a broader audience, and an easy-to-use platform, to facilitate writing, editing and translation work to collaborators. It is worth to highlight that ESA is doing well with dissemination activities, but they are mainly in English, so to avoid duplicities and to get the community portal brings added value, www.gaiaverse.eu will focus on providing well organised resources in several languages, not only materials prepared by GENIUS partners but also by other institutions participating in Gaia. Thus, we expect to help to strengthen dissemination done within Gaia, providing a channel focused on providing resources and highlighting news in a multilingual approach addressed to researchers, amateurs' astronomers and teaching professionals.

From ESTEC meeting onwards, development and implementation of the community portal began to take shape. Between December 2013 and January 2014, an internal evaluation phase to choose the suitable Content Management Strategy (CMS) was carried out. After several tests analysing different CMS, WordPress was elected as the best option to develop the community portal. It is an easy-to-use platform which allows multilingual editing and access to many users at the same time. This is a twofold goal, since the community portal will be edited from different countries. In that way, the portal must allow different users to be logged-in at the same time and has to be presented in many languages, functionalities that WordPress makes possible. Finally, as researchers themselves will create contents and will translate those with great interest for their country, it is important to emphasize the need of having an easy-to-use platform, where both editing and translation are easy to perform.

A survey (via VoteIt tool) has been run also to consult GENIUS members, as well as some DPAC and ESA members, on the main contents to be highlighted on the community portal. Thus, feedback on how to place and emphasize contents in the community portal was taken into account to prepare the first community portal layout. First users with editing rights to access the community portal were created late March, while the first public version of the community portal came to light in March 28th. Other editors were involved in April.

With the first public version of the community portal available, our main objective was to get users involved in the portal. Thus, it was decided to set up an editorial board representing centres from different languages and developing different tasks within the GENIUS project (see Task 7.1). First months using the platform have been useful to develop user tests and to plan further improvements taking into account users' experience. Thus, WordPress template will be modified to take more advantage of horizontal space and resources' search and visibility will be enhanced also. The advance version of the portal will be available in January 2015.

Task 7.3 - Community portal, outreach and academic activities

Task leader: UB

Contributing, partner: CNRS

There have been several meetings and teleconfs to set up the outreach and academic activities, some of them in coordination with the WP960 of the CU9, also devoted to the outreach activities of the Gaia mission (Vienna, July 2014). As the first step it was decided to compile all the existing outreach material, identify a contact person and ask for contributors to translate the material to other languages.

With the publication of the Community portal the project will have a platform to publish the latest news related to the Gaia mission and the project itself. A good coordination with ESA is essential, as it was exposed in the ESTEC meeting (December 2013). At the same time, the editorial board will take a pro-active attitude, contacting with different people to ask for contributions to the community portal.

3.2.2.7.3 Highlights

- Compilation of outreach material and academic activities from different teams.
- Set up of the editorial board of the Community portal.
- Survey (via VoteIt) to set the main contents of the web portal.
- Survey (via VoteIt) to choose the domain name (gaiaverse.eu).
- Development and implementation of the web portal.
- First public version of the portal available.

3.2.2.7.4 Deviations and impact on tasks and resources

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

3.2.2.7.5 Use of resources

The following table lists the person-month per participant in the first 12 months in WP7

	UB	**CSUC**	**CNRS**
WP7	2.3	2.04	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.

In this reporting period most of the management effort has been devoted to starting the work in the different GENIUS work packages and to establish its coordination and management tools, some specific for GENIUS and some shared with ESA and the Gaia DPAC.

Note: the ESA/DPAC systems are protected and access is restricted by a Non Disclosure Agreement. The GENIUS reviewers and project officer have been provided access to the ESA/DPAC information systems by a special agreement. The GENIUS participants are considered full CU9 members and also have access to these systems.

The first of such tools has been the GENIUS Twiki:

`https://gaia.am.ub.es/Twiki/bin/view/GENIUS/.`

This web site, based on the Twiki system, allows collaborative edition of its pages; each WP manager has been given edit access and the internal information of GENIUS has thus been jointly maintained, including the tracking of deliverables and milestones, meeting pages, GENIUS hiring and general project information.

The information in the GENIUS Twiki is complemented by the DPAC wiki pages, maintained by ESA and the DPAC consortium, and specifically on its CU9 section:

`http://wiki.cosmos.esa.int/gaia-dpac/index.php/CU9:_Catalogue_Access`

In these pages more technical and detailed information regarding the development of the archive systems is provided, as well as general information about the ongoing Gaia data reduction. Specific sections of the DPAC wiki are cited in some cases in this review document.

Thirdly, the technical GENIUS documentation is directly contributed to the CU9 for wide use. It is stored into the ESA/DPAC documentation system, LiveLink, and is accessible through ESA RSSD portal:

`http://www.rssd.esa.int/`

Finally, the code for the archive systems is integrated into the ESA/DPAC code repository (based on the *Subversion* code revision management system):

`http://gaia.esac.esa.int/dpacsvn/`

In this first year of GENIUS all of its participants have been given access to these systems and trained in its use, and they have become the baseline development and coordination tools for the project.

Besides the implementation and deployment of these tools in the consortium, several consortium meetings and teleconferences have allowed the coordination and progress tracking of the GENIUS developments. They are listed in the next section but we would like to highlight two main events:

- The GENIUS Kick-Off Meeting, held in Barcelona 4-5 December 2013. All GENIUS partners participated and it started the activities of the project: `https://gaia.am.ub.es/Twiki/bin/view/GENIUS/KickoffMeeting`
- The joint CU9-GENIUS plenary meeting held in Vienna, 7-8 July 2014. This large meeting counted with the participation of all the GENIUS partners and a large fraction of the CU9 members: `https://gaia.am.ub.es/Twiki/bin/view/GENIUS/PlenaryMeeting2014`

Regarding the organisation of GENIUS meetings we want to highlight here that one of the GENIUS goals regarding the organization of its work is to favour a good work-life balance (section 5.1 of the GENIUS proposal) by allowing for more but shorter trips and virtual international contacts. We have tried to follow these guidelines and we have also acquired a web-based videoconference system (WebEx) and a pair of microphones (as agreed in the kick-off meeting) that have facilitated the exchanges with the organization of teleconferences (both global and at WP level) and many shorter virtual contacts.

Related to this, we have also tried to keep a good gender balance in the project. As already stated in the GENIUS proposal, this is usually a hard task in computer science projects, where the IAU statistics show that a 16% female staff is the average. In the GENIUS hiring we have managed to improve this percentage to 28% in the newly hired personnel (see `https://gaia.am.ub.es/Twiki/bin/view/GENIUS/VacanciesGenius`.

Regarding the deliverables and milestones of the project, they have been included in the FP7 portal but can not be tracked due to problems in the newly implemented portal system. 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`

Finally, following the project structure defined in the proposal, we have set up an external advisory board to provide independent advice about the GENIUS project. This board is composed by:

William O'Mullane: Head of Operations Development Division at European Space Agency. http://www.cosmos.esa.int/web/gaia/gaia-people/william-o-mullane

Françoise Genova: director of Strasbourg astronomical data centre CDS. https://www.rd-alliance.org/about/organization/key-profiles/fran%C3%A7oise-genova.html

Tadafumi Takata: Associate professor of Astronomy Data Center, NAOJ

Mark Wilkinson: Royal Society University Research Fellow and Lecturer in Theoretical Astrophysics, University of Leicester http://www2.le.ac.uk/departments/physics/people/markwilkinson

The advisory group will start its activities in early 2015. They will receive the GENIUS reports and will participate in the mid-term meeting; from these inputs they will provide a report on the progress of GENIUS with feedback and advice for improvements.

3.2.3.2 Problems occurred and applied solutions

The following problems presented during the first GENIUS year:

- The Gaia launch was successful, removing one of the major risks of the project. However, as expected in such a complex endeavour as a space mission, the first months of commissioning revealed unexpected problems and features of the satellite (see for instance http://blogs.esa.int/gaia/2014/02/12/one-month-at-l2/). This has made the ESA/DPAC schedule for data releases change, having an impact on the GENIUS planning, tied to these data releases. The changes and solutions to this problem are discussed in detail in section 3.2.3.5.
- Problems with personnel: in some cases, the process of hiring of personnel for the project has been more lengthy than expected. This has been specially the case for a position at the Observatory of Besançon, where several tentative candidates refused the position in succession. This was mitigated by the initial 6-month period of project set-up included in the planning and finally solved with extended advertising of the position and specific contacts in the Gaia community, through the DPAC and ESA channels, allowing the filling of the position. On the other hand, a full-time postdoc position at the Paris Observatory was unexpectedly vacated a few months after its start, leading to some delays in the associated tasks reported in previous sections. We are now in the process of covering again this position.
- The CSIC partner had internal administrative problems in the allocation of the GENIUS budget. The centre participating in GENIUS, the *Centro de Astrobiologia*, was not allocated the initial payment funding until August 2014, in spite of it having been transferred by the University of Barcelona to the CSIC in due time. This has delayed the hiring of personnel in this institute, introducing some delay in the adaptation of the VOSA tool for Gaia, as described in previous sections. We will try to recover this delay during the first half of 2015 in time for the first Gaia data release, but in the worst case the tool will be ready for the second data release.

- We have made a special effort to try to correct the gender gap present in IT related projects like GENIUS, but this has been proven a hard task. As stated above, we have managed to improve the gender balance with respect to the average on this type of projects, but we are not reaching parity. We will keep working on this issue in the next hirings to try to further improve the balance.
- A part of the budget allocated to the UB, devoted to DB license acquisition and specific formation on DB systems, has not been spent due to changes in the technical specifications of the main archive. We would like to reallocate the funding to other tasks in GENIUS where we have found important shortcomings in the initial budget allocation. More details are given in section 3.2.3.5.

3.2.3.3 Changes in the consortium

The consortium has not been changed in this reporting period. We want only to remark that one of the partners has changed its name: the Centre de Supercomputació de CAtalunya (CESCA) has become the Consorci de Serveis Interuniversitaris de Catalunya (CSUC). This change has been duly notified to the European Commission and the new name will be used in the future in GENIUS.

3.2.3.4 List of project meetings, dates, venues and participants

Kind of meeting	Location	Date	Participants
Gaia Archive Workshop	ESAC (Madrid)	27-29/11/2013/	X. Luri, SAT members
GENIUS Kick Off Meeting	University of Barcelona	04-05/12/2013	F. Arenou, D. Hestroffer, L. Balaguer, N. Benitez, T. Branza (PO), A.G.A. Brown, L. Eyer (videocon), A. Gil, G. Gracia, N. Hambly, F. Julbe, X. Luri, E. Masana, A. Moitinho de Almeida (videocon), R. Smart, E. Solano, D. Pourbaix (videocon), M. Taylor, J. Torra, T. Via, N. Mowlavi (videocon), Y. Yamada (videocon)
Meeting to coordinate outreach activities, including portal contents approach	ESTEC	10/12/2013	A. Brown, S. Jordan, S. Voght, X. Luri, E. Masana, T. Via
Definition of the archive data model with the Science Archives Team	ESAC (Madrid)	22/01/2014	X. Luri, SAT members
GENIUS - NanoJASMINE Meeting. WP2. Discuss Science output of NanoJASMINE and JASMINE.	University of Tokyo (Japan)	27/01/2014	Y. Yamada
Coordination Meeting 6.3 Science Alerts testbed	Belfast	05-06/02/2014	X. Luri, N, Walton
GENIUS - NanoJASMINE Meeting. WP5. Confirm hardware and ground station status and evaluate impact of scientific performance.	University of Tokyo (Japan)	17/02/2014	Y. Yamada
Meeting to define the data mining requirements, use cases and possible architectures	CAB-INTA (Madrid)	17/02/2014	F. Julbe, L. Sarro, B. Montesinos
GENIUS CU9 Database and Server	Bologna (Italy)	25/02/2014	R. Smart

Kind of meeting	Location	Date	Participants
GENIUS - NanoJASMINE Meeting. WP2. Discussion of Science output of Nano-JASMINE, and model driven data analysis technique.	University of Tokyo and National Astronomical Observatory (Japan)	05-07/03/2014	Y. Yamada
Coordination meeting GENIUS -CU9 SAT	ESAC (Madrid)	04/04/2014	X. Luri, SAT members
GENIUS Discussion on Cross-Matching procedures	Roma (Italy)	22-25/04/2014	R. Smart
GENIUS AGIS Meeting. WP2	Dresden (Germany)	11-16/05/2014	Y. Yamada
GENIUS - NanoJASMINE Meeting. Discussion of defining requirement of combined catalogue of Gaia and Nano-JASMINE. WP2	National Astronomical Observatory (Japan)	26-27/06/2014	Y. Yamada
GENIUS - NanoJASMINE Meeting. WP2. Confirm hardware development status of Nano-JASMINE, and discuss science output of Nano-JASMINE and JASINE	University of Tokyo and National Astronomical Observatory (Japan)	24-25/04/2014	Y. Yamada
Joint CU9 - GENIUS Plenary Meeting	Vienna University Observatory	07-10/07/2014	X. Luri, F. Arenou, A.G.A. Brown, Y. Yamada, N. Hambly, A. Moitinho, A. Krone-Martins, M.D.D. Ferreira-Gomes, M. Taylor, N. Walton, G. Costigan. T. Via, G. Roldan, F. Julbe, S. Boudreault, L. Ruiz-Dern, H. Ziaeepour, C. Reyle, C. Babusiaux among others
GENIUS - NanoJASMINE Meeting. Discussion for Development of data analysis software. WP2	Tokyo University of Marine Science and Technology (Japan)	27-28/08/2014	Y. Yamada
GENIUS CU9 Data mining	ESAC (Madrid)	01-02/09/2014	F. Julbe, D. Tapiador, L. Sarro, B. Montesinos
Gaia Science Alerts Meeting. WP2	Warsaw (Polland)	09-14/09/2014	G. Costigan, A. Brown

3.2.3.5 Project planning and status

This first year of project has included, as defined in the GENIUS proposal, a six-month period of project set-up (previous to the start of the development) where all the coordination tools and procedures have been implemented, and the hirsing of personnel has been carried out. The implementation of the project started in parallel, and was at full speed at the end of this six month period.

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. This approach was based on the initial schedule for the Gaia data releases, where a first release was scheduled for mid-2015 and another one for mid-2016, with some intermediate versions of the archive in between based on simulated data.

However, the successful launch of Gaia was followed by a commissioning phase where several problems were discovered. None of these problems is critical and they will not prevent the success of the mission (although they have had some impact on the expected mission performances, see `http://www.cosmos.esa.int/web/gaia/science-performance`), but have forced a revision of the data release schedule. This revision has been discussed in the last months and formally approved in September 2014, and involves a delay of about one year in the official data releases, see: `http://www.cosmos.esa.int/web/gaia/release`, leaving the first one for mid-2016. This would have had an impact on the GENIUS planning, because no actual working versions of the archive would have been implemented until the late phases of GENIUS; a possible solution will be to simply test the GENIUS developments and the intermediate archive versions with the simulated data provided by the GOG simulator (see Section 3.2.2.6.2) but fortunately more recent developments have opened an alternative. An internal data release exercise has recently (November 2014) been introduced, with a limited number of stars but with Gaia results for the first year of operations. This exercise will be scheduled for mid-2015. Therefore, we will have a test with real data (with some limitations on the number of stars), again aligned with the GENIUS developments. Thus, we can expect to test with real data the first batch of operational GENIUS products in this mid-2015 exercise, and then the final products with the 2016/2017 data releases.

Besides this planning re-adaptation, we want to report a deviation from the original planning in spending and propose a set of changes to adapt to it. As already discussed in section 3.2.2.4.4, the original budget design was based on the assumption (valid at the time) that the main archive system at ESAC would use the Intersystems Cache DB or alternatively the Oracle DB system. Both these systems require commercial licenses for its usage, and even for academic or scientific use they are quite expensive. Consequently we allocated a budget (based on license pricing provided by Intersystems Cache and Oracle) for the acquisition of licenses for usage in GENIUS. However, the final implementation of the main archive at ESAC uses a different DB system, PostGres. This is an open-source system for which free licenses are available, therefore eliminating the licensing cost initially expected. Furthermore, the CSUC partner already has a good experience in this DB system and can provide all needed expertise in it without the need for expensive formation from Intersystems cache or Oracle. Therefore, the funding for these licenses has not been spent in the first year and we are now in a position to fully confirm that such licenses will not be needed for the next phases of the project. We thus propose to reallocate these funds to other areas of GENIUS

where important shortcomings have been detected:

- As explained in Section 3.2.2.4.4, The initial planning for the development of the data mining tools included a small test bed at CSUC for limited proof of concept, which has been operating there in this first year. However, the actual experiments have quickly reached the limits of the testbed, both in terms of disk storage and scalability, and full proof of concept will require in several use cases a more poweful testbed. Thus, we propose to reallocate about 40% of the above mentioned budget to the upgrading of the data mining testbed; this will allow to overcome the bootlenecks that the data mining activities would face in the next year if the current testbed is not upgraded. We also propose to devote the remaining 60% of the budget to increase the UB manpower in the data mining activities to support the increase in the scale of the experiments.
- Complementing this, we also propose to reallocate some of the FTEs of the CSUC partner currently in WP6 and move them to WP4. As explained in Section 3.2.2.6.2, thanks to the latest optimizations of the GOG simulator the effort needed for the generation of simulated data has been significantly reduced, and with it the need of resources at CSUC in WP6. On the contrary, a larger testbed with increased data handling and complexity will require more resources for its operation, and therefore more CSUC resources in WP4. Thus we propose a reallocation from one to the other.

A proposal will be made to the Project Officer after the first year review for the implementation of these two changes in the overall GENIUS planning.

3.2.3.6 Impact of deviations from the planned milestones and deliverables

The deviations from planned milestones and deliverables in this first year have been minor. Their impact has been recovered or will be in the next months with specific actions. The relevant effects on planning and schedule have been highlighted in the previous section.

3.2.3.7 Changes to the legal status of beneficiaries

No changes in the legal status of the beneficiairies.

3.2.3.8 Use of resources

The following table lists the person-month per participant in the first 12 months in WP1

	WP1	**WP2**	**WP3**	**WP4**	**WP5**	**WP6**	**WP7**
UB	3.8	-	-	10	-	1.3	2.3
CNRS	-	-	0	-	16.78	-	0
UEDIN	-	-	7.9	-	-	-	-
UL	-	7.07	-	-	-	-	-
CSUC	-	-	-	-	-	2.81	2.04
INAF	-	0.59	5.51	-	-	-	-
CSIC	-	-	0	0.25	0	-	-
UNIGE	-	-	-	.	0.7	-	-
ULB	-	-	-	-	4.5	-	-
FFCUL	-	0	-	2.9	0	-	-
UBR	-	-	-	3	-	-	-
UCAM	-	0	-	-	-	0.25	-
KU	-	1.0	-	-	0.5	-	-

The following table lists the actual efforts (**AE**) for the first 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	3.8/18	-	-	10/40	-	1.3/6	2.3/7
CNRS	-	-	0/1.8	-	16.8/77.6	-	0/1.8
UEDIN	-	-	7.9/43	-	-	-	-
UL	-	7.07/49	-	-	-	-	-
CSUC	-	-	-	-	-	2.81/12	2.04/12
INAF	-	0.59/16	5.51/18	-	-	-	-
CSIC	-	-	0/6	0.25/12	0/4	-	-
UNIGE	-	-	-	-	0.7/6.5	-	-
ULB	-	-	-	-	4.5/6	-	-
FFCUL	-	0/2	-	2.9/28	0/2	-	-
UBR	-	-	-	3/3	-	-	-
UCAM	-	0/4	-	-	-	0.25/12	-
KU	-	1/2	-	-	0.5/2	-	-

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 first 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**
D1.1	Kick Off Meeting	WP1	Dec 2013
D1.2	Semestral report 1	WP1	May 2014
D1.3	Semestral report 2	WP1	Nov 2014
D2.1	Requirement specification for catalogue and data service	WP2	Oct 2014
D2.2	Requirement specification for outreach facilities built into the system archive system	WP2	Oct 2014
D3.1	GENIUS ESAC-SAT Coordination and Interface Control document	WP3	Jan 2014
D4.1	Requirement specification document for the exploitation tools	WP4	Nov 2014
D4.2	Delivery of the first prototype of exploitation tools	WP4	Nov 2014
D5.1	Delivery of prototype of internal checking tools	WP5	Nov 2014
D6.1	Delivery of first simulated catalogue data	WP6	May 2014
D6.2	Deployment of first public science alerts prototype	WP6	Oct 2014
D7.1	Basic setup for the community portal internally available for working	WP7	Oct 2013
D7.2	First public version of the community portal	WP7	Mar 2014

3.3.2 Milestones in the first 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**
MS1	Kick Off Meeting (plenaary)	WP1	Dec 2013
MS2	Agreed testbed environment with ESAC-SAT and CU9	WP3	Jan 2014
MS3	Hiring of main developers	WP1, WP2, WP3, WP4, WP5, WP6, WP7	Jun 2014
MS4	GENIUS portal available at CSUC for internal use	WP6, WP7	Mar 2014
MS5	Archive user requirements document	WP2	Oct 2014
MS6	Requirements document for each subsystem	WP2, WP3, WP4, WP5	Nov 2014
MS7	Public Version of GENIUS portal	WP6, WP7	Mar 2014