

Gaia European Network for Improved User Services GENIUS

Collaborative project

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

Xavier Luri, Universitat de Barcelona

Part. no.	Organisation name	Country
1	Universitat de Barcelona (UB)	Spain
2	Centre National de la Recherche Scientifique (CNRS)	France
3	University of Edinburgh (UEDIN)	United Kingdom
4	Universiteit Leiden (UL)	Netherlands
5	Centre de Serveis Científics i Acadèmics de Catalunya (CESCA)	Spain
6	Istituto Nazionale di Astrofisica (INAF)	Italy
7	Consejo Superior de Investigaciones Científicas (CSIC)	Spain
8	Université de Genève (UNIGE)	Switzerland
9	Université Libre de Bruxelles (ULB)	Belgium
10	Fundação da Faculdade de Ciências da Universidade de Lisboa (FFCUL)	Portugal
11	University of Bristol (UBR)	United Kingdom
12	University of Cambridge (UCAM)	United Kingdom
13	Kyoto University (KU)	Japan

Foreword

This proposal is a resubmission of the original GENIUS proposal in the previous FP7 call. The main changes in the proposal are:

Taking into account the referees' comments: we have made an effort to correct the issues pointed out by previous referees. Specifically:

- The risk mitigation has been made more explicit in the requested cases
- Conflict resolution plan will be included in the Partnership Agreement
- Resource allocations in WP200 have been reviewed
- Recommendations for negotiation taken into account as far as possible in the present stage

Addition of two new partners: two new partners have been added to the proposal (University of Cambridge and Kyoto University) to cover two relevant tasks added to the proposal, science alerts from Gaia and strengthening of the cooperation with the only other two astrometric missions in the world aside from Gaia, the Japanese Nano-JASMINE and JASMINE.

Redefinition of WP300: one year after the submission of the original proposal the role of the ESA team in charge of the Gaia archive at ESAC has been more clearly defined, as well as its interaction with the scientific community. We have accordingly redefined WP300 to avoid duplication of efforts.

Better treatment of outreach and dissemination: we have improved the outreach and dissemination tasks in GENIUS, specifically by including provisions for a community web portal that should guarantee the visibility and availability of the GENIUS products.

List of acronyms

API	Application Programming Interface (software engineering)
ASTRONET	http://www.astronet-eu.org
BP/RP	Blue Prism / Red Prism (Gaia spectrophotometer context)
CU	Coordination Unit (in the DPAC context)
DBMS	Database Management System
DM	Data Mining
DPAC	Data Processing and Analysis Consortium (Gaia)
DPACE	DPAC Executive committee (Gaia)
EC	European Commission
ESA	European Space Agency
ESF	European Science Foundation
ESO	European Southern Observatory
ESAC	European Space Astronomy Center
FBG	Fundació Bosch i Gimpera (in UB context)
GAP	Gaia Archive Preparation group
GENIUS	Gaia European Network for Improved User Services (this proposal)
GAP	Gaia Archive Preparation group (in the ESA/DPAC context)
GEANT	http://www.geant.net/About_GEANT/pages/home.aspx
GOG	Gaia Object Generator
GREAT	Gaia Research for European Astronomy Training (FP7-ITN)
GST	Gaia Science Team
GTC	Gran Telescopio Canarias (Canary Islands, Spain)
HPC	High Performance Computing
HR	Hertzsprung Russell (diagram)
IPR	Intellectual property Rights
JAXA	Japan Aerospace Exploration Agency
JDBC	Java Database Connectivity
KD	Knowledge Discovery
LSST	Large Synoptic Survey Telescope
MDB	Gaia Main Database (in the DPAC context)
MLA	Multi Lateral Agreement (in the DPAC context)
NAOJ	National Astronomical Observatory of Japan
QSO	Quasi Stellar Object
PC	Project Coordinator
PI	Principal Investigator
PM	Project Manager
PS	Project Scientist
REG	Red Española de Gaia
RDBMS	Relational Database Management System
SSO	Solar System Object
TAP	Table Access Protocol (VO context)
UMR	Unite Mixte de Recherche (CNRS context)
VLT	Very Large Telescope (ESO, Chile)
VO	Virtual Observatory
VTK	Visualization ToolKit
WP	Work Package

Contents

1	Scientific and/or technical quality, relevant to the topics addressed by the call	6
1.1	Concept and objectives	6
1.1.1	Context: the Gaia mission	6
1.1.2	Mission structure	7
1.1.3	The Gaia archive & CU9	10
1.1.4	Collaboration with other space powers	10
1.1.5	Objectives	11
1.1.6	Compliance with the SPA.2013.2.1-01 objectives	12
1.2	Progress beyond the state-of-the-art	15
1.2.1	Archive interface	15
1.2.2	Virtual machines and light clients	15
1.2.3	Advanced statistics	16
1.2.4	Multidimensionality and visualization	16
1.2.5	Outreach and simplicity	16
1.2.6	Interoperability and reprocessing	17
1.2.7	Virtual observatory	17
1.3	S/T methodology and associated work plan	18
1.3.1	Overall strategy	18
1.3.2	Timing	20
1.3.3	Work description	22
	Table 1.3a - Work package list	22
	Table 1.3 b - Deliverables list	22
	Table 1.3c - List of milestones	25
	Table 1.3d I - WP100	26
	Table 1.3d II - WP200	27
	Table 1.3d III - WP300	31
	Table 1.3d IV - WP400	35
	Table 1.3d V - WP500	41
	Table 1.3d VI - WP600	47
	Table 1.3e - Summary of staff effort	50
1.3.4	Dependencies diagram	50
1.3.5	Risk analysis	51
	Summary (see below for a detailed discussion of each risk)	51
	External risks	51
	Internal risks	52
2	Implementation	54
2.1	Management structure and procedure	54
2.1.1	Management	54
2.1.2	Procedures and reporting	56
2.1.3	Partnership agreement	57
2.1.4	Recruitment strategy	57
2.2	Individual participants	58
2.2.1	Universitat de Barcelona (UB)	58
2.2.2	Centre National de la Recherche Scientifique (CNRS)	59
2.2.3	The University of Edinburgh (UEDIN)	60

2.2.4	Universiteit Leiden (UL)	61
2.2.5	Centre de Serveis Científics i Acadèmics de Catalunya (CESCA)	62
2.2.6	Istituto Nazionale di Astrofisica (INAF)	63
2.2.7	Consejo Superior de Investigaciones Científicas (CSIC)	64
2.2.8	Université de Genève (UNIGE) - Observatoire de Genève	65
2.2.9	Université Libre de Bruxelles (ULB)	65
2.2.10	Fundação da Faculdade de Ciências da Universidade de Lisboa (FFCUL)	66
2.2.11	University of Bristol (UBR)	67
2.2.12	University of Cambridge - Institute of Astronomy (IoA)	68
2.2.13	Kyoto University (KU)	69
2.3	Consortium as a whole	70
2.3.1	Third parties	72
2.3.2	Sub-contracting	73
2.3.3	Additional partners	74
2.4	Resources to be committed	74
2.4.1	Overview: budget	74
2.4.2	Overview: personnel	75
2.4.3	Universitat de Barcelona (UB)	76
2.4.4	Centre National de la Recherche Scientifique (CNRS)	76
2.4.5	The University of Edinburgh (UEDIN)	76
2.4.6	Universiteit Leiden (UL)	77
2.4.7	Centre de Serveis Científics i Acadèmics de Catalunya (CESCA)	77
2.4.8	Specialised partners	77
3	Impact	78
3.1	Expected impacts listed in the work programme	78
3.2	Dissemination and/or exploitation of project results, and management of intellectual property	80
3.2.1	Dissemination of GENIUS results	80
3.2.2	Exploitation and intellectual property	81
4	Ethics issues	83
5	Consideration of gender aspects	85
5.1	Providing for good work-life balance in GENIUS	85
5.2	GENIUS at university and faculties of computer science	85
6	Bibliography	86
7	PROPOSAL END	88

1 Scientific and/or technical quality, relevant to the topics addressed by the call

GENIUS in a nutshell

GENIUS is designed to boost the impact of the next European breakthrough in astrophysics, the Gaia astrometric mission. Gaia is a European Space Agency Cornerstone mission which is scheduled for launch in the final quarter of 2013 and aims at producing the most accurate 3D map of the Milky Way to date (Sec. 1.1.1). A pan-European consortium named DPAC, funded at the national level, is working on the implementation of the Gaia data processing (Sec. 1.1.2), of which the final result will be a catalogue and data archive containing one billion objects.

The archive system containing the data products will be located at the European Space Astronomy Centre (ESAC) and it will serve as the basis for the scientific exploitation of the Gaia data. The design, implementation and operation of this archive is a task that ESA has opened to participation from the European scientific community in 2012 (Sec. 1.1.3).

GENIUS is aimed at significantly contributing to the development of this archive based on the following principles: archive design driven by the needs of the user community that will scientifically exploit the Gaia results; 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; cooperation with the only other two astrometric missions in the world, Nano-JASMINE and JASMINE (Japan); and last but not least, the archive facilitates outreach activities (Sec. 1.2).

GENIUS is defined to fit 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 the relevant expertise in key development areas. Furthermore, GENIUS has the support of DPAC, several Gaia national communities in the EU member states, and will cooperate with the Japanese astrometric missions already mentioned. Letters attesting to this support have been sent to the coordinator.

GENIUS optimally fulfills the objectives of the SPA.2013.2.1-01 call (see Sec. 1.1.6 for a detailed review of its compliance)

1.1 Concept and objectives

1.1.1 Context: the Gaia mission

This proposal is devoted to the development of the archive of ESA's astrometric mission Gaia.

Gaia aims to study the origin, formation, and evolution of the Milky Way and its components [17]. The Milky Way is composed of a disk of about 50000 light years in radius, containing stars of many types and ages as well as interstellar gas and dust; a spheroidal halo of some 100000 light years in radius, containing very old stars; a bar and a bulge in the centre; and the ubiquitous dark matter. The disk contains spiral structure, with the spiral arms being the preferred location for star forming regions. Although these general features are rather well known, much remains to be elucidated, including the detailed number and structure of the spiral arms, the disk warp, the detailed shape and rotation of the bulge, disk, and halo, the dynamics and the kinematics of the Galaxy, and the distribution of dark matter. In addition, the process by which the Galaxy was assembled, presumably from small building blocks, and the history of star formation are not well understood.

To address these and many other questions, one needs a deep all sky survey covering a significant volume of the Galaxy, and providing 3D positions and velocities at high accuracy as well

as the physical properties of the objects observed. The goal of the Gaia mission is to create the largest and most accurate three-dimensional survey of our Galaxy and beyond by providing unprecedented positional, proper motion, radial velocity, and spectroscopic data for about one billion stars in our Galaxy and throughout the Local Group.

In addition to the in-depth study of the Milky Way, the Gaia survey will also contribute to many other areas of astronomy and astrophysics. Just to mention the most relevant ones [17]: star formation history, stellar structure and evolution, stellar variability, stellar ages and the age of the universe, the distance scale, binaries and multiple systems, planetary systems, solar system objects and even some fundamental physics applications such a test of General Relativity. Furthermore, since the Gaia survey will be uncensored it will also contain some millions of extragalactic objects (galaxies and QSOs) that will contribute to extragalactic astrophysics and to the redefinition of the optical astrometric reference frame.

The launch of Gaia is currently scheduled for end 2013¹ and the nominal mission ends 5 years later. The detailed design and construction phase started in February 2006, with development and construction progressing as planned. Gaia is a survey mission and the operating principle is the same as that of its successful predecessor, the Hipparcos satellite of the European Space Agency (ESA), which has provided a rich harvest of science since the publication of its catalogue in 1997 [15]. Although in retrospect Hipparcos seems like a modest project, it was the first space astrometric mission and it revolutionized our view of the Galaxy, despite the reduced number of observed stars. Gaia builds on the experience acquired with the Hipparcos mission; **it is an entirely European mission supported by a wide and motivated scientific and technological community distributed throughout Europe, which holds the leadership in the space astrometry domain.**

Gaia performs wide-angle astrometry, allowing the determination of absolute trigonometric parallaxes. The size of the primary mirrors and the fields of view and the use of CCD detectors in the focal plane enable an enormous jump in scientific performance compared to the Hipparcos mission. While Hipparcos yielded precisions of 1 mas at 9th magnitude and measured about 120000 stars, Gaia will yield 25 μ as at 15th mag and **will survey a billion objects**, aiming to reach completeness at $V_{\text{lim}} \sim 20\text{--}25$ mag (the exact limit depending on the colour of the object), providing accurate distances and motions — including the radial velocities for a large fraction of the objects — and detailed physical properties of each star observed. For full details on the Gaia performance, see http://www.rssd.esa.int/index.php?project=GAIA&page=Science_Performance.

1.1.2 Mission structure

The Gaia satellite and mission operations are fully funded by ESA. The Gaia management structure is summarized in Fig. 1.1 and includes a project manager and a project scientist. The project manager is in charge of supervising industrial development and ground segment management. The project scientist supervises the accomplishment of the scientific goals in the design phase and is assisted by a group of scientists external to ESA and representing the scientific community.

The processing of the acquired data and production of the “Gaia catalogue” is the responsibility of the Data Processing and Analysis Consortium (DPAC², see also [10]), comprising more than 430 members in 11 countries who are joining their efforts to overcome the challenging problem of managing the 1 Petabyte of Gaia data. DPAC activities are supported by national funding agencies that have signed a multilateral agreement with ESA to ensure the stability of the scientific teams for the necessary interval of time. Processing will be done on the premises of six Data Processing

¹The earliest launch date at the time of writing this proposal is in October 2013, but can still move some weeks back or forth depending on the Soyuz launch schedule

²See <http://www.rssd.esa.int/gaia/dpac>

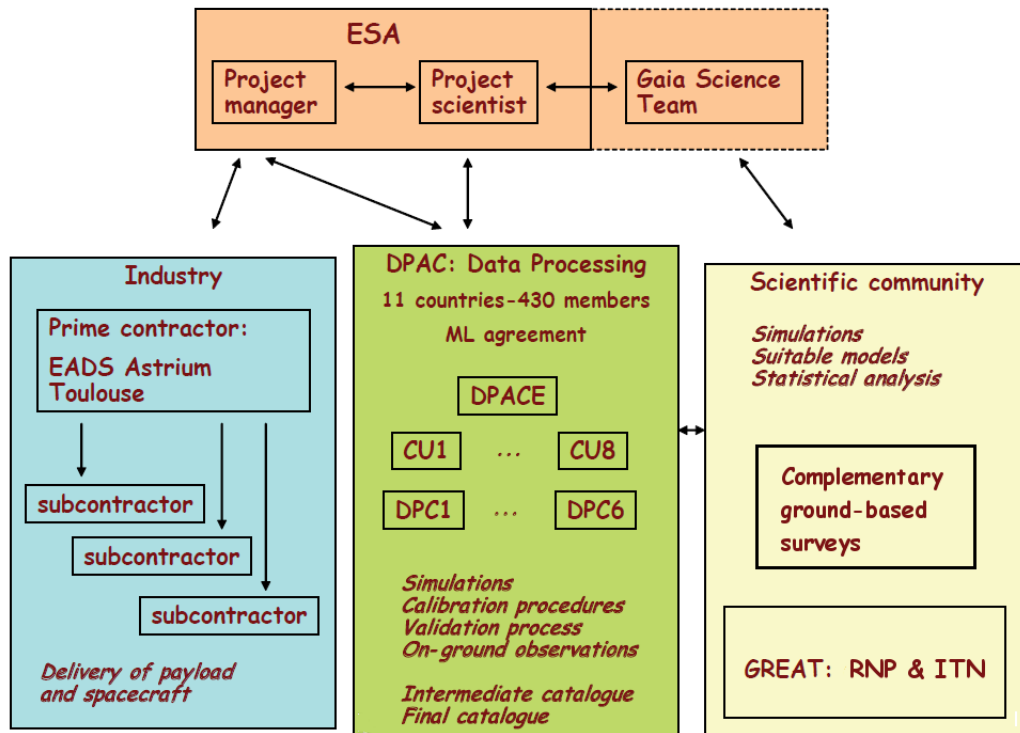


Figure 1.1: The present structure of the Gaia mission. GENIUS will provide the bridge between the DPAC and the scientific community.

Centres also made up of scientists and engineers funded by the agencies who have signed the multilateral agreement. As depicted in Fig. 1.2, the work within DPAC is presently organized into eight coordination units, comprising distinct fields of competence and having separate leaders.

Neither ESA nor DPAC has direct responsibility for the scientific exploitation of the Gaia data, which instead is to be conducted by the scientific community at large. It is up to this community to organize the best use of Gaia's products. Since the DPAC community has deep knowledge of the Gaia mission and extensive expertise in astrometry, synergy with the community at large is obvious and necessary for making use of the enormous legacy that Gaia will provide. To that end, several networks are currently in place, such as the GREAT (Gaia Research for European Astronomy Training) Research Networking Program (RNP), funded by the European Science Foundation, the GREAT Initial Training Network (ITN), funded by the European Commission, and the REG (Red Española de Gaia, Spanish Gaia Network), just to mention a couple of examples. These networks aim to strengthen the cooperation between teams, thereby joining expertise in observations, the Gaia mission, theoretical analysis, model development and interpretation, and statistical treatment. Complementary ground-based observations covering the areas of science alerts, infrared photometry, and high resolution spectroscopic surveys are being studied or are underway. **Proposals for new wide field multi-object spectrographs for 2- to 4-m class telescopes, which will be used to collect spectroscopy complementary to Gaia, are being pursued.**

Note that the participants in this proposal play a significant role in DPAC and the Gaia project in general, which strengthens the relevance of GENIUS for the scientific exploitation of the Gaia data:

- The GENIUS coordinator, X. Luri, is the manager of the Coordination Unit 2 (data simulations), as well as member of the executive committee of DPAC. He was also one of the

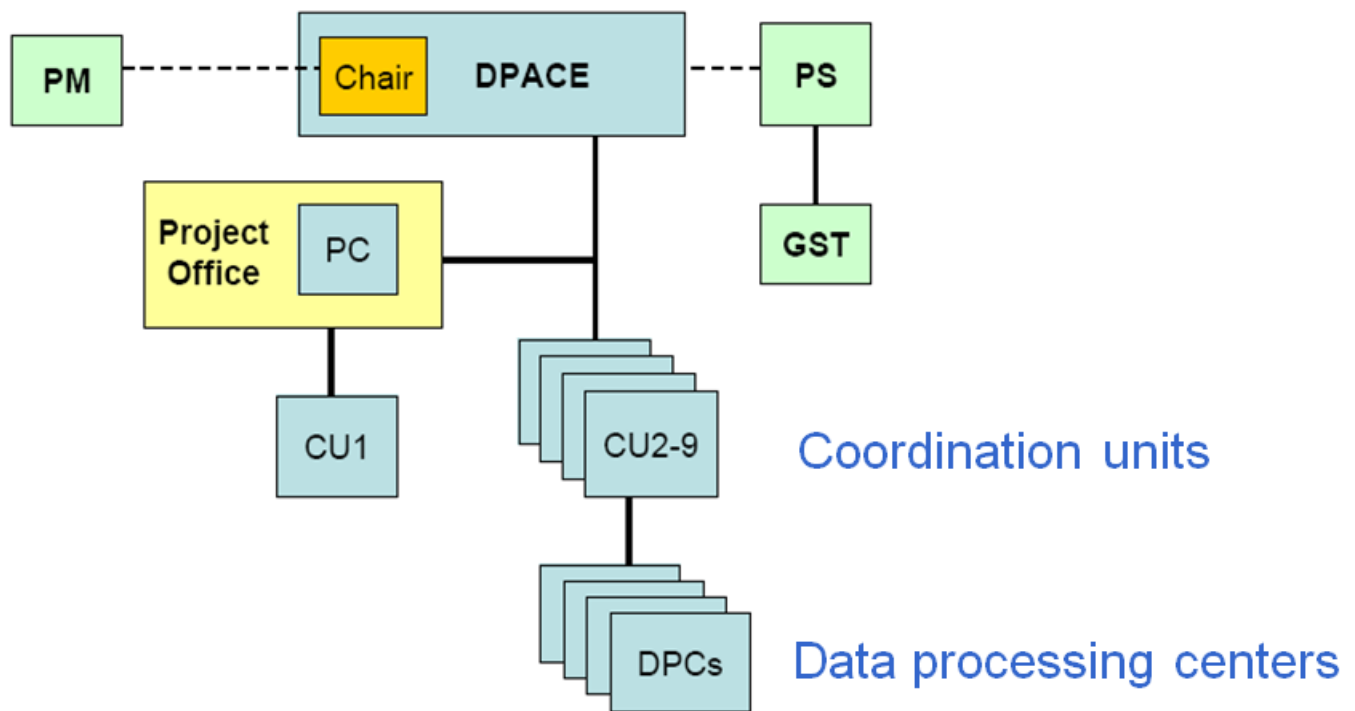


Figure 1.2: The structure of the DPAC. The DPAC Executive (DPACE) interfaces with both the Gaia Project Manager (PM) and Project Scientist (PS), who is advised by the Gaia Science Team (GST). The DPACE is supported by a project office headed by the Project Controller (PC).

authors of the “Gaia Concept and Technology Study Report” [16] outlining the mission’s scientific case on which Gaia was accepted within ESA’s scientific programme, and has been a member of the Gaia Science Advisory Group (1997–2000), the Gaia Data Analysis Coordination Committee (2005–2006), the Gaia Science Team (2001–2007) and the DPAC Executive Committee (2007–). He is also the leader of the answer of the European community to ESA’s Announcement of Opportunities for the creation of the Coordination Unit 9, that will have the responsibility to implement and operate the Gaia archive (see Section 1.1.3).

- A.G.A. Brown (WP200 manager) is chair of the DPAC Executive and member of the Coordination Unit 2 (data simulations) and the Coordination Unit 5 (Photometric Processing), where he is active in the Development Unit 11 (BP/RP flux extraction and initial data treatment). Brown is a member of the Gaia Science Team (2006 onwards) and has also served on the Gaia Data Access and Analysis System Steering Committee (2004–2005) and the Gaia Data Analysis Coordination Committee (2005–2006).
- N. Hambly (WP300 manager) is member of the Coordination Unit 3 (Core Processing), and of the Coordination Unit 5 (Photometric Processing), working on the initial data treatment, intermediate data update, and First-Look systems and calibration issues.
- F. Arenou (WP500 manager) is member of Steering Committee of the Coordination Unit 4 (Object processing), where he manages the Development Units 38 (test data) and 39 (multiple star combiner). Arenou coordinated two scientific Working Groups (On-board Detection

and Double and Multiple Stars) during the scientific preparation phase (2001–2005) and has been a member of the Gaia Science Team (2001–2005).

- N. Walton (UCAM-IoA) is member of the Gaia Science Team and coordinator of the GREAT ESF (European Science Foundation Research Networking Programme <http://www.great-esf.eu/>) and GREAT-ITN (ITN network programme EU FP7, <http://great.ast.cam.ac.uk/great-itn/>).
- Luri, Arenou, Brown, Hambly and Walton are all members of the Gaia Archive Preparation working group and work package managers in the answer of the European community to ESA's Announcement of Opportunities for the creation of the Coordination Unit 9 (see Section 1.1.3).

1.1.3 The Gaia archive & CU9

As described in the previous section the task of the processing of the Gaia data has been delegated by ESA to DPAC. This delegation explicitly excluded one of the tasks needed to make the data available to the scientific community: the definition, implementation and operation of the Gaia archive. The way ESA is handling this task is two-fold:

- Firstly, **the Gaia archive will be hosted at ESA's European Space Astronomy Centre (ESAC)**, where the agency will locate an engineering team to work on the design, deployment and operation of the archive.
- Secondly, ESA has issued an Announcement of Opportunity (similar to the one issued for the formation of DPAC, but more specific) for the participation of the scientific community in these tasks.

This Announcement of Opportunity will lead to the creation of a new Coordination Unit inside DPAC (named CU9) that will combine both aspects in a single structured cooperation for the development of the Gaia archive. **This proposal is oriented to fit into this cooperation, providing targeted resources for key enabling activities for the definition and implementation of the Gaia archive.**

ESA's AO was issued on 19 November 2012³ but the preparatory work for CU9 started well before. To advance on the definition of its role, requirements, and tasks, a working group was created (the Gaia Archive Preparation group, GAP) including members of ESA, DPAC and the wider scientific community. The group has already generated documentation that has been used as guidelines for this proposal, as described in the next section.

1.1.4 Collaboration with other space powers

ESA and DPAC have established a collaboration with the Japanese Nano-JASMINE mission⁴, a technical demonstrator for the JAXA/NAOJ astrometric mission JASMINE⁵. Currently this collaboration covers only the data reduction of Nano-JASMINE but GENIUS includes a Japanese partner (Kyoto University) to enhance this cooperation, aiming to allow the sharing of archive technologies and to make the Nano-JASMINE, JASMINE and Gaia archives fully interoperable or even incorporating all catalogues into a common archive, allowing its combined use for validation of

³<http://sci.esa.int/science-e/www/object/index.cfm?fobjectid=51116>

⁴<http://www.jasmine-galaxy.org/nano/nano-en.html>

⁵<http://www.jasmine-galaxy.org/index.html>

the respective catalogues, among other applications. **The funding for this Japanese partner is oriented to allow the Nano-JASMINE/JASMINE researchers to travel and participate in the GENIUS activities, ensuring the cooperation between the missions.**

A letter of interest from the Nano-JASMINE technical manager of data analysis software development and catalogue release and a letter of support from the Nano-JASMINE P.I. have been sent to the coordinator of GENIUS and are available on request.

1.1.5 Objectives

The objective of the GENIUS proposal 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 the future CU9.

According to the current planning the first releases of the Gaia catalogue, and therefore the first operational versions of the Gaia archive, will be around mid-2015, and the release of the final Gaia archive should take place around 2021. Therefore, with a planned duration of 3.5 years starting at end 2013, GENIUS is conceived to contribute from the start to the CU9.

The proposed contribution is based on the discussions and recommendations of GAP, and comprises a four-level set of objectives that is mapped onto the Work Package structure presented in Sec. 1.3.3:

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 will be 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 fulfill the needs of its users.

Furthermore, we intend to provide special attention to what we have called the *Grand challenges* (see WP 200 for more details), relevant scientific cases of use of the Gaia data that require an intensive and/or complex access to the archive.

Related milestones: 3, 5, 8, 9, 10, 11, 12, 16 (Sec. 1.3.3)

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. Our goal is to provide just such a contribution for archive systems development, as described in detail in WP 300.

Related milestones: 2, 3, 6, 8, 9, 12, 13, 14, 16 (Sec. 1.3.3)

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 WP 400, will thus cover three areas: visualization tools, data mining tools, Virtual Observatory tools and outreach tools.

Related milestones: 3, 6, 8, 10, 12, 16 (Sec. 1.3.3)

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 WP 200. In particular, this validation will require 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 WP 500 description.

Related milestones: 3, 6, 8, 11, 12, 14, 16, 17 (Sec. 1.3.3)

1.1.6 Compliance with the SPA.2013.2.1-01 objectives

We present here how the GENIUS objectives relate to the topics addressed by the call, specifically to:

Activity: 9.2. Strengthening the foundations of Space science and technology

Area 9.2.1: Research to support space science and exploration

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

1. *ESA has supported many science missions (see in particular data available at the European Space Astronomy Center (ESAC)), but data analysis beyond the execution of the mission remains limited. Collaborative proposals in the field of further data analysis and data exploitation are of particular importance.*

The Gaia archive will be hosted at ESAC and the ESA team already working on the design of the archive is hosted there. GENIUS aims to closely cooperate with them in the framework of the ESA/DPAC CU9, as described in previous sections, specifically aiming to develop tools allowing a better, deeper and more fruitful analysis and exploitation of the archive.

2. *A focus is to be given to research, analysis and presentation of data obtained from space missions, exploiting such space mission data in combination with data collected from ground based observations/data. [...] Proposals should clearly demonstrate how their proposed combination of data sets, from multiple instruments or mission sources, including combinations of space and non-space based data e.g. from terrestrial space observatories like European Southern Observatory (ESO) or the Canary's Islands observatories, leads to strong synergies, and adds value to the data obtained in space.*

Being focused on Gaia, which as described in Sec. 1.1.1 will make a substantial contribution to many areas of astronomy and astrophysics, GENIUS perfectly suits this goal of the present call. Several ground based observation programs are in preparation to complement the Gaia data, for instance the Gaia-ESO survey, a public spectroscopic survey that will observe more than 10^5 stars using FLAMES and UVES at the VLT: <http://www.eso.org/~rsmiljan/gaiaeso/Welcome.html>. A letter of support from G. Gilmore, co-PI of the Gaia-ESO Public Spectroscopic Survey, has been sent to the coordinator.

Furthermore, as we have already mentioned, we have included a Japanese partner to ensure the coordination and synergies with Nano-JASMINE and JASMINE, the only other space astrometry missions in preparation aside from Gaia, aiming to ensure the interoperability (or even integration) of the archives. A letter of interest from the Nano-JASMINE technical manager of data analysis software development and catalogue release and a letter of support from the Nano-JASMINE P.I. have been sent to the GENIUS coordinator.

3. *Research and analysis projects are intended to strengthen cooperation on scientific problems, which are for instance relevant to our solar system, internal constitution of stars and stellar evolution, exoplanets, galaxies and interstellar media.*

As described in Sec. 1.1.1 Gaia will contribute to all these areas and more, and the implication of the user community in the development of GENIUS will enhance its capability of exploiting the Gaia data.

4. *Projects should enhance the effectiveness and productivity of the European scientific community, and promote the contribution of space assets to scientific and technological knowledge, through:*

- (a) *mobilising the best expertise, in particular academic researchers and scientists, in various fields of science for the analysis and interpretation of space data, selecting the most innovative and challenging objectives in emerging scientific fields;*

The composition of the GENIUS project gathers a wide range of expertise relevant for the exploitation of the Gaia mission, and the involvement of the user community widens this range even more. Furthermore, the selection of the so-called *Grand challenges* corresponds to the request of selecting innovative and challenging objectives.

- (b) **supporting the preparation of data exploitation of missions under development, which might need special tools for data acquisition and processing, in light of the wealth and variety of data they will collect, including ground-based auxiliary data where relevant;**

GENIUS covers precisely these areas, specifically aiming to the developments of tools for access, analysis and validation of the Gaia data, as well as developing the archive and allowing an efficient access to it. Furthermore, as already mentioned, several ground based observation programs are in preparation to complement the Gaia data, for instance the aforementioned Gaia-ESO survey and other projects under development, such as the WEAVE, 4MOST, and MOONS multi-object spectroscopy surveys.

- (c) **developing comprehensive presentation and visualisation techniques, preferably in 3D, in order to better understand the interrelations between different sources (sensors) and modes (temporally, spatially)**

The development of visualization tools, including 3D, for a better and more comprehensive exploitation of the Gaia catalogue is included in WP400 (see Section 1.3.3). See also Section 1.2.4.

5. *This topic is open to international cooperation and should focus on downstream R&D activities complementing space missions, such as the effective scientific exploitation of existing data. Cooperation with international partners from third countries (ICPC) , or countries which have signed an agreement with the EU covering Science and Technology, as well as other space-faring nations (e.g. US, **Japan**) will help to expand the use of data, the corresponding data processing and management methods in third countries, and **enhance research partnerships with emerging or established space powers**. Therefore such international partners will be eligible to participate and to be funded and this aspect will be taken into account in the evaluation.*

The essence of GENIUS is precisely to enable the implementation of advanced features for the Gaia archive which will enhance the scientific exploitation and legacy value of the data produced by Gaia, an ESA space mission.

As described in 1.1.4, GENIUS includes a Japanese partner (Kyoto University) to enhance this cooperation aiming to allow the sharing of archive technologies and to make the Nano-JASMINE, JASMINE and Gaia archives fully interoperable or even incorporating all catalogues into a common archive. Letters of support from the mission managers have been sent to the GENIUS coordinator and are available on request.

6. *Projects are expected to contribute to the much needed coordination and exploitation of existing and future data collections from space missions, and thereby enhancing the possibility to base research on data sets providing comprehensive or full coverage, while at the same time addressing the potential need for further analysis of existing data sets. It is also expected that the projects will facilitate access to, and appropriate use of data for those scientists who were/are not part of the team having obtained the space mission data (e.g. principal investigators).*

GENIUS aims to contribute to the coordination and exploitation of the Gaia data, to open up the possibility of re-analyzing or re-processing (parts of) the Gaia data, and to ensure archive interoperability with existing and future archives. Furthermore, since the Gaia data will not have proprietary rights it will be open to the wide scientific community as soon as it is available, and therefore the GENIUS contributions will benefit all potential end users of the Gaia archive.

7. *Furthermore, projects are expected to add value to existing activities on European and national levels, and to raise the awareness of coordination and synergy efforts among stakeholders.*

As described in Sec. 1.1.2 the Gaia data processing is being carried out by the DPAC consortium, funded by national agencies, and the scientific community is organising itself for the exploitation of the Gaia data at European level (GREAT, including 90 groups from 17 countries) and national level (e.g. REG with 132 members from 22 Spanish Universities and Research Centres). GENIUS will clearly add value to these existing initiatives and will work in coordination with them, as highlighted by the fact its members are also part of these consortia and networks. Letters of support from DPAC and the REG have been sent to the GENIUS coordinator and are available on request.

1.2 Progress beyond the state-of-the-art ⁶

1.2.1 Archive interface

The basic interface of the archive to be developed at ESAC can follow the well-proven model of the existing archives, but we can also consider to adopt emerging tendencies like the use of natural language in the interface – for example, one could then imagine an interaction like “plot magnitude against colour for all stars with parallax between 1 and 1.5”, similar to the queries one can do using Wolfram Alpha⁷. In any case the users should have the last word in this in order to avoid providing an interface that they do not actually use. This will be one of the tasks in WP 200, gathering the user inputs to guide the archive development.

1.2.2 Virtual machines and light clients

A monolithic one-piece client-server archive design is typical of many existing astronomical archives, but there are other emerging models that are worth exploring if we are to provide flexible and scalable research solutions to our scientists.

The latest generation of research applications are increasingly employing a Web2.0 user interface. This allows an ‘anywhere, any-time’ work-flow as users only need access to a web browser to undertake (or indeed, resume) their work. Web2.0 functionality has improved to the point where the boundaries between local and remote applications are very blurred. Users no longer need to worry about where an application exists and how it is installed; rather they focus on how it is used and how to get the most from it. Examples of this can be seen emerging in many fields, astronomy-specific examples can be found at CDS Strasbourg, CADC Canada, and WFAU Edinburgh. In addition, the proliferation of complex mobile devices based on Android and iOS opens new opportunities for application developers to exploit new functionality in innovative ways provided a suitable infrastructure exists (see below).

As archives increase in size and complexity the need for a flexible infrastructure to contain them grows ever more urgent. Virtualisation of data storage and processing is an emerging solution that provides just this flexibility. Virtualised infrastructures are scalable, resilient to point failures and support a distributed architecture. Whether the data is stored in traditional databases or in the newer “NoSQL” solutions, virtualisation and distribution of the infrastructure is proving itself the optimal solution to increasingly demanding use. Investigations by the CADC (in the CANFAR project⁸) and others have confirmed that the latest VM technologies allow VMs to run at near bare-metal speeds once the VM has been instantiated thus removing the only serious objection to VMs, namely the overhead in running them.

Sitting between archives and applications are the Virtual Observatory protocols and standards. These provide an internationally-agreed abstraction layer between the two that greatly simplifies the development and deployment of data and applications. Standards exist for accessing tabular data, images, spectra, time-series etc. as well as intermediate infrastructure requirements such as data and application discovery, inter-application communications and localised data storage. Virtualisation also has a role to play here as it is possible to provide users with complex, insulated environments in which to undertake their research. By providing users with their own Virtual Machine for them to configure and use as they wish — and by running those VMs in a cloud environment — all the benefits of scale and flexibility are obtained whilst still ensuring ‘rogue’

⁶this section has strongly benefited from the discussions in the above described Gaia Archive Preparation group and the documents produced therein. A good recollection of these ideas can be found in [12].

⁷<http://www.wolframalpha.com/>

⁸<http://canfar.phys.uvic.ca/>

experiments are contained within their own ecosystem. A good example of this is CANFAR, and this is something all future archives would do well to explore.

These and other potential possibilities for the future of astronomical archives will be explored in WP 200 and WP 300.

1.2.3 Advanced statistics

Many astronomers today research small portions of the sky and most archives are well capable of answering queries for such matters. The spatial nature of astronomy means that this type of query will always be required, and certainly the Gaia archive will include them, but we are in the era of mega surveys producing huge amounts of data across large sky areas (or the entire sky like Planck and Gaia). Nowadays an astronomer interested in a particular object may want to search for similar objects in several surveys. In other words, interoperability is a mandatory feature and has therefore been included among the objectives of GENIUS as described above.

Furthermore, with large homogeneous data sets, statistics becomes important; providing them for very large data sets is hard and needs to be automated to be useful: existing tools like the Virtual Observatory do not yet provide anything capable of doing this⁹. If we take as an analogy that most current astronomers are digging with shovels, some are realizing they need a bulldozer but no one has built it yet. Precomputing some statistics and allowing the user to query those to reduce the size of the data set of interest is a starting point. Ultimately the user should be able to use data mining tools like cluster analysis, pattern recognition and N-point correlation function calculations in any subset of the data. In other words, data mining should be an integral component of the archive, as described in WP 400.

1.2.4 Multidimensionality and visualization

We have referred above to the Gaia catalogue as being 3D, the spatial nature of an astrometric catalogue reminding one of its 3D distribution. However, the dimensionality of the Gaia data is in fact much higher — one may think of any catalogue as a multi-dimensional space of all attributes in the the catalogue where each attribute represents a dimension.

If we could visualise this multidimensional space of parameters, similar objects would appear close to each other, but dealing with more than three dimensions (or four if one counts time) is not intuitive. In many cases one gets representations of three dimensional data in two dimensions, and nowadays 3D devices are starting to become popular. It is probable with advances in 3D technology that people will arrive at ways of explaining a few more dimensions in data sets, and such an advance could result in more requests for representing and sorting of data in higher dimensional multi spaces and to represent this in 3D.

In any case it is clear that the availability of powerful and tailored visualization tools for the Gaia archive will contribute to its scientific exploitation. For this reason the development of such (presently not available) tools has been included in WP 400.

1.2.5 Outreach and simplicity

Whether we like it or not, Europe is far behind the USA regarding the outreach of astronomy or space-related activities. One way to improve the situation in the Gaia case would be to build a community portal including outreach oriented facilities, a query interface simple enough for school children to use at the outset and with more advanced features available for secondary to

⁹The complexity of Gaia even in the Virtual Observatory era had been anticipated a decade ago, see [13]

university students and amateur astronomers. Ideally, of course, the outreach user should be able to customize the system for personal taste by choosing a range of applications or widgets which they wish to use regularly. Also simpler interfaces run better on mobile devices, which are becoming a ubiquitous access point to knowledge. In addition there should be lessons and learning activities centred around the archive – for instance the SDSS has several good examples built into the same website the Sloan astronomers use. On this note for proper connection to people the system must be properly multi-lingual from the ground up. The possibilities for the implementation of these ideas will be explored in WP 400.

1.2.6 Interoperability and reprocessing

Astronomical catalogues have moved on-line from paper relatively recently, the first examples just allowing searching. However, the nature of astrophysical research means that the data is constantly being reinterpreted using new data, and moreover, reprocessed using better models and tools (for instance, ESO has been offering on the fly reprocessing of image data for some years now) and therefore the classical static nature of the data is changing.

This trend implies that interoperability (the need to cross-match data of several archives and sources) and reprocessing will become more and more relevant, and has been thus included in WP 200 and WP-500. Furthermore, the large size and richness of the Gaia archive implies that the final reduction scheduled for 2021 may not be the last word on its contents. One can easily imagine that an improved treatment of the astrometric data of binary systems or the light curves of variables is developed, potentially improving the orbital parameters of the former and the periods of the latter for instance. Including the facilities for reprocessing in the design of the Gaia archive will not only improve its quality but also much extend its useful lifetime.

The possibility of reprocessing is also tied to another of the objectives of GENIUS, the quality control and validation of the data (WP 500), that in the case of reprocessing will be even more crucial.

1.2.7 Virtual observatory

Astronomy has traditionally been at the forefront of the development of on-line services. National and international ground- and space-based observatories produce terabytes of data which are publicly available all around the world from data centres. This information is typically complemented with added-value services such as high-level data products or analysis tools. Results published in electronic journals are also available on-line. The already existing archival information, together with the all-sky surveys foreseen in the coming years, will produce coverage of large areas of the sky in a wide wavelength range.

Although this situation will clearly benefit existing multiwavelength astronomy, an important limiting factor is the lack of standardization (different access and retrieval protocols, data models, formats, policies, . . .) among the astronomical archives and services. This makes it very inefficient to identify, compile and transfer the data of interest from more than one resource. Also, added to this lack of interoperability is the management of the volume of data generated by the all-sky surveys, which will reach the Petabyte level in the coming years.

These are the challenges of the Virtual Observatory (VO), an international initiative whose main goal is to guarantee an easy and efficient access and analysis of the information hosted in astronomical archives and services. The VO concept goes one step further than just giving access to distributed computational resources or to the data. It also permits operations on the data and returns results. VO is opening new lines in the astronomical research and it is considered both

from the technical and scientific point of view the framework where to settle the astronomical archive-related activities.

The importance of the Virtual Observatory as a key research infrastructure has also been recognized at European level by ASTRONET, an ERA-NET created by a group of European funding agencies and funded by the European Commission with the objective of consolidating and reinforcing the world-leading position that European astronomy has attained at the beginning of this 21st century. The ASTRONET Roadmap Executive Summary states that:

“Systematic archiving of properly calibrated observational data in standardised, internationally recognised formats will preserve this precious information obtained with public funds for future use by other researchers, creating a Virtual Observatory. The Virtual Observatory will enable new kinds of multi-wavelength science and presents new challenges to the way that results of theoretical models are presented and compared to real data.”

This is a remarkable result keeping in mind that the Virtual Observatory is a relatively young concept, since it appeared around 2000, and illustrates that the astronomical scientific community is conscious of the change in paradigm that is occurring in the way research is done with the emergence of the VO.

Two important objectives of GENIUS are to ensure that the Gaia archive provides the infrastructure for VO applications (WP-300) and also the tools to make use of VO facilities (WP-400).

1.3 S/T methodology and associated work plan

1.3.1 Overall strategy

The overall philosophy for the work plan is to have the requirements analysis and specification progress in parallel with the implementation thereof. We made this choice because of the complexity of the features for the archive to be implemented, especially the *Grand Challenges*, and because of the relatively short duration of the project. The work flow we foresee for the implementation of a given feature set of the Gaia data archive is shown in Fig. 1.3.

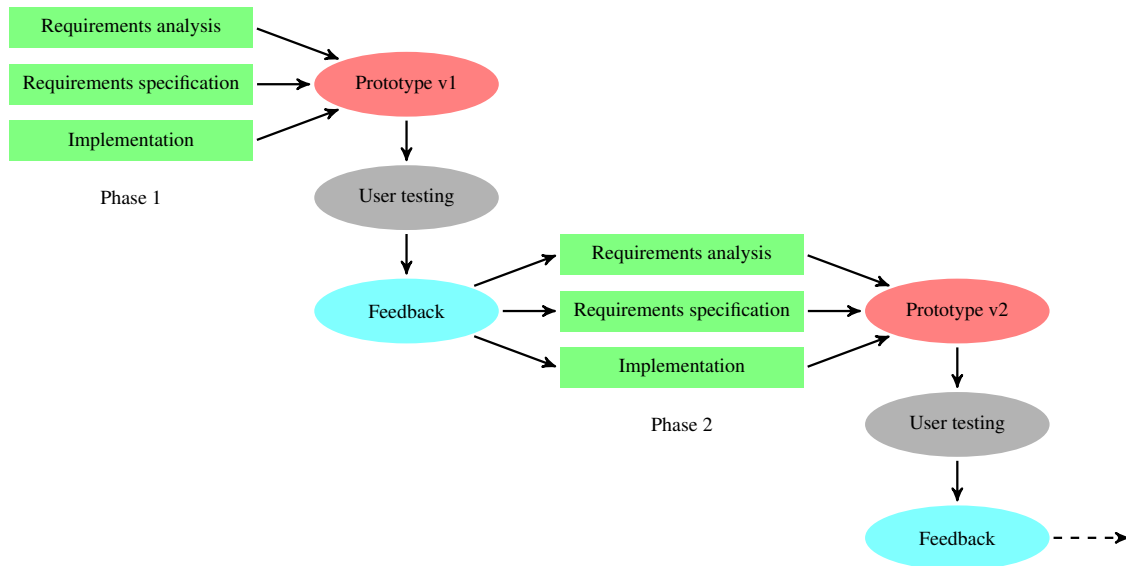


Figure 1.3: Iterative work flow for GENIUS.

We expect that two to three parallel phases are needed in order to arrive at a satisfactory understanding and implementation of a given data archive feature. Doing this with a waterfall process

would be impossible to fit within the duration of the project. The foreseen work flow is akin to the cyclic development approach used by DPAC and also adopted for CU9, into which GENIUS will integrate. Close coordination between the work packages will be essential. As show in Fig. 1.3 at certain points during the project the requirements and the corresponding implementation will be ‘frozen’ and then testing will take place in order to verify that we have implemented what was specified and, more importantly, to verify that the end user wishes are satisfied. This last point should be taken to mean that even if an implementation is exactly as specified it may still not be what the user had in mind when using the archive in practice. This is a natural consequence of the vagaries of specifying scientifically motivated requirements. The cyclic approach is very well suited to dealing with this problem, and will be coordinated with the cyclic approach also adopted for the overall CU9 development.

1.3.2 Timing

The GENIUS duration is **42 months**, divided into two stages:

Preparatory stage: the first six months will be 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.

Development stage: the remaining 36 months will be devoted to development, with the personnel in place and the coordination with DPAC/CU9 fully established. It is during this phase that the iterative approach show in Fig. 1.3 will be used.

ID	Task Name	1st Half	2nd Half	1st Half	2nd Half	1st Half	2nd Half	1st Half	2nd Half	1st Half
1	Gaia Launch									
2	Provisional date first Gaia data release (conservative estimate)									
3	GENIUS									
4	WP100 Management of GENIUS									
5	WP100 Management									
6	Kick-off meeting									
7	Advertise positions									
8	Hiring of main developers									
9	Archive user requirements document									
10	Requirements document for each subsystem									
11	Mid term review									
12	User prototype archive review									
13	Exploitation tools review									
14	Validation tools review									
15	Prototype archive open to community									
16	Stress test									
17	Load of actual Gaia data									
18	Completion meeting and final external review									
19	GENIUS products availability									
20	WP200 Tailoring to the end user community									
21	WP210 Technical coordination									
22	WP220 Analysis and working out of requirements gathered by GAP									
23	WP230 Confronting complex models with complex catalogues									
24	WP240 Seamless data retrieval across archives and wavelength domains									
25	WP250 The living archive									
26	WP260 Re-processing of data									
27	WP300 Aspects of archive system design									
28	WP310 Technical coordination									
29	WP320 Aspects of Archive Interface design									
30	WP330 VO infrastructure									
31	WP340 Data centre collaboration									
32	WP350 Cloud-based research and data mining environments									
33	WP400 Tools for data exploration									
34	WP410 Technical coordination									
35	WP420 Visualization tools									
36	WP430 Data mining									
37	WP440 VO tools and services									
38	WP450 Tools for outreach and academic activities									
39	WP500 Tools for data validation and analysis									
40	WP510 Technical coordination									
41	WP520 Looking for trouble: definition of problem cases, validation scenarios and tools									
42	WP530 Simulation vs reality: from models to observables									
43	WP540 Confronting Gaia to external archives									
44	WP550 Data demining: outlier analysis									
45	WP560 Transversal tools for special objects									
46	WP600 Support activities									
47	WP610 Technical coordination									
48	WP620 Simulated catalogue data									
49	WP630 Science alerts tested									
50	WP640 Community portal infrastructure									
51	Agreed testbed environment with ESAC-SAT and CU9									
52	GENIUS portal available at GESCA for internal use									
53	Public version of GENIUS portal									
54	Handover of GENIUS portal to CU9									

1.3.3 Work description

Table 1.3a - Work package list

WP no.	WP title	Activity type	Lead part. no.	Lead part. short name	Person months	Start month	End Month
100	Management	MNG	1	UB	32	1	42
200	Tailoring to the end user community	RTD	4	UL	71	1	42
300	Aspects of archive system design	RTD	3	UEDIN	66.8	1	42
400	Tools for data exploitation	RTD	1	UB	87.8	1	42
500	Tools for data validation and analysis	RTD	2	CNRS	96.1	1	42
600	Support activities	RTD	1	UB	42	1	42
Total					395.70		

Table 1.3 b - Deliverables list

Del. no.	Del. name	WP no.	Nature	Dissemination level	Delivery date
100.1	Kick-off meeting (plenary)	100	O	PU	1
100.2	Partnership agreement	100	O	PP	2
100.3	Semestral report 1	100	R	PP	6
100.4	Semestral report 2	100	R	PP	12
100.5	Semestral report 3	100	R	PP	18
100.6	Mid-term meeting	100	O	PU	21
100.7	Semestral report 4	100	R	PP	24
100.8	Semestral report 5	100	R	PP	30
100.9	Semestral report 6	100	R	PP	36
100.10	Completion meeting	100	O	PU	41
100.11	Final report to the EC	100	R	PP	42
200.1	Requirements specification for catalogue and data archive	220	R	PU	12
200.2	Requirements specification for outreach facilities built into the archive system	220	R	PU	12
200.3	Requirements specification for generic projection module	230	R	PU	18
200.4	Requirements specification for data retrieval across archives	240	R	PU	18
200.5	Conclusion of requirements update gathering exercise	220	R	PU	24
200.6	Requirements specification for incorporating new information into an existing archive	250	R	PU	30

Continued on next page

Del. no.	Del. name	WP no.	Nature	Dissemination level	Delivery date
200.7	Requirements specification for the archiving of raw and intermediate data	260	R	PU	36
200.8	Requirements specification for the archiving of the original software with which the archive was produced	260	R	PU	36
200.9	Requirements specification for model comparison and optimization tools	230	R	PU	40
300.1	GENIUS/ESAC–SAT Co-ordination and Interface Control document	310	O	PU	3
300.2	Web2.0 user interface demonstration prototype deployment	320	R	PU	18
300.3	Gaia Data IVOA compliance document	330	R	PU	42
300.4	Deployed web services, code and documentation	330	R	PU	42
300.5	Data centre report and analysis document	340	R	PU	42
300.6	TAP+ code and documentation	340	R	PU	42
300.7	Deployed CANFAR–style VM research environment and produced reports and documentation	350	R	PU	42
400.1	Requirement specification document for the exploitation tools	420/ 430/ 440	R	PU	6
400.2	Opening of community portal	450	O	PU	6
400.3	Delivery of first prototype of exploitation tools	420/ 430/ 440	P	PP	12
400.4	Delivery of second prototype of exploitation tools	420/ 430/ 440	P	PP	24
400.5	Delivery of third prototype of exploitation tools	420/ 430/ 440	P	PP	36
400.6	Delivery of exploitation tools user manuals and technical documentation	400	R/O	PU	36
400.7	Deployment of exploitation tools on the first actual Gaia archive	400	O	PU	42
400.8	Handover of the community portal to the Gaia community	450	O	PU	42
500.1	Delivery of prototype of internal checking tools	520	P	PP	12
500.2	Delivery of prototype of statistical tools	550	P	PP	18
500.3	Delivery of internal consistency checking tools	520	O	PP	24
500.4	Delivery of statistical tools	550	O	PP	24
500.5	Delivery of model-based validation tools	530	O	PP	24

Continued on next page

Del. no.	Del. name	WP no.	Nature	Dissemination level	Delivery date
500.6	Delivery of prototype of external validation tools	540	P	PP	24
500.7	Delivery of external validation tools	540	O	PP	36
500.8	Delivery of special objects tools	560	O	PP	36
500.8	Deployment of validation tools on the Gaia archive	500	O	PU	42
600.1	Basic setup of the community portal internally available for working	640	O	PP	3
600.2	First public version of the community portal	640	O/R	PU	6
600.3	Delivery of first simulated catalogue data	620	O/R	PU	6
600.4	Deployment of first public science alerts prototype	630	P	PU	12
600.5	Delivery of second simulated catalogue data	620	O/R	PU	18
600.6	Deployment of second public science alerts prototype	630	P	PU	24
600.7	Upgraded public version of the community portal	640	O/R	PU	24
600.6	Delivery of third simulated catalogue data	620	O/R	PU	30
600.7	Deployment of third public science alerts prototype	630	P	PU	36
600.8	Final (complete) version of the community portal. Handover to the Gaia community	640	O	PU	42

Table 1.3c - List of milestones

Milestone no.	Milestone name	WP involved	Expected date	Means of verification
1	kick-off meeting	100	1	Organised by GENIUS board
2	Agreed testbed environment with ESAC-SAT & CU9	300	4	GENIUS board should be notified
3	Hiring of main developers	200, 300, 400, 500, 600	4	GENIUS board should be notified
4	GENIUS portal available at CESCA for internal use	600	6	GENIUS board should be notified
5	Archive user requirements document	200	12	Document approved with External Advisory Board & CU9
6	Requirements document for each subsystem	300, 400, 500	12	Documents approved with External Advisory Board & CU9
7	Public version of GENIUS portal	600	12	GENIUS board should be notified
8	Mid term review	All	20	Positive report from external board
9	User prototype archive review	200, 300	24	Test and report by selected users
10	Exploitation tools review	200, 400	24	Test and report by selected users
11	Validation tools review	200, 500	24	Report by DPAC & CU9
12	Prototype archive tools open to community	All	24	Verified by GENIUS board and external board
13	Stress test	300	35	Verified by GENIUS board and external board
14	Load of actual Gaia data	300, 500	38	Verified by GENIUS board & CU9
15	Completion meeting & final external review	100	41	Organised by GENIUS board
16	GENIUS products availability	All	42	Verified by GENIUS board & CU9
17	Handover of GENIUS portal to the Gaia community	600	42	GENIUS board should be notified

Table 1.3d I - WP100

WP number: 100		Start: Kick-off meeting				
WP title:	Management					
Activity type:	MGT					
Participant number:	1	2	3	4	12	
Participant short name:	UB	CNRS	UEDIN	UL	UCAM	
Person-months per participant:	24	2	2	2	2	
Objectives: This package provides the overall administrative management of GENIUS, as described in Sec. 2.1.						

Description of work:

This work package includes the administrative tasks to fulfil the EC requirements and rules as well as the global administrative and coordination tasks inside the consortium, including financial management, intellectual property management and project documentation. It also includes the coordination with institutions and bodies relevant for the development of the Gaia archive like ESA, DPAC and GREAT, as well as the representation of GENIUS in meetings or committees related to this coordination. These tasks will be carried out by:

- The GENIUS coordinator, assisted by a hired project manager (to be devoted part time to the GENIUS tasks) (6 +18 person-months)
- The coordinators of the four main partners through the GENIUS executive board (2 person months per partner). In addition the coordinator of the UCAM (IoA) node will provide the interface to the wider GREAT community.

Deliverables

The deliverables of this work package are the final report for the EC, the partnership agreement (Sec. 2.1.3), the semestral reports for the external advisory board (Sec. 2.1.1) and the organisation of the GENIUS plenary meetings (Sec. 2.1.2).

Del. Num.	Description	Month
100.1	Kick-off meeting (plenary)	1
100.2	Partnership agreement	2
100.3	Semestral report 1	6
100.4	Semestral report 2	12
100.5	Semestral report 3	18
100.6	Mid-term meeting (plenary)	21
100.7	Semestral report 4	24
100.8	Semestral report 5	30
100.9	Semestral report 6	36
100.10	Completion meeting (plenary)	41
100.11	Final report to the EC	42

Table 1.3d II - WP200

WP number: 200		Start date: Kick-off meeting				
WP title:	Tailoring to the end user community					
Activity type:	RTD					
Participant number:	4	6	10	12	13	
Participant short name:	UL	INAF	FFCUL	UCAM	KU	
Person-months per participant:	47	16	2	4	2	
Objectives:						
<p>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.</p> <p>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.</p> <p>The GAP requirements gathering process has revealed a number of advanced requirements (the <i>Grand Challenges</i>) 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 <i>Grand Challenges</i> will be researched in this work package:</p> <ul style="list-style-type: none"> • Confronting complex models with complex data archives (WP-230) • Seamless data retrieval across archives and wavelength domains (WP-240) • The living archive (WP-250) • Re-processing of archived (raw) data (WP-260) 						

Description of work:

WP-210: Technical coordination (2 UL staff months): This work package oversees the work conducted withing WP-200. It includes progress tracking and reporting, ensuring that deliverables are ready on time, and taking action in case of delays in the work. The latter action consists of re-assessing the priorities of the efforts spent on the different work packages if needed. The efforts in this work package will feed into developments in the other GENIUS work packages so

coordination with the respective work package leaders is also part of this WP.

The technical coordination of WP200 will be done by Brown at UL.

WP-220: Analysis and working out of requirements gathered by GAP: (9 UL +2 FFCUL +4 UCAM +2 KU staff months) Under the auspices of the GREAT network, GAP, and the Gaia science team, the astronomical community was given the opportunity to specify how they might wish to access the Gaia catalogue and data archive. This was done through usage examples in order to get an overview of what the future archive users may want. These data access scenarios¹⁰ need to be turned into precise specifications for the data archive which will serve as input to the activities in the WPs 300/400/500. This task will be undertaken in this work package. As mentioned above the examples provided by the community also revealed a number of advanced usage scenarios requiring a complicated interaction with a substantial fraction of the entire data archive. These will be addressed specifically by WP-230–260.

In addition to satisfying science user requirements the archive should also be ready to support outreach activities. So part of the work in this WP is to analyze outreach cases and formulate requirements for building outreach facilities into the Gaia archive.

This package will be carried out by the personnel hired at UL. The group at FFCUL will contribute 2 staff months of effort to provide their expertise for the analysis of the user requirements related to visualization aspects. The group at KU will contribute 2 staff months to a collaborative effort of conducting a requirements gathering and analysis exercise in the context of the Japanese Nano-JASMINE mission. In particular the requirements on providing a combined Nano-JASMINE/Hipparcos catalogue (improved proper motions) will be investigated. The KU group will benefit from the GAP experience and in turn we expect that lessons learned from the requirements analysis for Nano-JASMINE can also be applied to the Gaia case. The UCAM group will devote 2 staff months to the organisation of the update of the requirements from the GREAT community. In addition the UCAM will contribute 2 staff months to the analysis of user requirements specific for ‘scienc alerts’.

WP-230: Confronting complex models with complex catalogues: (12 UL staff months) Modern astronomical surveys offer the possibility of testing our understanding of the universe against vast data sets collected over the entire sky. In particular the Gaia catalogue will be highly constraining for models of the Milky Way or of the properties of stars. The models must explain the data collected across all stellar populations over a large fraction of the volume of our Galaxy. Testing stellar evolution models against single clusters or Galaxy models against star counts along a single line of sight will no longer be sufficient. These tests will have to be made against the entire catalogue in order to extract the maximum scientific return. Such an undertaking is very difficult because of the large amount of data involved, the large range in observational errors (due to the survey depth), the correlations between errors on the different quantities and between sources, and the often non-linear relation between the measured quantities and the natural model parameters (for instance parallax is measured rather than distance). It has therefore been argued over the recent years (see e.g., [2, 7]) that the only truly robust way to deal with this challenge is to project models into the data space (i.e., use ‘forward modelling’) and thus predict the catalogue data. A good model will thus provide the correct ‘predicted catalogue’.

To facilitate (and encourage) such a forward modelling approach we want to provide the corresponding tools on the data archive side. The following concepts will be worked out and turned into detailed requirements:

- Provide tools to project models into the catalogue’s data space. For example, turn a Galaxy

¹⁰The requirements gathering process and the collected data access scenarios are summarized in [3], available online at <http://www.rssd.esa.int/llink/livelinek/open/3125400>.

model into predicted astrometry, radial velocities, stellar population properties (ages, metallicities), or turn synthetic spectra from stellar models into predicted photometric measurements. The tools should encapsulate our knowledge of the instruments that produced the catalogue. This effort can build on the substantial instrument modelling expertise built up within Coordination Unit 2 of DPAC.

- Provide tools for comparing the predicted and the observed catalogue or data. The comparison will likely be done in a Bayesian framework so the following could be foreseen: a likelihood generator that is aware of the catalogue's error properties, including correlations; tools for specifying priors; posterior likelihood optimizers. Users should also be able to contribute their own optimization tools.

The forward modelling facilities will also be very valuable in the context of the data validation approach taken in WP-530.

This package will be carried out by the personnel hired at UL.

WP-240: Seamless data retrieval across archives and wavelength domains (16 INAF staff months): Although the Gaia catalogue on its own will be a very powerful tool, it is the combination of this high accuracy archive (especially the astrometry) with other archives that will truly open up amazing possibilities for astronomical research. An example application would be to query the Gaia catalogue for sources brighter and fainter than the survey limit of Gaia, where behind the scenes the work is done to combine Gaia and other sky surveys. In this way our reach across the Galaxy can be extended by combining the greater depth of surveys like LSST, Pan-Strass, SDSS, and EUCLID, with very accurately calibrated photometric distance indicators. The latter will be one of the Gaia results. Another example is the combination of accurate stellar distances, and extinction measurements with data on the gas and dust in the Milky Way's interstellar medium in order to build up a 3D picture of the ISM. In addition data on the velocity of the gas will enable us to constrain the gravitational potential in which the gas moves, and through combination with the stellar phase space data much more tightly constrain the Galaxy's mass density.

Many other examples can be provided but the point here is that the advanced inter-operation of archives does not simply mean 'cross-matching' but providing truly seamless data retrieval, leaving the user with the feeling of working with one single data archive. The data retrieval should work not only across data archives but also across wavelength domains as illustrated with the ISM example above. This WP can possibly build on developments that have already taken place in the context of the Virtual Observatory and the resulting requirements will feed into WP-330 and WP-440 and will also benefit the efforts planned for WP-540.

This package will be carried out by the person hired by INAF. The management of this WP and the coordination between INAF and UL will be done by Smart of INAF-OATo (2 staff months), while Spagna, also at INAF-OATo, will contribute his expertise on cross-matching (2 staff months).

WP-250: The living archive (12 UL staff months): A concept closely related to the previous item is that of making the Gaia data archive a 'living entity'. By this we mean that it should be possible to incorporate new information into the archive. Examples are complementary ground-based spectroscopy, updated classifications or parametrizations of stars based on independent information, better distance estimates for faint stars (e.g., photometric distance indicators calibrated on stars with accurate parallaxes), etc. The seamless integration with archives from other large sky surveys forms a natural part of the living archive idea.

The questions to investigate here are: how do we incorporate new information into the Gaia archive in a controlled manner? This means vetting of the new information, tracing the history of the information related to a source as well as the history of source classifications and parametrizations, and making the new information available in a transparent manner.

This package will be carried out by the personnel hired at UL.

WP-260: Re-processing of archived (raw) data (12 UL staff months): The Hipparcos Catalogue publication included the so-called intermediate astrometric data. The intermediate data are residuals of the observables with respect to the primary astrometric solution and the derivatives of these observables with respect to the astrometric parameters. These data allow users to re-process the Hipparcos astrometric data, notably to improve the astrometry of binaries and very red giant stars. Re-processing of already published data is gaining increasing popularity (as illustrated by the reprocessing of SDSS multi-epoch data described in [9]) and allows for much extending the scientific value and lifetime of existing data archives. Examples of re-processing that could be foreseen for the Gaia data archive are: the re-processing of intermediate data for groups of stars in order to derive a common radial velocity or parallax, the re-processing of data for objects that are discovered or confirmed to be binaries following a data release, or the re-determination of astrophysical parameters for stars following future improvements in stellar atmosphere modelling.

On a more ambitious level the study described in [18] built on improved insights into the attitude modelling for the Hipparcos spacecraft to perform a re-processing of the *entire* Hipparcos data set. The resulting new version of the Hipparcos catalogue features very much reduced error correlations and improved astrometric accuracies (by up to a factor of 4) for the bright stars. In principle also for Gaia the re-processing of *all* the raw data might be warranted at some point in the future.

The research questions underpinning the requirements specification in this case are:

- How do we archive the raw and intermediate data products for long term usability? This includes calibration methods and their parameters as well as the original processing software.
- How do we present, communicate, and facilitate the use of intermediate data or raw data?

This package will be carried out by the personnel hired at UL.

Deliverables

Del. Num.	Description	Month
200.1	Requirements specification for catalogue and data archive	12
200.2	Requirements specification for outreach facilities built into the archive system	12
200.3	Requirements specification for generic projection module.	18
200.4	Requirements specification for data retrieval across archives.	18
200.5	Conclusion of requirements update gathering exercise.	24
200.6	Requirements specification for incorporating new information into an existing archive.	30
200.7	Requirements specification for the archiving of raw and intermediate data.	36
200.8	Requirements specification for the archiving of the original software with which the archive was produced.	36
200.9	Requirements specification for model comparison and optimization tools.	40

Table 1.3d III - WP300

WP number: 300		Start date: Kick-off meeting				
WP title:	Aspects of archive system design					
Activity type:	RTD					
Participant number:	3	7	2	6		
Participant short name:	UEDIN	CSIC	CNRS	INAF		
Person-months per participant:	41	6	1.8	18		
Objectives:						
<p>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 made will be motivated by the real user requirements identified by WP 200 – in particular, the massive, complex queries defined by the Grand Challenges – and by other initiatives, such as the GREAT project, and will be made with full recognition of the constraints imposed by the ESAC archive system, with which it must interface effectively. Prototypes will be prepared and tested in cooperation with the end user community and with the ESAC science archive team through the DPAC CU9. A core principle will be 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.</p>						

Description of work:

WP-310: Technical coordination (4 UEDIN staff months):

In addition to managing the resources deployed on the other WP-300 work packages, and producing reports on those activities, this work package oversees the design and specification of all work conducted under WP-300, to ensure that it adequately addresses the requirements identified within the GENIUS project and from external sources, such as the CU9 and GREAT. The key thing here is to ensure maximum science return by enabling science exploitation through appropriate use of information technologies.

This WP also includes the assurance of compliance with the deployment of the archive at ESAC. Since the Gaia archive will be designed and run at in this centre, it is essential that the techniques and technologies prototyped in this project are consistent with what can be ultimately implemented there. An important aspect of WP-310 is to ensure the injection of the relevant requirements for this in the design and evaluation phases, and that all GENIUS system design work is tackled with full awareness of the constraints imposed by ESAC infrastructure and practice. A key deliverable is therefore a formal, documented co-ordination and interface agreement between GENIUS and the Science Archive Team (SAT) at ESAC through the CU9. This work will be undertaken by Hambly of UEDIN.

WP-320: Aspects of archive interface design (12 UEDIN staff months):

The Gaia mission will produce a wide variety of data products, leading to a complex archive. A crucial issue for the exploitability of the Gaia data set is, therefore, an archive interface that supports a sufficiently rich range of functionality and is sufficiently easy to use for users to do their science with it effectively. The task of this WP is to prototype archive interface components that meet these user requirements, as developed by the CU9 and GREAT. Since any candidate archive DBMSs to be employed at ESAC support access from Java via Java Database Connectivity (JDBC), it is possible to develop archive interface prototypes independent of the backend DBMS.

UEDIN has recently been prototyping the use of Web 2.0 technologies for the delivery of an intuitive, but richly-functioned user interface to sky survey archives with a complicated schema,

and this appears promising for Gaia: functionality like making schema information readily available to users as they develop their queries, and, even, using code completion to help write them, can make archive use much more effective.

The interface is able to offer users the ability to explore data interactively: they can execute a query, generate summary plots (e.g. scatter plots, histograms, etc), realise their query was not quite making the desired selection, and then easily tweaking the query and executing it again. This reflects the iterative method of working that scientists naturally adopt, which is clearly revealed in analyses the query logs from sky survey archives such as the WFCAM Science Archive [6], curated by UEDIN, and this iterative workflow can be made to run efficiently using a combination of client- and server-side technologies.

What is most important is that the functionality prototyped is that prioritised by scientists, and that any testbed developed here helps the user community to further refine their expressed requirements. For example, while GAP has successfully engaged the Gaia user community via a call for ‘usage scenarios’ under the auspices of GREAT (and these form the inputs to WP200), iteration of requirements with these key consumers has not been considered so far. This process will drive the further development of user interface design – e.g. in determining which additional graphical capacities to implement, and to assess how sophisticated a caching mechanism is required to support the division of datasets between the client and the server – and we propose to use the interfaces developed by this WP for an initial deployment as a testbed for the community to further assess its requirements. The work will be undertaken by Read (UEDIN).

WP-330: VO infrastructure (6 CSIC +1.8 CNRS +18 INAF +9 UEDIN staff months):

The past decade has seen a huge amount of activity in defining, standardising and implementing the global ‘Virtual Observatory’. From the outset, large-scale mission data sets from ground and space were anticipated as being the cornerstone of the VO. This work has reached a level of maturity whereby most of the basic interoperability standards are in place¹¹ and it is possible to build project-specific services on top of them and to see where the further development of standards is needed in support of particular projects.

Our goal in WP-330 is a focused programme of VO consolidation and development work concerning server-side components (as opposed to client-side applications; see WP440) to provide the particular VO infrastructure required for Gaia exploitation. This will involve the following strands of work:

- i) Assessment of compliance with VO standards (Solano, 6 sm CSIC): to test, and implement the Virtual Observatory standards and protocols necessary to make Gaia data fully VO compliant. We will define the list of VO standards applicable to Gaia data; implement VO standards in Gaia simulated data; and document using simulated data and IVOA standards and protocols as inputs. The main deliverable will be a specification for VO-compliant Gaia data.
- ii) Deployment of specific web services (Berthier, 1.8 sm CNRS): the SkyBOT¹² service suite will provide VO-compliant tools for the treatment of solar system bodies within Gaia data, while Miriade¹³ computes positional and physical ephemerides of known solar system bodies in a VO-compliant manner.
- iii) VO-Dance (Smareglia, 18 sm INAF):The VO-Dance suite provides a lightweight method of publishing data to the VO. Its components can be distributed as disk images to be run on a virtual machine, so we shall assess its use as a means whereby users can integrate their own datasets with Gaia data.
- iv) VOSpace (Voutsinas, 9 sm UEDIN): Support for an extension to the current VOSpace func-

¹¹<http://www.ivoa.net/Documents/>

¹²<http://vo.imcce.fr/webservices/skybot/>

¹³<http://vo.imcce.fr/webservices/miriade/>

tionality so that, in addition to providing users with file storage space addressable by VO access protocols, they can also have database storage space on the same basis. This will provide users with a personal database facility like the SDSS *MyDB* systems, which they are able to address in a VO-complicant manner. For example, a user will be able to direct the result set from one VO query into their personal database, and then use it as the target for a subsequent query, possibly also involving other datasets in the VO, using the TAP Factory system of WP-340 below.

WP-340: Data Centre Collaboration (10 UEDIN staff months):

With the Table Access Protocol (TAP¹⁴) the VO provides a standard means of querying tabular data sets, and with the advent of the TAP factory [8] it has become possible to execute multiple, distributed TAP queries. In a traditional IVOA TAP scenario, single TAP endpoints provide the means for VO clients to present the user with a data resource schema and then to service an ADQL query on that resource, but it is then up to further, separate client–end manipulations to join data for multiwavelength science. TAP Factory takes this further by combining TAP with the Open Grid Service Architecture Data Access Infrastructure (OGSA–DAI) middleware to provide a means of creating TAP end-points on–the–fly, and, thereby, facilitating the cross-querying of distributed resources by TAP clients.

Such a system supports one of the fundamental usage scenarios for the VO. A user can select a set of data resources published using TAP on which to execute a distributed query. From the metadata exposed by the individual TAP services, TAP Factory is able to create a new TAP endpoint on–the–fly for the distributed query and present the user with the metadata of the virtual data federation thus generated. The user can then pose a query against this virtual federation as if querying a single TAP service, and, when coupled with the MyDB–like personal database of WP-330, it enables users to create sophisticated sets of cross–catalogue queries, as required for the full exploitation of Gaia data. The key point here is that a data resource can be incorporated into a virtual federation without requiring any action on the part of the staff of the data centre that curate it; so, in the case of Gaia, it is possible for higher level services like these to be developed and deployed, without requiring any action from (or placing any obligations on) the staff at ESAC.

A basic prototype of this system has been produced by UEDIN, but it needs further development in several related regards before it is capable of supporting the scientific exploitation of Gaia. Firstly, the efficiency with which the system can execute a distributed query over the virtual federation constructed by TAP Factory depends on the metadata available to OGSA-DAI's Distributed Query Processor (DQP) for the purposes of constructing a good query execution plan. For example, if DQP knows the distribution of values of the attributes used in join clauses in the distributed query, it can make an informed decision about how best to move data in executing the query, and whether to perform any server-side pre-processing before doing so. Taking full advantage of these capabilities will require an extension to the TAP standard, to expand the range of metadata exposed by a TAP service, and this can be best progressed through the IVOA standardisation process by the demonstration of powerful prototypes performing realistic science analyses.

The efficiency of the distributed queries can be improved further by collaboration between data centres. A naive spatial cross-match query executed between distributed multi–TB data sets will remain expensive, given network speeds, but several strategies exist that can ameliorate this situation and this work package will assess, through quantitative analysis – and, where possible, direct experimentation – the optimal configuration of the multi-wavelength datasets required for the scientific exploitation of Gaia. For example, to determine which external catalogues should be co-located with a copy of the Gaia archive, for which should “cross-neighbour” tables be pre-computed to facilitate queries between data sets that remain geographically separated, and for which can cross-matches be performed on-the-fly with sufficient speed. The work will be under-

¹⁴<http://www.ivoa.net/Documents/TAP/>

taken by Read and Voutsinas of UEDIN.

WP-350: Cloud-based research and data mining environments (6 UEDIN staff months)

Research environments such as that provided by CADC with CANFAR¹⁵ represent state-of-the-art solutions to the large and growing range of research and data mining demands being placed upon astronomical archives. CANFAR offers scientists a rich, yet bounded, environment based on virtual machines (VMs), within which a scientist can deploy the software they need for their individual research and have it run in a manner that does not risk the stability of the archive or the research of other scientists. VM images can be created and stored by individual scientists or research consortia, and deployed when, and in the numbers, necessary for the job at hand, so that the available data analysis hardware can be employed effectively, but with the flexibility needed to match the differing needs of multiple user groups.

As archives increase in size and complexity, data analysis will shift to the data centre, and the CANFAR initiative is showing how this can work in practice. Of particular relevance to this project is the recent work¹⁶ deploying the Skytree scalable data mining software within the CANFAR cloud, which has demonstrated how such the provision within a data centre of such a virtualized environment can support the large-scale data mining analyses envisaged for Gaia by WP-400. CANFAR is the pioneer in this domain, but further R&D work is needed to shape a system that will be suitable for Gaia: e.g. further integration with VO protocols (see WP-330 above), and creation of a more sophisticated packaging system for deployable software.

The work of WP-350 will centre on the prototyping the deployment, configuration and enhancement of a virtualized data analysis environment for Gaia. Starting with the existing CANFAR system, it will identify best practice and requirements for further development, some of which can be prototyped within WP-350. Comparison with other solutions for Gaia analysis within the data centre will be undertaken and conclusions reported. This work will be undertaken by Read at UEDIN.

Deliverables

Del. Num.	Description	Month
300.1	GENIUS/ESAC–SAT Co-ordination and Interface Control document	3
300.2	Web2.0 user interface demonstration prototype deployment	18
300.3	Gaia Data IVOA compliance document	42
300.4	Deployed web services, code and documentation	42
300.5	Data centre report and analysis document	42
300.6	TAP+ code and documentation	42
300.7	Deployed CANFAR–style VM research environment and produced reports and documentation	42

¹⁵<http://canfar.phys.uvic.ca/>

¹⁶https://sites.google.com/site/nickballastronomer/research/canfar_skytree

Table 1.3d IV - WP400

WP number: 400		Start: Kick-off meeting				
WP title:	Tools for data exploitation					
Activity type:	RTD					
Participant number:	1	7	10	11	2	
Participant short name:	UB	CSIC	FFCUL	UBR	CNRS	
Person-months per participant:	43	12	28	3	1.8	
<p>Objectives:</p> <p>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 will be needed. This work package is devoted to the development of such tools, in close coordination with WP200 to ensure that they are tailored to the actual needs of the scientific user community. It will include:</p> <ul style="list-style-type: none"> • 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 WP 200, that will involve complex and massive exploration of the data. <p>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.</p>						

Description of work:

The UB team leads this work package and will contribute most of the resources devoted to it. The personnel at the UB (see Sec. 2.2.1), led by the GENIUS coordinator X. Luri, will provide its extensive background on astrometry in general and the Gaia data in particular, and its knowledge and experience on the use of astronomical data. In addition, an experienced software engineer will be hired with the GENIUS funding and devoted full time to WP400 to provide the technical expertise necessary for the developments in this work package with the support of the UB staff. Some funding will also be devoted to specific tasks along the schedule, to employ part time software engineers already working for DPAC developments in the UB team.

WP-410: Technical coordination (2 staff months, UB):

In addition to managing the resources deployed on the other WP-400 work packages, and producing reports on those activities, this work package oversees the design and specification of all

work conducted under WP-400, to ensure that it adequately addresses the requirements identified within the GENIUS project and from external sources, such as the CU9 and GREAT.

This WP also includes the liaison with Gaia and Science Archive team members at ESAC for the coordination in the development of exploitation tools working on the Gaia archive.

WP-420: visualization tools (28 FFCUL + 8 UB staff months)

This Work Package addresses the development of visualization tools and solutions, adapted to the large size and complexity of the Gaia archive. This includes interaction with the data, resulting in seamless visual queries to the archive.

The full understanding of the Gaia catalogue data requires a rich set of visualization tools, that will help in the human interpretation of the data and knowledge discovery from its internal relation. To achieve that, the visualization package should support a wide variety of visualization algorithms including geometrical, volumetric methods and also advanced topological and modelling algorithms (i.e. polygon reduction, contouring, or glyphs) among others. Besides that, we must consider modern concepts of displaying (statistical) data, moving beyond simple histograms or plots towards visual knowledge inspiration and persuasive presentation components (i.e. voxels, hixels, texels representations). It will be also important to go one step forward in current research areas such as visualization of the uncertainties (errors, and their models must be seamlessly integrated and never ignored), user interactivity or cosmetics (essential for outreach, WP-450).

The core components of the visualization framework that interact with different (N-dimensional) graphic widgets and the algorithms will have to be provided as part of this package. Internal (server-side) parallel processing of massive data sets and provision for easy human interaction will have to be considered. From the hardware infrastructure the visualization package will have to allow for a flexible definition underlying the client and server-side egressing technologies and platforms.

Although Gaia data will be multi-dimensional, visual exploration in Astronomy is mostly done using 2D representations. This reduced dimensionality has a price: It easily hides features and relations in the data and can produce cluttered views. Multiple 2D panels are often used as a solution, but the linkage between data in different panels is frequently not clear. Curiously, 3D visualization, with the gain of an extra visual dimension, is not widespread in Astronomy, where most of the data are individual entities (stars, galaxies, asteroids). It is almost exclusively used in simulations of astrophysical fluids and fields, which are extended bodies. The reason is a lack of good tools for 3D selection and interaction with point clouds. 2D interfaces, such as a mouse and keyboard, are not adapted for this kind of interaction. This is one of the most critical inhibitors of the advantages of using the extra third dimension in scientific research. There is clearly a need of developing an adequate tool for 3D interactive visualization supporting human-computer interfaces other than the mouse and keyboard.

Besides our own developed components, the analysis for the reuse and extension of widely accepted (astronomical) visualization software will be considered as part of the WP tasks. In particular the tools that support VO formats will be targeted (i.e. TOPCAT, VOSpec) in coordination with WP-440. Those tools are already using a set of different astronomic formats and allow the inclusion of several user defined formats. They also provide widgets for higher dimensional visualisation, statistics algorithms or visual comparison that will be adapted to visualise the contents of the Gaia archive and compare it against other archives. Other existing tools will have to be examined, in particular the ones that deal with parallel visualization on large clusters (i.e. using MapReduce), the open-source ParaView coprocessing library (that uses VTK) or VisIVO, a current parallel processing capable visualization tool well known in astronomy.

The tasks in this sub-work package include the contributions of the FFCUL specialised partner. The team at FFCUL will provide expertise in the development of visualization tools. Their activity

in visualization studies and developments for space and earth observation further allows GENIUS to take advantage of the synergies with fields other than astronomy. Furthermore, as described in Sec. 2.3.3, the team at Vienna University led by J. Alves has expressed its interest in future collaborations with GENIUS on this issue (letter of interest sent to the coordinator).

The following tasks have been identified for the visualisation WP:

- Define the list of requirements and feasible use cases to be covered by visualization.
- 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 WP430 the requirements for data mining visualization.
 - Define in collaboration with WP440 the infrastructure technology compatibility and extensions to use VO standards and services.
- Implement, test and monitor the visualisation and interaction tools (widgets and algorithms).

WP-430: Data mining (DM) (30 UB + 2 CSIC staff months)

The Gaia catalogue will represent an unmatched opportunity to apply data mining techniques and algorithms as tools for knowledge discovery in a domain where there is no alternative to automated methods based on statistical learning (human exploration is certainly not feasible except for very limited subsets of data). The application of the data mining algorithms in order to extract new knowledge from the data is mandatory for a full scientific exploitation of the Gaia data. The main focus will be on Knowledge Discovery which is expected to reveal patterns and relationships within the astronomical data that can lead to the detection of new types of objects or isolated, exotic objects that represent rapid stages of stellar evolution and/or new astrophysical scenarios. Also, modelling tasks will arise from the discovered patterns. In that sense, the capability of automated dimensionality reduction (feature extraction, feature selection) and the development of key learning algorithms (clustering, outlier analysis, swarm intelligence, ...) implemented for parallel processing are foreseen as important.

From the architecture point of view, the DM module will have to scale to the entire Gaia data set and allow for a flexible definition of the underlying infrastructure (Cloud Computing, High Performance computing (HPC), GRID computing, and other emerging technologies). The initial approach we plan is an architecture where the mining algorithms are accessed following the paradigm of Software as a Service (SaaS) over a service oriented architecture. However, the package should also be compatible with future definitions of data mining processes, that are expected to include more complex mining work flows supporting asynchronous notifications from those services.

The tasks in this sub-work package are mainly under the UB partner, and also include the contribution of the CSIC specialised partner. Through the CSIC the team of L. Sarro will provide to GENIUS its expertise in Data Mining in astronomy, including the synergies with his work in the area inside the Gaia DPAC (see Sec. 2.2.7).

The following tasks have been defined for the data mining WP.

- Define the list of requirements (in coordination with WP200) and feasible use cases to be covered.

- Define the architecture to support the mining processes listed in the requirements.
- Define the framework to allow users to develop their own implementations of the mining algorithms.
- Define the proper data models for the data mining based on the requirements. In particular:
 - Define in collaboration with WP420 (Visualisation) the requirements for dimensionality reduction.
 - Define in collaboration with WP300 the infrastructure technology compatibility for the data mining work flows needed by the requirements
- Parallelise existing algorithms or libraries for Data Mining in distributed environments

WP-440: VO tools and services (10 CSIC + 3 UBR staff months)

Besides novel modes of access to the entire Gaia archive and the emerging needs on visualisation (WP420) and data mining (WP430) it is anticipated that the more traditional archive access mode – in which a potentially complex query downloads a data set of modest size for interactive client-side processing – will continue to be important. The most efficient way to support this model is to provide a seamless interface for Gaia data acquisition from existing analysis tools in which astronomers already have expertise. We therefore intend to extend the following existing VO applications with Gaia-specific data acquisition tools:

- TOPCAT (Tool for OPERations on Catalogues And Tables)¹⁷ is an interactive graphical application for exploration, analysis and manipulation of tabular data, especially source catalogues, which works well with moderately large data sets (up to a few million rows and a few hundred columns; more details are given in 2.2.11). TOPCAT already offers a number of service-specific load dialogues (e.g. VizieR, Millennium Simulation), and a Gaia option would be added alongside these. Additionally, investigations will be made of whether the existing practical limits on dataset size can be increased. TOPCAT is in regular use by certainly hundreds and perhaps thousands of astronomers worldwide, and has users in 24 of the 27 EU member states. Providing direct access to Gaia data from this tool will be a highly effective way to facilitate an entry point for its exploitation.
- VOSpec : Gaia will produce a large set of spectra (spectrophotometric data for all the objects and high-resolution spectra for all objects up to $G \leq 17$). VOSpec is a ESA-VO tool that can handle spectra in the VO context. It offers multi-wavelength spectral analysis and spectral widgets. The inclusion of Gaia-specific modules are foreseen for the users that have to work with spectra processing in Gaia.
- VisIVO : (Visualization Interface to the Virtual Observatory) is an open-source tool developed following the VO standards and recommendations. Data is retrieved by connecting to a VO service and loaded locally for manipulation or visualization. It can deal with multidimensional data sets of both observational and simulated data. It offers parallel processing facilities that will need to be extended to fully exploit the access to the Gaia data.
- VOSED: is a tool developed in the framework of the Spanish VO to ease the generation of Spectral Energy Distributions (SEDs). VOSED is able to build SEDs gathering information from the spectroscopic services available in VO. These datasets can be complemented with photometric information from a number of VizieR Catalogues as well as with data provided by the user.

¹⁷<http://www.star.bris.ac.uk/~mbt/topcat/>

- VOSA¹⁸: a tool to query photometric catalogs accessible through VO services, query VO-compliant theoretical spectra and calculate the associated synthetic photometry and derive physical parameters from the model that best reproduces the observed data.

The tasks in this sub-work package include the contributions of the CSIC and UBR specialised partners. At CSIC the team led by E. Solano (Spanish Virtual Observatory, see Sec. 2.2.7), will provide VO support and at UBR M. Taylor (main developer of TOPCAT and other VO tools, see Sec. 2.2.11) will provide the TOPCAT integration.

The following tasks have been defined for this sub-work package:

- Define the list of services and tools specifications to be covered using VO for Gaia. In particular:
 - Define in collaboration with WP420 (Visualisation) the requirements for VO tools and services.
 - Define in collaboration with WP430 (Data mining) the requirements for VO tools and services.
- Design and Implement VO services and tools for the Gaia data.
- Test and optimise, and validate of the VO tools and services providing performance monitoring.
- Define/implement the query extensions necessary to query the catalogue to fulfil the specifications.
- Obtain user feedback and update the tools and services if necessary
- Write documentation

WP-450: Community portal, outreach and academic activities (3 UB + 1.8 CNRS staff months)

The basic dissemination of the Gaia data is already ensured by the usual ESA dissemination activities through its three-level web portal system, including the archive portal to be provided at ESAC. However, in order to increase the visibility and dissemination of the Gaia products we propose to make an additional effort by building a GENIUS community portal aimed to provide enhanced dissemination tools.

This portal should, on the one hand, act as an entry point and hosting site for advanced services for the scientific community, tied to the tools developed in GENIUS (and beyond). On the other hand, and although not explicitly included in the FP7 call, we believe that the task of bringing astronomy to the general public and the provision of resources for teaching astronomy based on actual Gaia data is a worthy contribution to dissemination of space mission data on a global scale.

The work in this work package will be carried out in coordination with WP200 in order to ensure the gathering of the inputs and contributions of the widest possible community. It will use a relatively small fraction of resources aimed to coordinate and integrate into the portal the inputs of the wider Gaia community (DPAC, GREAT, REG, etc.) for the professional, outreach and academic users leaving the (larger) resources for the design, implementation and maintenance of the portal to be included in WP600 (to be provided by the CESCO partner). Special care will be taken to coordinate the activity of this portal with the official Gaia sites at ESA, where one could even consider an integration of the portal if feasible.

¹⁸<http://svo.cab.inta-csic.es/theory/vosa/>

Specifically, outreach will mainly rely on the outside contributions of several teams (in GENIUS and more generally in CU9) that have expressed an interest in outreach, in particular the UB team and the CNRS team at the Observatoire de Paris-Meudon and the team at Vienna University led by J. Alves (Sec. 2.3.3) have expressed its interest in this task.

A first description of the outreach ideas for the community portal follows:

Main home: entry point, including a description of the main characteristics and highlights of the Gaia mission and the contents of the catalogue.

Tools for teaching: a set of tools will be developed for teaching purposes. They will be mainly visualization tools of specific areas of the catalogue data, derived from the visualization tools developed under WP420. To facilitate the data access this site is expected to work with a reduced, light-weight version of the Gaia catalogue (for instance restricted to astrometry and some photometric data, or only including the catalogue information up to a given bright magnitude), but the access to the full catalogue will be open to all for those willing to go beyond these tools.

A few examples of the possible tools are: A 3-D display of the solar neighbourhood from any point of the space, a dynamic sky tool showing the sky at different epochs based on the Gaia proper motions, a simple tool to generate HR diagrams of clusters or specific populations.

News: the goal is to present to the general public the last news of the mission, from the operational phases, to the catalogue releases or the scientific works from the Gaia data. This section will most likely be based on a feed off the news sections of the main Gaia and DPAC web sites, specifically tailored for the needs of GENIUS with the addition of internal news.

Science alerts: Gaia will generate a lot of scientific alerts during the 5 years of operation (i.e detection of supernovae or new solar system objects). Using a feed from the main Gaia Science Alerts system (see WP600) these can be presented in the GENIUS site adding specific links to the catalogue data.

Once GENIUS is finished we envisage a handover of the community portal to the wider Gaia community, for instance through one or more of the the national partners involved in GENIUS.

Deliverables

Del. Num.	Description	Month
400.1	Requirement specification document for the exploitation tools	6
400.2	Opening of the community portal	6
400.3	Delivery of first prototype of exploitation tools	12
400.4	Delivery of second prototype of exploitation tools	24
400.5	Delivery of third prototype of exploitation tools	36
400.6	Delivery of exploitation tools user manuals and technical documentation	36
400.7	Deployment of exploitation tools on the first actual Gaia archive	42
400.8	Handover of the community portal to the Gaia community	42

Table 1.3d V - WP500

WP number: 500		Start date: Kick-off meeting				
WP title:	Tools for data validation and analysis					
Activity type:	RTD					
Participant number:	2	8	9	7	10	13
Participant short name:	CNRS	UNIGE	ULB	CSIC	FFCUL	KU
Person-months per participant:	75.6	6.5	6	4	2	2
Objectives:						
<p>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, . . .) is daunting, and would be impossible without tools adapted to work on such a massive and data-diverse archive. This work package aims at 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 will rely 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 WP 400, will also produce tools for the use of the scientific community in its analysis of the Gaia data. This work package will undertake the following tasks:</p> <p>520. Looking for trouble: definition of problem cases, validation scenarios and tools</p> <p>530. Simulation versus reality: from models to observables</p> <p>540. Confronting Gaia to external archives</p> <p>550. Data <i>demining</i>: outlier analysis</p> <p>560. Transversal tools for special objects</p>						

Description of work:

Despite the precautions taken when building the data processing algorithms, completely avoiding errors in the astrometric, photometric, spectroscopic or classification data in a one billion source catalogue, with many intricate data for each, is indeed an impossible task. Still, provisions should be made for ensuring the highest quality for the Gaia Catalogue through a data validation before each release.

While every Gaia DPAC Coordination Unit (CU) has indeed implemented unit tests and verification tests, a validation between CUs, and a comparison with external data can offer, perhaps not a final word, but at least a useful complementary insight. The present section details the tools, either interactive or automated, devoted to validation purposes. As much as possible, the validation tools will rely on requirements, methods and tools developed in the other work packages in order to validate not only the data but also the other tools developed within GENIUS.

WP 510: Technical coordination (2 CNRS staff months)

The objective of this work package is to ensure that WP 500 meets its objectives within budget and on schedule. Tasks will include co-ordinating and supervising activities to be carried out, monitoring project progress, monitor quality and timing of deliverables, reporting back to the GENIUS executive board.

The manager of this Work Package will be responsible for management and progress reports and ensuring a good coordination with the other Work Packages and with CU9 needs.

WP 520: Looking for trouble: definition of problem cases, validation scenarios and tools (10 CNRS staff months)

A basic verification of the Catalogue content should ensure that the field contents are as expected, that all fields are within valid ranges and fields present as indicated (e.g. spectroscopic epoch data should be present when and only when indicated). Blind automated tools for fulfilling these simplest basic tests are thus needed. Besides, a consistency of this content with documentation is mandatory.

Complementing this formal validation of the Catalogue output, more complex tests should be elaborated, and the associated tools should be developed. For instance, the fact that Gaia is a complete observatory in orbit, combining astrometric, photometric and spectroscopic information implies some redundancy which can be exploited for validation purposes; for example, photometry should be consistent with spectroscopy. Other intrinsic correlations between parameters can be used to build those tests, such as e.g. the dependence of proper motion on distance.

On a forward modelling side, it is of interest to wonder what kind of problems could occur and what consequences this would have on observed parameters. Some expected problems which would produce errors in the Catalogue are the following:

- Calibration or instrumental problems
- Classification errors
- Data Processing shortcuts or approximate models

This work package will accordingly define validation scenarios, and implement the corresponding tests. Some illustrative examples can be given:

- It is expected that photometric calibration problems would introduce a spurious variability for stars. Consequently, the analysis of stellar variability either spatially or versus time can validate the data or exhibit calibration problems.
- On the astrometric side, any annual thermal or calibration effects would introduce a parallax bias, as was already studied for Hipparcos, so the parallax zero-point should be studied as in, e.g., [1].
- Bad cross matching of Solar System Objects (SSO) would produce spurious SSOs or stars, so the distributions of the distances to the nearest neighbour, from SSO observations to nearest non-SSO, is a useful test.

Summarising the above comments, the corresponding work packages would then be the following:

521. Formal validation of the Catalogue field content as function of the object type
522. Internal consistency tests
523. Tests based on what is known to produce effects on given parameters
524. Generation of validation reports with diagnostics filtering

WP 530: Simulation versus reality: from models to observables (14 CNRS staff months)

The CU2 DPAC Coordination Unit has provided a very valuable tool: the Universe Model. Indeed, this model, initially based on the Besançon Model of our Galaxy, has been complemented with an extinction model, multiple stars and variability models, etc., and now represents the best simulated sky one could hope to test the DPAC algorithms against.

In turn, this model can be used to validate the Gaia data. In a first step, the (astrometric, photometric, spectroscopic or classification) observable parameters which are predicted by the model should be computed, in the form of statistics: distribution, confidence intervals and correlations between parameters, by object type, by region and by time.

Certainly, differences between what is predicted and what is observed are expected (or even desired) and the comparison between model and observed data requires clustering tools (WP 551) and robust implementations (WP 553). Clearly, for several parameters, checks will have to be made separately for different classes of sources and it would also be desirable that scientists are able to apply their interpretative skills to the comparison of model versus data.

531. Statistics of the parameters deduced from models

532. Build tools checking that all Catalogue fields have ‘reasonable’ distributions, i.e. consistent with what is obtained in WP 531

The CNRS UMR 6213 is the one responsible of the CU2 Universe Model and will then very efficiently tackle these tasks.

Projecting models into the observable domain (such as the Universe model mentioned above) is a task in common with WP 230, the difference being that validation expects to retrieve from data already known specific structures while scientific users of the Catalogue will expect to find new extra ones. The development of the needed tools will consequently be done in close cooperation with WP 230.

WP 540: Confronting Gaia to external archives (17 CNRS +4 CSIC +2 KU staff months)

One of the first uses of the Gaia data will be the cross-matching to external archives where the astrometry will allow 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, what is important is that a photometric analysis should show the consistency between Gaia data and external data.

The problem — which is actually not a problem but one of the strengths of Gaia — is that there is no comparable all-sky survey with a comparable angular resolution and multiple star discovering power. Although the cross matching will be based on the VO tools elaborated in WP 400, the methodology to do this in practice a) in dense areas, b) with multiple objects handling, c) taking into account all properties of Gaia on the one hand and of the other Catalogues on the other hand, implies the need of developing tools allowing both input from the users and intelligence in the data pairing.

Besides, the validation mentioned here is supposed to be the validation of Gaia data, *not* that of the external archives, although this certainly would be interesting on the scientific side (and will thus be disseminated to provide input to further scientific analysis). Robustness is thus mandatory in front of the lack of data, the lack of precision, and the high level of systematics which are expected in external archives and which could wrongly be interpreted as problems within Gaia data. Robustness shall be achieved thanks to tools developed in WP 550.

541. Multi-wavelength cross-matching tools

CSIC will devote 4 man months in this Work Package. Consultancy from the INAF partner will also be useful for this task as INAF-OATo is responsible for the IGSL cross-matching algorithms in DPAC. This WP will also benefit from the developments in WP 240.

542. Photometric and classification analysis tools

543. Cross-validation tools with Nano-JASMINE data

Similar to Gaia, the Nano-JASMINE (N.J.) astrometric results will need to be validated, and many tools defined in the whole WP 500 package can indeed be used for this purpose.

Besides, it is planned to combine N.J. data with Hipparcos data to hugely improve the proper motion precision of the stars in common, thanks to the long time base between both missions. A cross-validation is needed before combining data, which will incidentally allow to detect long period binaries.

As N.J. uses the Astrometric Global Iterative Solution (AGIS) developed for Gaia by ESAC and Lund Observatory, a useful insight on the AGIS behavior with N.J. (e.g. validation of the estimation of the correlations between astrometric measurements) will be obtained when the more precise Gaia data is available. Finally, the Gaia data will also allow to test the results obtained with the validation tools applied to the N.J. data.

WP 550: Data *demining*: outlier analysis (8 CNRS +2 FFCUL staff months)

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.

While a first risk already handled is the presence of problems or systematic errors in the Catalogue, another issue is an incorrect interpretation of data features. Indeed, although three of the WP managers of the current proposal drew the attention of the community [4] to the precautions to be taken with the analysis of the Hipparcos data, this did not prevent incorrect exploitation of the astrometric data. In this respect, being able to show that objects are *not* outliers is perhaps as important. Tools dealing with extreme values are thus needed.

In an interactive discovery phase, the data analysis via graphics developed as in WP 400 should allow tolerances in order not to detect noise instead of substructures. Still, at some point special sub samples will be detected, thanks to clustering tools. At that point, what is needed is an immediate characterization (statistical analysis and classification) of the properties of this sub sample with a subsequent visualization (e.g. 3D spatial maps).

551. Clustering and sub-population statistical characterisation tools

552. From graphics to diagnostics, from diagnostics to graphics

The FFCUL node will efficiently contribute to this task

553. Robust tools using truncated, censored or correlated data

WP 560: Transversal tools for special objects (24.6 CNRS +6.5 UNIGE +6 ULB staff months)

Some special objects need a special treatment, in particular those having a time dependence such as multiple or variable stars or solar system objects. Moreover, these objects may greatly benefit from a reprocessing of the Gaia data using external epoch data. Dedicated sub-work packages led by experts of the models used in these fields are thus required here and their specialized tools will also contribute to WP 520-550.

- Detection of new objects in the Solar System is foreseen with Gaia. While in numbers they are very few (about one per 10^7 Gaia objects), in terms of classes they are scientifically valuable: one expects to detect Near Earth Objects (NEOs) inside the orbit of the Earth or bright outer Solar System objects. Thus, real time validation has to be provided for the dedicated ground-based support Gaia-FUN-SSO network; this to reduce as much as possible false alerts and also to validate data inserted in the global data analysis scheme. Specific software needs to be developed for automating alerts, transformation and dissemination of data for use by observers, and making all data and alerts VO compliant. This task will have to combine ground-based and space-based Gaia data. One also needs to compute the orbital elements and compare them to the elements of the known population of asteroids and comets, so as to perform orbital adjustment, taking into account a full dynamical model, and at the same time validate the inversion process from the limited Gaia sample (which usually corresponds to less than an orbital period around the Sun). These data are mandatory to ingest in the input auxiliary database for small solar system bodies which has to be maintained and regularly updated during the space mission.

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 can be corrupted because of a close approach to a star; in such case the information has to be provided to the group having the task to analyse stellar data. Furthermore, one will either validate the rejection of corrupted data, or retain the data as possibly valuable additional scientific input (entering so WP 550). Such analysis has to be performed on solar system objects directly observed by Gaia as well as other objects that will not be observed by Gaia but are catalogued in SSO data bases (e.g. the planets, dwarf planets, large satellites and irregular ones, and asteroids fainter than magnitude 20).

- Considering multiple stars, it should be noted that about half of the Gaia Catalogue will consist of sources that are actually non-single stars of which a significant but only much smaller fraction will be detected. Assessing the quality of the data reduction for the majority of stars not detected as non single stars (NSS) thus appears complicated (as the model fit to the astrometric measurements may be incorrect), whereas it may prove easier when a more correct astrometric model is already known, which is the case for detected NSS.

Two types of validations are identified: on the one hand, validations relying upon the statistical behavior of the solutions leading to the catalogue, i.e. purely standalone validations e.g. goodness of fit; on the other hand, validations based on a comparison with some auxiliary data, e.g. speckle observations or spectroscopic orbits. These validations may thus allow a better insight into the properties of the observations, associated uncertainties, and data reduction done in the astrometric (CU3), photometric (CU5) and spectroscopic (CU6) data reduction chains.

- Finally, the stellar variability should be studied in detail: while a certain fraction of the sources are expected to be intrinsically variable in flux, some assumed constancy of many other sources is also what permits the principles of the data reduction. Conversely, an unexpected variability can also be the signature of an acquisition or data reduction problem. Because variability is transversal to the validation process, this work package will develop tools validating the astrometric, photometric, spectro-photometric and spectroscopic reduction from the point of view of time series and variability. The study will take several directions: for example studying variable sources to determine if some of their variability behavior is due to the instrument or the reduction, or to detect if constant sources, or small

amplitude variables have some residual effects coming from the satellite or perturbations from that data acquisition mode, or reduction method. To see such effects, it is then important to gather several sources and to take averaged quantities.

A large list has already been established of all effects that should be studied to see if there are some residual effects in the (spectro)-photometric, and spectroscopic time series.

561. Solar system objects

The competence of the CNRS UMR 8028 will prove useful if not mandatory for this task.

562. Multiple stars

The ULB node contribution, in charge of the CU4 NSS handling in DPAC, will be needed for this Work Package.

563. Variability and time series

This WP is where the competence of the UG node (in charge of the coordination unit, CU7, responsible for variability processing in DPAC) will help to build tools related to the variability analysis.

Deliverables

Del. Num.	Description	Month
500.1	Delivery of prototype of internal checking tools (WP 520)	12
500.2	Delivery of prototype of statistical tools (WP 550)	18
500.3	Delivery of internal consistency checking tools (WP 520)	24
500.4	Delivery of statistical tools (WP 550)	24
500.5	Delivery of model-based validation tools (WP 530)	24
500.6	Delivery of prototype of external validation tools (WP 540)	24
500.7	Delivery of external validation tools (WP 540)	36
500.8	Delivery of special object tools (WP 560)	36
500.9	Deployment of validation tools on the Gaia archive	42

Table 1.3d VI - WP600

WP number: 600		Start date: Kick-off meeting				
WP title:	Support activities					
Activity type:	DTD					
Participant number:	1	5	12			
Participant short name:	UB	CESCA	UCAM			
Person-months per participant:	6	24	12			
<p>Objectives: This work package aims to provide support activities needed for the development of the tasks in the rest of WPs:</p> <ol style="list-style-type: none"> 1. The provision of simulated data mimicking the actual Gaia catalogue; this mock-up data will be used for testing the system, from technical tests to user trials for validation. 2. The provision of a testbed for science alerts; the prototypes of the science alerts system will be installed in it for testing and validation and made accessible to the test users. 3. The development and implementation of the basic infrastructure for the community portal (hardware, content management system, design, etc.). 						

Description of work:

The work is divided into three sub-work packages:

610 - Technical coordination (1 UB staff month): as described in 2.1 the management of this work package will be done in direct coordination with the GENIUS executive board, since the tasks included here are of a global nature and its supervision needs a global view of the project. The management of this work package will be carried out by J. Torra.

620 - Simulated catalogue data (12 CESCA + 5 UB staff months): the products of the project should be available in time for the first release of Gaia data (around mid 2015, according to present plans) so they can be used for the actual archive at ESAC. At that stage the systems and tools should be fully tested. For this purpose this sub-work package is devoted to the provision of simulated catalogues allowing to fill the database with realistic data that will allow the testing and development of the archive prototypes and the tools, and will also allow the test users to use the system in realistic conditions.

These simulations will be generated by the already available Gaia simulator, and specifically by a tool named GOG (Gaia Object Generator). GOG has been developed at the University of Barcelona in the context of the DPAC CU2 and has been providing simulated catalogue data to the DPAC for several years. Its development will continue at the UB at least until 2015 and will be used by the UB team to generate the mock-up catalogues until then. These catalogues can be shared for the DPAC development or specifically generated for GENIUS as needed. The GOG team at the UB under the supervision of the work package manager will take the responsibility for these simulations.

The running of the simulations will take place at CESCA. This center provides supercomputing resources on which the Gaia simulator has already been run on many occasions. These resources, including the hardware and technical support, will be provided in this case to ensure the availability of simulated data for testing in GENIUS.

We envisage to provide three major deliveries, listed below, of full mock-ups of the Gaia catalogue (one billion objects), although smaller partial catalogues for small scale testing will be generated on an as-needed basis.

These simulations will be conducted using the hardware and software at CESCO. For this purpose CESCO has several supercomputers with different architectures which will allow running the software needed to carry out these simulations. As an example CESCO will provide a SGI UV 1000, which is a shared-memory machine with 224 processors, for a total of 1,344 processing cores. It has 6.14 TB of memory and 112 TB of disk storage. CESCO will also provide, through the supercomputing team, the necessary technical support in order to make the most optimal use of the available hardware and software. Suffice it to say that this is a team of experienced professionals who have been involved with Gaia project since 2000.

630 - Science alerts testbed (12 UCAM staff months): 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 supernovae (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 this WP 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 up 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 WP300. Finally the alerts testbed will plug into the portal testbed developed in WP640.

With the termination of the GENIUS WP630 testbed activity, the full functionality will be deployed for community use - providing enhanced access to science alert data from 2015 onwards.

640 - Community portal infrastructure (12 CESCO staff months): The development and implementation of the community portal, based on a personalization and customization of a content management system (like Drupal or similar), will be done following a user-centered design (UCD) methodology. The portal will include information about the project, documents, links to references, alerts, news, tutorials, etc., as well as all the necessary contents that will be defined in the first steps of the portal analysis. In addition, the basic infrastructure to host the portal will be delivered, including the consequent maintenance of the needed software and hardware.

Once the portal is ready, content provision will be continuously done to spread information on the project's status and to provide up to date and useful information (news, images, videos, ...) created within the project for each community (general public, science community, ...) and with the researchers' guidance.

At the same time, the use of social networks will help to increase awareness of the project and to attract people interested in Astronomy to the portal. The community manager will publish real time information in social networks, which will generate traffic to the portal and will help to spread the project and its achievements.

Deliverables

Del. Num.	Description	Month
WP	Description	Month
600.1	Basic setup of the community portal internally available for working	3
600.2	First public version of the community portal	6
600.3	Delivery of first simulated catalogue data	6
600.4	Deployment of first public science alerts prototype	12
600.5	Delivery of second simulated catalogue data	18
600.6	Deployment of second public science alerts prototype	24
600.7	Upgraded public version of the community portal	24
600.8	Delivery of third simulated catalogue data	30
600.9	Deployment of third public science alerts prototype	36
600.10	Final (complete) version of the community portal. Hand-over to the Gaia community.	42

Table 1.3e - Summary of staff effort

Participant no. & short name	WP100	WP200	WP300	WP400	WP500	WP600	Total person-months
1 UB	24	–	–	43	–	6	73
2 CNRS	2	–	1.8	1.8	75.6	–	81.2
3 UEDIN	2	–	41	–	–	–	43
4 UL	2	47	–	–	–	–	49
5 CESCA	–	–	–	–	–	24	24
6 INAF	–	16	18	–	–	–	34
7 CSIC	–	–	6	12	4	–	22
8 UNIGE	–	–	–	–	6.5	–	6.5
9 ULB	–	–	–	–	6	–	6
10 FFCUL	–	2	–	28	2	–	32
11 UBR	–	–	–	3	–	–	3
12 UCAM	2	4	–	–	–	12	18
13 KU	–	2	–	–	2	–	4
Total:	32	71	65.8	87.8	96.1	42	395.7

1.3.4 Dependencies diagram

Figure 1.4 shows the dependencies between the various work packages in GENIUS. The arrows indicate a ‘depends on’ relationship. For example WP300 depends on WP200 for the requirements specifications while WP200 depends on WP300 for testing the implementation of the requirements. WPs 300/400/500 are all enabled by the simulated data provided through WP600.

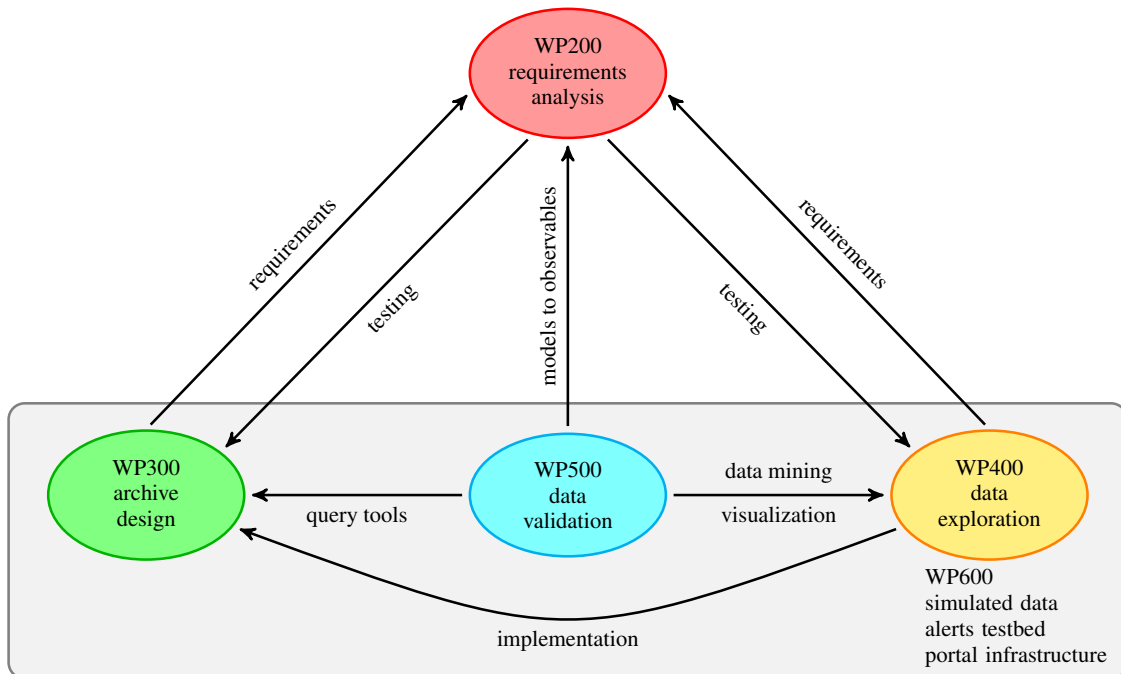


Figure 1.4: The dependencies between the work packages in GENIUS.

1.3.5 Risk analysis

Summary (see below for a detailed discussion of each risk)

Risk	Mitigation of probability	Mitigation of impact
External risks		
Gaia failure	N/A	Transfer of knowledge and tools to astronomical data centers, specially to ESAC
Gaia launch delay	N/A	Resynchronisation of GENIUS development with Gaia releases
Catalogue release delay	N/A	Resynchronisation of GENIUS development with Gaia releases
External coordination problems	Participation of partners in DPAC and GST	Executive Committee and External Advisory Board to forewarn and act in advance
Internal risks		
Late recruitment	Six month preparatory stage already allocated in the schedule and early advertising	Executive Committee to track recruitment
Developers leave project	Motivation of staff	Executive committee to supervise personnel status
Hardware testbed availability	Participation of partners in CU9; representation of ESAC-SAT in the External Advisory Board	Computing resources available at CESCO partner in case of need
Unexpected complexity	Rigorous tracking of WP development by the Executive Committee	Drop or delay low priority tasks

External risks We list below the identified risks caused by factors external to GENIUS. We explore the consequences and possible mitigation actions.

- **Gaia failure:** the Gaia mission is scheduled for launch in end-2013 and therefore, being a mission not yet in operation, there is the possibility of some type of problem (i.e. launcher failure) causing the mission to fail and in this case the basic goal of GENIUS would be jeopardized. This is a risk completely out of our control, and possible mitigation depends on the final outcomes.

If ESA decides to rebuild Gaia, as it did when the Cluster mission failed ¹⁹ the problem would be equivalent to a large launch delay; the GENIUS developments would still be fully useful, but provisions would have to be made for a much later availability of the Gaia archive.

Otherwise, the research and development within GENIUS is focused on the Gaia mission but the results that will be obtained and the expertise that will be built up can readily be transferred to archives from other large sky surveys and space missions. Specifically, since GENIUS will closely cooperate with the Science Archives Team at ESAC — where many mission archives are hosted, and more will be hosted in the future — and several large astronomical data centers through CU9 (for instance the CDS), the technologies developed

¹⁹<http://sci.esa.int/science-e/www/object/index.cfm?fobjectid=31258>

can remain as in-house technologies there for present and future space astronomy missions. Furthermore, the technologies would still be applicable (though at a smaller scale) to the Japanese JASMINE missions.

Thus even if Gaia would not be rebuilt the efforts of the GENIUS collaboration would not go to waste.

- **Gaia launch delay:** likely delays in the Gaia launch are presently estimated to be of the order of a few weeks (the integration schedule of the spacecraft having now converged and being under full control of the main contractor, Astrium). Such a delay would not pose much of a problem for the GENIUS developments, and will only require to review the synchronisation of the GENIUS completion with the first release of the Gaia catalogue. Since the GENIUS heritage is anyway planned to be integrated in the wider works of the DPAC CU9, this can be easily handled. Furthermore, since the launch will be in the early phases of GENIUS the handling of this problem can be already accommodated in the initial planning.
- **Gaia catalogue release delay:** even if the Gaia launch is on schedule, the first release of the Gaia data may be delayed due to internal problems in DPAC. The effect of this would be similar to the effect of a launch delay discussed above and, furthermore, since coordination with the DPAC work is an essential component of GENIUS (see Sec. 1.1.3), such delay can be accommodated in the planning as soon as/if it is detected in the DPAC planning.
- **Coordination with external actors:** as described in 1.1.3, the development of the Gaia archive will be a shared responsibility with external actors (specially ESA and DPAC). The joint work with these actors may imply a risk if the coordination is not properly handled, but this is strongly mitigated by the fact that the GENIUS participants are already deeply involved in the relevant consortia and committees.

Internal risks This project represents a research and software development effort and the main internal risk is delays in the deliverables of GENIUS. We list below what we consider to be the major possible causes of delays and describe the mitigating actions.

- **Late recruitment of developers:** the impact will be a lack of human resources in the early stages of the project and thus delays in the deliverables. This risk is mitigated by the fact that we have explicitly allowed for a six month period to recruit the developers. In addition we intend to start advertising the project as soon as possible if we are successful with this proposal.
- **Developers leave the project early:** The impact will again be delays due to a drop in staff effort and the period needed to get possible replacements up to speed. This risk is mitigated by making sure we hire people with a strong interest in the type of work involved in GENIUS.
- **Hardware testbed availability:** the development of the GENIUS tools will be made in the framework of CU9. Therefore, these tools will be designed for and integrated in the Gaia archive at ESAC. Such a development requires a hardware testbed for the archive (that is included in the CU9 plans) but whose availability may be subject to delays, downtimes or restrictions. However, the participation of the ESAC Science Archive Team in CU9 will mitigate this risk ensuring the coordination with the CU9 in general and GENIUS in particular.
- **The work packages are more difficult to execute than anticipated:** This risk is certainly the most likely to occur and can again lead to delays in the project deliverables. We will

manage this risk through a rigorous tracking of the progress of each WP and having decision points (such as the mid-term review) where we will re-assess the development priorities and drop activities if needed. The dropped activities will release staff resources that can be re-allocated to the higher priority activities. To take an example: in WP200 the sub-packages 220, 230, 240 are considered very important even for an early Gaia data release in order to provide the users with features they will certainly expect. The packages 250 and 260 on the other hand provide features that are likely much more relevant toward a final release of the Gaia catalogue and data archive and these could thus be dropped or given lower priority in case of delays in the progress of WP200. Similar considerations hold for the other WPs.

2 Implementation

GENIUS in a nutshell

The GENIUS pan-European cooperation (Fig. 2.2) is centred around **four main partners** — providing most of the resources, assuring the continuity of personnel for the duration of the project, and taking the responsibility for the development of the core tasks — with the support of a supercomputer centre providing computing resources and expertise for the implementation of a web portal. Additional partners provide expertise for specialised tasks within the project (Sec. 2.3). Third parties having expressed an interest in future collaborations with GENIUS are listed in Sec. 2.3.3.

The management of GENIUS is based on the principles of simplicity and efficiency, with four management layers (Fig. 2.1): The global technical management is controlled by the GENIUS board, composed of the high-level work package managers, and an external advisory board provides independent overview and advice, as well as coordination with institutes and institutions participating in the development of the Gaia archive (Sec. 2.1.1).

2.1 Management structure and procedure

The management of GENIUS is based on the principles of simplicity and efficiency. We will not introduce unnecessary committees or management structures and will rather target simple and clear management procedures.

2.1.1 Management

GENIUS is based on a four-level management structure described in Fig. 2.1.

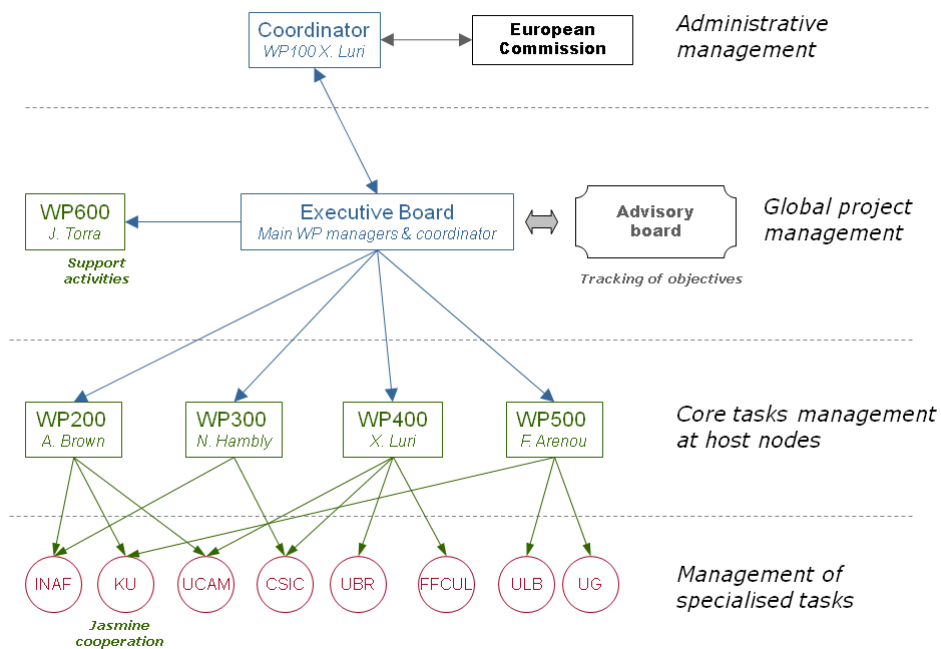


Figure 2.1: The GENIUS management structure

This top-bottom structure ensures that appropriate control and reporting takes place at all the relevant levels, and that information flows appropriately for the coordination of the overall effort:

Administrative management: this first level takes care of the administrative management of GENIUS, including financial management and reporting to the European Commission. It is carried on by the GENIUS coordinator (X. Luri) with the help of the part-time project manager hired under WP100.

Global project management: this second level is in charge of the global technical control of the project, ensuring that the tasks are properly carried out and remain on schedule, and that the objectives of GENIUS are fulfilled. It is carried out by the **GENIUS executive board**, formed by the managers of the high-level work packages: F. Arenou, A. Brown, X. Luri and N. Hambly. Notice that we have chosen to set up a small board for the overall control of GENIUS for practical reasons; the size of a committee is always a trade off between expertise and efficiency, and in our case the GENIUS structure (see Sec. 2.3) is based on a four-node core for its main tasks, with the rest of contributors tied to one or more of these main partners for their specific tasks. This core structure is mapped into the executive board, providing full visibility of the work and leaving at the same time an agile and operational committee.

In addition to the executive board, an **external advisory board** will be formed. Its role will be to provide complementary expertise in the relevant areas and an independent overview of the GENIUS progress (see Sec. 2.1.2). It will be formed by leading experts in astrophysics and engineering, including representatives of institutions and consortia related to the goals of GENIUS (a budget to cover the travel expenses of the advisory board has been included in the UB management costs). Specifically:

- A representative of the Gaia Science Team
- A representative of ESA (covering the representation of the Science Archive Team at ESAC and the mission team)
- An independent scientist with experience in fields close to the Gaia science case (to be chosen for instance from among the GREAT community)
- An independent computer science expert
- A senior scientist with experience in the preparation of the Hipparcos catalogue, the predecessor of Gaia

Additionally, the link in this committee with the wider Gaia scientific community organised in GREAT will be provided by N. Walton (included in WP100).

Notice that, as shown in Fig. 2.1, WP600 has been located at this level of management. This is due to the fact that the tasks of this WP are of global nature (provision of simulations and community portal) and its supervision needs a global view of the project. The aim is for the activity of this work package to be coordinated directly at the executive board level to which the WP manager (J. Torra) will report.

The coordination with the Nano-JASMINE and JASMINE missions through the Japanese partner at Kyoto University will take place through WP200, but we expect some higher level coordination with the Japanese missions through the executive board to allow a fluid global interaction with these space projects.

Core tasks management: the third level of management takes place at the main work packages. Each WP manager has the responsibility of supervising the execution of the core tasks assigned to their host institution, the work of the nodes carrying out specialised tasks attached to the work package (see Fig. 2.2), and report to the executive board accordingly.

We do not make explicit here the management procedures for each work package beyond the specifications given in 2.1.2, since the work in each of the work packages WP200 to WP400 will be integrated in the already existing and well established teams in each host institution. In particular, the work at UB and CNRS will join the already in-place teams working on the Gaia DPAC, benefiting from the synergies with the development of the Gaia data reduction system, the work at UEDIN will fit in the existing team working on astronomical archives, and the UL work will fit into the rich and multidisciplinary environment of the astronomy department.

Specific tasks management: finally, the fourth level of management corresponds to the control of the execution of specialised tasks in a number of additional nodes. A coordinator for the GENIUS tasks is defined for each of these nodes (see Sec. 2.2) and will report directly to the manager of the corresponding high-level package.

2.1.2 Procedures and reporting

The following procedures and reporting strategy will be applied to GENIUS management:

- The GENIUS coordinator will report to the European Commission following the rules applicable for collaborative projects in the framework of the FP7 cooperation work programme 2013.
- GENIUS will hold at least three plenary meetings open to the participation of all its members: kick-off meeting, mid-term meeting and completion meeting. The participation of the node coordinators (or a representative) will be mandatory. Other plenary meetings may be held if needed.
- Specialised meetings or workshops will be organized on an “as-needed” basis by the high-level WP managers.
- The executive board will hold teleconferences to track the status of the tasks at least once a month. The minutes of these teleconferences will be made available to the members of the consortium.
- The executive board will hold face to face meetings at least twice a year. A status report will be produced from these meetings and will be sent to the external advisory board for feedback and recommendations.
- The executive board shall ensure the coordination of the GENIUS work with the wider activities concerning the development of the Gaia archive, in particular ensuring the coordination with ESA and DPAC (see Sec. 1.1.3). This can take place in a natural way through the corresponding members of the external advisory board, but other coordinating mechanisms can be defined if needed.
- The nodes in charge of the specialised tasks shall report to the corresponding high-level WP managers as agreed internally, but should produce at least a final report for the given task once it is completed, along with the technical documentation associated.

2.1.3 Partnership agreement

In order to formalise in detail the internal workings of GENIUS and the roles and responsibilities of the partners (described in this section) a Partnership Agreement (PA) will be drawn up and agreed at the start of the programme. It will define the decision taking procedures in GENIUS and the mechanisms for conflict resolution, that will be channelled through the Executive Board, as well as the management of Intellectual Property Rights as described in 3.2.

2.1.4 Recruitment strategy

The recruitment policy for GENIUS will conform to the principles of the European Charter for Researchers and the Code of Conduct for their recruitment. It will take place in a globally coordinated way during the 6-month setup phase described in Sec. 1.3.2, placing an emphasis on individual excellence and capacity for team working and taking special care to ensure equal opportunity and gender balance as described in Sec. 5.

2.2 Individual participants

2.2.1 Universitat de Barcelona (UB)

General description:

The Departament d’Astronomia i Meteorologia (DAM) is the astronomy department of the Universitat de Barcelona. About 30 senior researchers cover areas of observational, computational and theoretical astrophysics. A team of 15 engineers is at present working within the Gaia DPAC. The UB has a Master’s degree in Astrophysics, Particle Physics and Cosmology and the current PhD programme involves 40 PhD students. The UB also coordinates the Spanish Network ‘Red de Explotación Científica de Gaia’ (125 research participants).

Role:

The UB will act as the coordinator of GENIUS and will also manage WP 400 and 600. It will host the kick-off of GENIUS as well as some specialised workshops.

Key competences and facilities:

Membership of Gaia DPAC (X. Luri) and GST (C. Jordi); lead for Gaia data simulations; Gaia core and photometric processing; large computational resources at the Centre de Serveis Científics i Acadèmics de Catalunya (CESCA) and Barcelona Supercomputing centre (BSC); responsibility for Gaia Initial Data Treatment in collaboration with the company GMV (Madrid), ESAC and DPAC; Galactic structure; kinematics and dynamics of our galaxy; stellar astrophysical parameter determination and classification of stars; multi-variate statistical techniques and data mining; development and implementation of large computationally demanding processes; participation in the Spanish MEC Programme Consolider-Ingenio 2010 ‘Supercomputación y e-Ciencia’

Key persons:

The coordinator, **Dr. Xavier Luri**, professor at the faculty of physics, ICC, is one of the authors of the “Gaia Concept and Technology Study Report” [16] outlining the mission’s scientific case on which Gaia was accepted within ESA’s scientific programme, has been a member of the Gaia Science Advisory Group (1997–2000), the Gaia Data Analysis Coordination Committee (2005–2006) and the Gaia Science Team (2001–2007).

Prof. Jorge Torra was the scientist in charge at the UB for the ELSA (European Leadership in Space Astrometry) Marie-Curie Research Training Network, supervising two of the ELSA ESRs. He was also a member of the Steering Committee of ELSA. Prof. J. Torra has participated in 55 R&D projects since 1981, 34 of them as PI. Participant in two ESA contracts for the development of the Gaia Data Processing Prototype. Participation in Hipparcos, Integral and Minisat 02 missions. About 80 publications in international refereed papers. Supervisor of 7 PhD theses Deputy Manager of Gaia DPAC Coordination Unit 3 (Core Processing). Manager of the Red de Infraestructura de Astronomia. Spanish Delegate to the ESO Council (2007–2011).

Dr. Francesca Figueras is the scientist in charge at the UB for the GREAT-ITN (Gaia Research for European Astronomy Training) Maria Curie Actions (FP7-PEOPLE-2010-ITN264895, 4.2M€, 2011–2015) supervising two of the ESRs; she is head of the Executive Committee of the network ‘Red Española de Explotación Científica de Gaia’.

Dr. Carme Jordi is member of the Gaia Science Team (GST) since 2002. She is the manager and coordinator of the Gaia UB-CU5 group and member of the Steering Committee of the ESF-RNP GREAT. She was co-coordinator of the Gaia Photometry Working Group for the elaboration of ‘Gaia Concept and Technology Study Report’ [16] and the definition of the photometric system. Dr. C. Jordi has participated in 68 R&D and/or outreach projects since 1983, 12 of them as PI. Participant in two ESA contracts for the development of the Gaia Data Processing Prototype. Participation in Hipparcos, Integral and Minisat 02 space missions.

2.2.2 Centre National de la Recherche Scientifique (CNRS)

The CNRS (National Center for Scientific Research) is a government funded research organization, under the administrative authority of France's Ministry of Research, and is the largest fundamental research organization in Europe. Three Mixed Units of Research (UMR) of the CNRS will participate in the current project.

GEPI (Galaxy-Star-Physics-Instrumentation), CNRS UMR 8111, is one of the five Departments of the Paris Observatory, founded in 1667, with more than seven hundred people over the whole spectrum of research in Astronomy and Astrophysics. The main research topics of the GEPI Laboratory in Meudon are the formation and evolution of stars in our Galaxy as well as in numerous other galaxies, which constitute the luminous matter component of the Universe. This research calls upon many disciplines, from chemistry to physics, from instrumentation to data-processing engineering, and from project management to financial management within an international framework. **Dr. Paola Di Matteo** is an expert in numerical simulations and will be a key person in WP 500, as well as **Dr. Carine Babusiaux**, deputy manager of the DPAC data simulation unit CU2.

The 'Institut de mécanique céleste et de calcul des éphémérides' (IMCCE), CNRS UMR 8028, is an institute of Paris Observatory. IMCCE researchers focus on studies of the Solar System and planetary systems in the domains of celestial mechanics, astrometry, planetology and mathematics. IMCCE is also under the umbrella of the French Academy named Bureau des longitudes. Dynamics and astrometry of natural satellites and small Solar System bodies are a main topic of research at IMCCE. Also, IMCCE is in charge of providing the national ephemerides in France. Therefore, many studies performed here in celestial mechanics and astrometry lead to applications in the ephemeris domain. **Dr. Jérôme Berthier**, team leader of a DPAC CU4 development unit, is an expert in the dynamics of asteroid systems and VO aspects and will contribute to this latter aspect in WP300 and WP400, while **Dr. Daniel Hestroffer**, director of the IMCCE, will manage the tools specialised in Solar System Objects validations in WP500 together with **Dr. William Thuillot**, former IMCCE director and head of the Gaia-FUN-SSO ground-based follow-up network.

The UTINAM institute (Universe, Transport, Interfaces, Nanostructures, Atmosphere and environment, Molecules), CNRS UMR 6213, is associated with the University of Franche-Comté (UFR) and placed under the joint supervision of the UFR of Sciences and Techniques and the Observatory of Sciences of the Universe, both at Besançon. The Institute studies the structure and dynamics of isolated systems and interactions with complex environments, over a wide range of space and time scales. The large scale systems studies include our galaxy, the Solar System and single, and double astronomical objects. An important associated activity is the establishment and dissemination of accurate standards for the measurement of time and frequency, as well as the discovery of useful extragalactic astrometric references for spatial measurements. **Dr. Annie Robin** and **Dr. Céline Reylé**, leaders of the DPAC CU2 simulated universe model, are the most competent to take charge of the aspects of the Gaia data validation using a Galaxy model.

These three CNRS UMR have been involved in the definition of the Gaia mission, the simulation of its scientific content and the data processing within DPAC, devoting in total about 15 FTE (permanent staff or personnel funded by CNES) since the beginning of DPAC.

WP500 devoted to the data validation will be led by **Dr. Frédéric Arenou**, CNRS senior research engineer at GEPI, whose PhD thesis was devoted to the validation of the Hipparcos astrometric data[1]. He is a member of the Steering Committee of DPAC CU4 and former member of the ESA Gaia Science Team (2001–2005). Within DPAC he manages two development units of CU4, one concerning the production of simulated multiple systems, the second devoted to the multiple solution combiner and is also active within the CU2.

2.2.3 The University of Edinburgh (UEDIN)

The participants from the University of Edinburgh²⁰ are members of the Institute for Astronomy's Wide Field Astronomy Unit (WFAU). Formed in 1999 during the reorganisation of the Royal Observatories in the UK, the Unit is still housed in offices at the Royal Observatory Edinburgh on Blackford Hill along with the rest of the IfA, and co-located with the UK's Astronomy Technology Centre.

Role:

WFAU is tasked with working on aspects of archive system design and infrastructure within the GENIUS collaboration. WFAU has a strong track record in the curation and dissemination of large-scale survey datasets consisting of 10s of billions of rows of catalogue data occupying 10s of terabytes: these include the legacy all-sky digitised Schmidt photographic optical surveys culminating in the terabyte-scale SuperCOSMOS Science Archive²¹; the current generation of infrared sky surveys in both hemispheres, namely the United Kingdom Infrared Deep Sky Survey presented in the WFCAM Science Archive²² and the VISTA Public Surveys presented in the VISTA Science Archive²³; and most recently the Gaia-ESO Survey²⁴, a large-scale ground-based spectroscopic public survey with the VLT to complement the Gaia mission. WFAU serves these datasets through both state-of-the-art interactive user interfaces and VO services²⁵. Over the last decade, WFAU has gained much experience in the complete archive project lifecycle, from requirements gathering and analysis, system analysis and design including aspects of hardware design, software engineering to professional standards, and archive systems deployment and maintenance. Throughout this work, WFAU has remained committed to a science-driven philosophy aimed at maximising scientific impact through appropriate use of information technologies. Furthermore, WFAU plays a key role in the Gaia DPAC collaboration, currently contributing ~ 3 FTEs to critical aspects of the core processing in the ground segment for Gaia data.

Key persons:

Dr. Nigel Hambly is WFAU's science archive architect, and is an experienced astronomer/developer. Both the SSA and the WSA archive systems were designed by Hambly [5, 6]. He has much experience in managing small research and development teams for both hardware and software processing and archiving systems. He is also a member of the Gaia Archive Preparation group within DPAC, as well as being a senior developer for the Gaia DPAC Coordination Unit 5.

Mr. Mike Read is a survey astronomer, survey database and systems user interface specialist with extensive scientific and technical experience. He is in charge of WFAU's survey dataset user interfaces along with survey production quality control, and also deals with integration of server-side VO infrastructure and client-side applications.

Mr. Stelios Voutsinas is an IT specialist who has joined WFAU fresh from a successful Masters study into server-side VO infrastructure. His latest work concerns a Web 2.0 interface to VO services, in particular employing TAP as a front-end to distributed query web services.

Mr. Mark Holliman is a VO infrastructure specialist and expert data centre systems administrator, with extensive experience in publishing WFAU's data holdings to the VO and in developing and maintaining WFAU's network of 10s of high-end processors and 100s of terabytes of online storage.

²⁰The University of Edinburgh is a charitable body, registered in Scotland, with registration number SC005336

²¹<http://surveys.roe.ac.uk/ssa/>

²²<http://surveys.roe.ac.uk/wsa/>

²³<http://surveys.roe.ac.uk/vsa/>

²⁴<http://surveys.roe.ac.uk/ges/>

²⁵<http://vo.roe.ac.uk/index.htm>

2.2.4 Universiteit Leiden (UL)

General description:

Leiden Observatory is the astronomy department at Leiden University. Research covers areas of observational and theoretical astrophysics, instrumentation, and laboratory astrophysics. Undergraduate courses are given at all levels. The graduate programme currently involves 70 PhD students. Leiden Observatory shares a building with the Lorentz centre, an international centre for workshops in the sciences which offers excellent facilities for hosting workshops, conferences, and schools. In particular, office space is available for bringing researchers together for extended periods.

Role:

UL will lead WP-200 and host one or two post-doctoral researchers who will carry out the tasks in WPs 220, 230, 250, and 260.

Previous experience:

WP-200 is focused on the scientific analysis of the requirements on the Gaia archive and translating those into precise requirements that can be used for the implementation of the ideas put forward by astronomers. Leiden Observatory provides an environment where astronomers can be consulted directly on their wishes for data archives. The many astronomical research areas represented at Leiden make it an especially rich environment in this respect. Brown has been involved in numerous aspects of translating the scientific requirements on the Gaia spacecraft and data processing systems into requirements that can be understood by engineers and IT specialists. In addition he has been involved for a number of years in the discussions on advanced requirements on the future Gaia archive and was a member of the Gaia Archive Preparations group.

Short profile of staff members involved:

WP-200 will be led by **Dr. Anthony Brown**, who is the chair of the Gaia Data Processing and Analysis Consortium Executive and a member of the Gaia Science Team. From 2006 to 2012 he was also the manager of one of the development units of coordination unit 5 of the DPAC. In the latter role he has led a team of astronomical software developers in Leiden and at various institutes in Italy. The development unit that was led by Brown is in charge of the pre-processing software for the photometric data processing pipeline within DPAC.

Brown's research interests include: the structure and evolution of open clusters and OB associations; using Gaia's data to understand the structure and origin of the Milky Way galaxy; the evolution of open clusters in the Milky Way; and use of large catalogues and data archives to constrain complex models. Brown has previously been a scientist in charge for the Marie-Curie RTN ELSA, supervising a PhD student and post-doc. The graduate student completed his PhD on the in-depth study of the radiation damage problem and its mitigation for the Gaia mission. The post-doc worked on the development of a detailed physical model of the spacecraft attitude for Gaia. This model is used within CU2 for the generation of realistic attitude disturbances in the telemetry simulations. Brown is currently scientist in charge within the GREAT ITN, supervising a PhD student, and also serves on the Steering Committee of the GREAT ESF Research Networking Programme.

2.2.5 Centre de Serveis Científics i Acadèmics de Catalunya (CESCA)

General description:

CESCA is a public consortium created in 1991 and recognized as a scientific and technical singular facility by the Ministry of Science and Innovation. The objective of CESCA is to manage infrastructures based on information and communication technologies, to provide services to universities and research institutes.

Role:

CESCA is the large scale facility of reference for this project: its main tasks will be the set up, configuration and management of the testbed infrastructure (WP600) of the project as well as contributing to the design of the archiving system by providing advice and knowledge on virtual environments and cloud infrastructures for WP300.

Key competences and facilities:

CESCA has a wide experience in managing complex HPC systems, storage resources and networks. CESCA's infrastructure includes large HPC systems with more than 1 800 cores, high availability cloud-ready clusters based on blade technology and virtualization for hosting web applications, database servers like Oracle, PostgreSQL and MySQL, and a centralized storage system providing multiple protocols (NFS, Fiber Channel and iSCSI). CESCA also manages the *Anella Científica* (Scientific Ring), a flexible and reliable infrastructure that connects almost a hundred institutions involved in the world of R&D and innovation. It is also connected to RedIRIS and, through it, to GEANT2. CESCA's infrastructure (for instance storage resources) is reachable through broadband networks.

CESCA participates in national and international projects and currently, for instance, is member of Spanish NGI (National Grid Initiative), of the DPAC consortium of the Gaia Project, of MEGHA RedIRIS project group for promoting cloud services.

Key persons:

Dr. Joan Cambras has a degree in mathematics from the University of Barcelona (UB) with specialization in applied mathematics, and a postgraduate degree in 'Fundamentals and Applications of Artificial Intelligence' by the Universitat Politècnica de Catalunya (UPC). At CESCA since 1991: currently he is manager of the Applications and Projects Department.

Dr. Alfred Gil has a degree in Chemistry by the UB and a PhD in Chemistry by the Universitat Rovira i Virgili (2007). He was awarded a Marie Curie research fellowship (2001) at CNRS in Lyon (Laboratoire de Chimie). **Dr. David Tur** has a degree in Physics from the UB and a PhD in Nanostructures and Nanotechnologies from the University of Milano Bicocca. Both are HPC Support Scientist.

Mr. Ricard de la Vega has a Bachelor in Computing Engineering by the Universitat Oberta de Catalunya (UOC) and a Bachelor in Software Technical Engineering at the UPC. He has been working at CESCA since 2003. Currently, he is the head of the Portals and Repositories servicer.

Ms. Teresa Via holds a Master Degree in Communications & Marketing from the UOC and a Degree in Journalism from the Universitat Autònoma de Barcelona (UAB). At CESCA since 1999, she is currently the the Marketing and Communication manager.

Mr. Gorka Roldan has a Telecommunications Engineering Degree by the UPC. He has been working as Project Coordinator at CESCA since 2007.

2.2.6 Istituto Nazionale di Astrofisica (INAF)

General description:

Organization name: Istituto Nazionale di Astrofisica, with participating institutes: INAF-Osservatorio Astronomico di Roma (OAR) with ASI-Science Data Center (ASDC); INAF-Osservatorio Astronomico di Torino (OATo); INAF-Osservatorio Astronomico di Trieste (OATs); INAF is the main Italian research institution in the field of Astronomy and Astrophysics. INAF comprises 19 observatories and institutes spread across the country and employs more than 1200 people, about 500 of which are research astronomers. INAF promotes, realizes and coordinates, also within programs of the European Union and international bodies, research activities in many astronomical fields. INAF scientific research activity is divided into five themes: Galaxies and Cosmology; Stars, stellar populations and interstellar matter; The Sun and the solar system; Relativistic astrophysics and astroparticles; and Advanced technologies and instrumentation. In particular INAF is active in observational cosmology, Galaxy structure studies, stellar evolution, star formation, young stellar objects and interstellar matter, stellar population and the evolution of the Galaxy, and supernovae and their remnants.

Previous experience:

INAF groups at OAR, OATo and OATs are deeply involved in data processing and analysis for both space- and ground-based experiments. ASDC is a multi-mission science data centre cooperating with INAF, INFN, ESA and NASA. A specific agreement between ASI and INAF is on-going to manage the scientific activities and scientific personnel in ASDC. In ASDC there is a large expertise in the development of web based tools for the multi-wavelength data analysis and for the interactive on-line analysis of the data stored in the ASDC multi-mission archive. INAF coordinates a national Virtual Observatory initiative (VObs.it). It participates in the Euro-VO initiative and plays an active role in the projects funded by the EU Framework Programmes: VO-TECH, EuroVO-DCA (in which it led WP5: 'Coordination with computational grid projects'), EuroVO-AIDA (in which it led WP5: 'Service Activities for higher education and outreach') and played an active role in other work-packages, and EuroVO-ICE (in which it leads WP5: 'Coordination of education and outreach activities'). VObs.it is a member of the world-wide International Virtual Observatory Alliance (IVOA) and is represented on its Executive Committee.

Role and Short profile of staff members involved:

INAF-OATo will coordinate WP-240 and INAF-OATs will participate in WP-340; the participation will be coordinated by Richard Smart who is an astronomer at the INAF-OATo since 1994 studying astrometric reduction techniques, low mass stars, brown dwarfs, galaxy kinematics and the production of large catalogues. In this framework, he and colleagues from OATo and ASDC are cooperating in the development of database mining techniques. Presently he is responsible for the production of the Initial Gaia Source List and the Attitude Star Catalogue for CU3 in the DPAC as well as deputy for the Ground Based Optical Tracking working group.

Key Persons:

Dr. Richard Smart, INAF-OATo, INAF coordinator for the project;

Dr. Riccardo Smareglia, INAF-OATs, head of Italian Center for Astronomical Archive (IA2);

Dr. Marco Molinaro, INAF-OATs, expert in the use and the development of the VObs;

Dr. Alessandro Spagna, INAF-OATo, expert in cross-matching of catalogues containing sources with heterogeneous astrometric and photometric properties.

2.2.7 Consejo Superior de Investigaciones Científicas (CSIC)

General description:

The Agencia Estatal Consejo Superior de Investigaciones Científicas (CSIC) is the largest public multidisciplinary research organisation in Spain. CSIC collaborates with national and International universities, public RTD organisations, SMEs, companies and other institutions of a scientific/technological nature. It has a staff of more than 13,500 employees, among these about 3,500 are permanent researchers and more than 4,000 are pre- and post-doctoral researchers.

The CSIC has 135 institutes or centres distributed throughout Spain, including 51 Joint Research Units with universities or other research institutions. There is also a delegation in Brussels. It has considerable experience in both participating and managing RTD projects and training of research personnel. Under the 6th Framework Programme CSIC has signed 404 actions (of which 37 coordinated by CSIC). CSIC has been the 5th organisation in Europe in project execution and funding in the 6th Framework Programme. Under the 7th Framework Programme CSIC has signed as of today more than 440 projects (including 42 coordinated by CSIC as well as 22 ERC)".

Key competences and facilities:

The CSIC/CAB is also the major actor of the VO development at national level. The CSIC/CAB leads the Spanish Virtual Observatory Project, an initiative funded by the Ministry of Science and Innovation to enhance the participation of the Spanish institutes in the VO framework through technical (VO compliance of astronomical archives, implementation of theoretical services) and scientific (VO science cases) developments, provision of added-value services (data mining and VO tools) and organization of meetings to explore ways of collaboration and exploit synergies among teams.

Key persons:

Dr. Enrique Solano received his astronomy PhD from Universidad Complutense de Madrid in 1994. In 1990 he joined the International Ultraviolet Explorer (IUE) project working in the Image Processing Department. In 1995 he took up a position as IUE Resident Astronomer until the end of the project (February 1998). In September 1998 he started to work at LAEFF for the ESA-INTA INES Transfer Programme at the end of which (June 2000) INTA took full responsibility for the INES Archive. Since January 2002, he has been leading the INTA Astronomical Data Centre.

Dr. Solano is the Principal Investigator of the Spanish Virtual Observatory project (approved in June 2004 and renewed in 2011) and responsible for the VO initiatives in the ASTROMADRID (funded by the Madrid Regional Government) and CONSOLIDER-GTC (an excellence project funded by the Ministry of Science and Innovation) projects. He is a member of the IVOA Executive Committee and the IVOA Committee on Science Priorities and co-chair of the IVOA Application Working Group. He has been the Spanish representative in the Board of the Euro-VO DCA and EuroVO-AIDA projects. He is also co-author of a significant number of VO-science refereed papers.

Support personnel:

Dr. Luis Sarro: Dr. Luis M. Sarro obtained his degree in Physics (sp. Astrophysics, Universidad Complutense de Madrid, Spain) in 1993 and his PhD. in 1998 (Universidad Autónoma de Madrid, Spain). He is a full-time lecturer at UNED University (Spain) in the Artificial Intelligence Dpt. since 2011, and has focused his research interests in the Data Mining of large astronomical Databases, including the CoRoT database of stellar time series and the preparation for the Gaia mission through the Data Processing and Analysis Consortium.

Scientific profile: Data Mining in astronomical databases, Supervised and unsupervised classification of variable stars, Outlier detection techniques, Non parametric density estimation, Explosive events in the solar Transition Region, Bayesian parametric inference, MCMC methods, efficient methods for posterior sampling, Pre-Main Sequence disks, Ultra cool dwarves.

2.2.8 Université de Genève (UNIGE) - Observatoire de Genève

General description:

Geneva Observatory is the Astronomy Department of Geneva University. Activities of the Observatory and accompanying research centres cover photometric and spectroscopic observations, variable stars, exoplanet searches, theory of stellar evolution and Galactic dynamics, observational cosmology, high energy astrophysics, instrument building, and data management of Space missions. Undergraduate courses are given at all levels. The graduate programme involves PhD students related to photometry, variable stars, and activities relative to data reduction in astronomy.

The Integral Science Data Centre, an institute attached to Geneva Observatory, has been in charge of reduction of the data transmitted by the INTEGRAL satellite (space gamma ray telescope) of the European Space Agency (ESA) and their subsequent distribution to the users. The activities extend to other space missions, such as COROT (seismology and planetary transits), Gaia (astrometry and photometry), PLANCK (cosmology).

The involvement of Geneva in Gaia has been within the Data Processing and Analysis consortium. The coordination unit 7 (variability processing) is managed from Geneva and the corresponding Data Processing Centre is based there.

Key persons:

The manager of the Geneva node in this proposal, **Dr. Laurent Eyer**, obtained his PhD in 1998 at the Observatoire de Genève, Switzerland, working with Michel Grenon on the analysis of the variability data from Hipparcos, resulting in the Hipparcos Variability Annex and Light Curves (Volume 12 of the Hipparcos Catalogue). During his first postdoc, at the Catholic University of Leuven, he worked on exploiting Hipparcos findings on certain variable stars. A subsequent post-doctoral position at Princeton was focused on the analysis of the OGLE data, providing experience in the handling of massive data sets of direct relevance to Gaia.

Following the award of a major grant by the Swiss Government in May 2003 for involvement in the Gaia mission, Laurent Eyer returned to Geneva, where he led the Variability Working Group (2003-2005), whose goals were to develop the algorithms necessary for the automatic detection and classification of variable stars for Gaia. As a member of the DACC (2006) Eyer was involved in preparations leading to the formation of the Gaia Data Processing and Analysis Consortium. Within the DPAC he is manager of CU7, Variability Processing, and he is a member of the DPAC.

2.2.9 Université Libre de Bruxelles (ULB)

General description:

The Institute of Astronomy and Astrophysics (IAA) is part of the Department of Physics of the Université Libre de Bruxelles. It is in charge of some aspects of the validation of the binary star solutions provided by DPAC. Binaries represent substantial activity in the IAA, ranging from ground-based observations (with the Hermes spectrograph) to theoretical modelling of their evolution, from light curve fitting to astrometric and spectroscopic orbit determination. IAA is also the curator of the IAU database of spectroscopic orbits. The derivation of spectro-astrometric orbits and their assessment has been a regular activity of some IAA members for the past ten years, thus developing some expertise in such scientific validations.

Key persons:

Whoever will be hired for this job will have the strong support of **Dr. Dimitri Pourbaix** who is the leader of the DPAC coordination unit CU4 in charge of the processing of the double and multiple stars observed by Gaia. Pourbaix is a member of the DPAC.

2.2.10 Fundação da Faculdade de Ciências da Universidade de Lisboa (FFCUL)

The Fundação da Faculdade de Ciências da Universidade de Lisboa (FFCUL) is a private non-profit organization created in 1993 in order to manage financial and administratively its R&D projects. FCUL acts as a third party in those R&D projects based on a scientific agreement in force since FFCUL's establishment.

The Laboratory for Systems, Instrumentation and Modeling in Science and Technology for Space and the Environment (SIM) is a research unit of the Faculty of Science of the University of Lisbon. SIM is a key Centre for space science in Portugal with strong activities in astronomical instrumentation and astrophysics. The SIM laboratory has a national scope with offices in Lisbon and Porto. Our instrumentation activities include leading the Portuguese participation in the Gaia DPAC (which includes 5 Centres) and in the ESO/GRAVITY experiment; we participate in the ESO/ESPRESSO spectrograph. R&D in image reconstruction, adaptive optics and astrophotonics is also undertaken.

Dr. André Moitinho de Almeida

Relevant astronomical research: Galactic Structure, large scale star formation, star clusters, brown dwarfs.

Computing: coordinates the Portuguese participation in the Gaia DPAC. Responsible for assessing biases and works in the data mining infrastructure of the the Gaia variability processing pipeline. Participation in the development of a knowledge discovery tool (including visualization) for large data sets (ESA contract - delivered). On-going ESA contract for assessing visualization needs and available facilities for visualization in climate change research and astrophysics.

Instrumentation: Management of Work Packages (WPs) for a) Design and implementation of the optomechanics of the Coudé Train for ESPRESSO, an ultra stable high resolution optical echelle spectrograph for the ESO VTL. Phase A approved by ESO. b) Design and implementation of the cryogenic system and cold mechanics of NAHUAL - an ultra stable high resolution infrared echelle spectrograph for the 10 meter GTC telescope.

Consulting: co-chair of the ESO OPC (telescope time allocation committee) and chair of one of the Stellar Astrophysics panels (2004-2006). Member of the Calar Alto telescope time allocation committee (2009-2010). Member of the ESA Astronomy Working Group (2008-2010). Chair of the Science review panel for the ESA PLATO mission (December 2010).

Dr. Alberto Krone-Martins

Relevant astronomical research: star clusters, brown dwarfs, data analysis, galaxies.

Computing: responsible for the design and development of the pipeline for analysis of extended objects for the Gaia DPAC. Designed and implemented automatic tools for analyzing membership of star clusters and associations based on astrometric data and genetic algorithms, kernel density estimators, models, convergence point, etc. On going design of an automatic an unsupervised tool for analyzing membership on star clusters with photometric data. Implemented the code for simulation of galaxies in the Gaia Universe Model. Designed and developed an ultra high energy cosmic rays propagation code with hadronic interactions in a parallel environment. On-going ESA contract for assessing visualization needs and available facilities for visualization in climate change research and astrophysics.

Consulting: member of the BraVO - Brazilian Virtual Observatory - Board (2010-). Referee ad-hoc for the SOAR 4m Telescope TAC (2011-). Referee for the Journal of Computational Interdisciplinary Sciences (2011-). Member of the committee responsible for the specification, acquisition and deployment of a large computer (>2300 cores) for the IAG-USP (2009-2011). Vice-president of the CASP - Sao Paulo Astronomy Club (2007-2011).

2.2.11 University of Bristol (UBR)

General description:

The Astrophysics Group of the University of Bristol is a part of the HH Wills Physics Laboratory. The research of the Group is focussed on (a) the formation and evolution of galaxies and clusters of galaxies, (b) the mechanisms underlying active galaxies and their effects on their environments, and (c) the physics of the formation of planets. Theoretical, computational, and observational work is done in the Group, using a wide range of space-based and ground-based telescopes, and local and national supercomputing facilities. The Group is responsible for Physics with Astrophysics BSc and MSc degrees in the University of Bristol, and also takes national and international students for its MSc and PhD programmes. Researchers in the Group have close contact with other physicists working within the HH Wills Physics Laboratory, and also interact with colleagues in the Schools of Earth Sciences, Computer Science, and Mathematics.

Role:

The TOPCAT application will be extended to provide seamless access to Gaia catalogue data.

Key competences and facilities:

TOPCAT is an interactive graphical application for exploration, analysis and manipulation of tabular data, especially source catalogues, which works well with moderately large data sets (up to a few million rows and a few hundred columns). It combines ease of use with flexible and powerful facilities for visualisation, crossmatching, calculations and other general purpose table operations. Much of its development has been within Virtual Observatory (VO) projects; it presents a user-friendly interface to many VO services and interfaces, and is recognised within the VO community as one of the core VO applications. It already offers a number of service-specific load dialogues (e.g. VizieR, Millennium Simulation), and a Gaia option would be added alongside these. Additionally, investigations will be made of whether the existing practical limits on dataset size can be increased.

TOPCAT is in regular use by certainly hundreds and perhaps thousands of astronomers worldwide, and has users in 24 of the 27 EU member states. Providing direct access to Gaia data from this tool would be a highly effective way to facilitate its exploitation.

Key persons:

Dr Mark Taylor is a Research Fellow in the Astrophysics Group of the School of Physics at the University of Bristol, who has worked on various astronomy software items since 1998. He has been responsible for almost all aspects of the development of TOPCAT and its related products STILTS and STIL, including design, coding, testing, documentation, user support, and ongoing identification and prioritisation of requirements in collaboration with the astronomy community. He is also the lead author of the SAMP interoperability protocol, and the Chair of the Applications Working Group of the International Virtual Observatory Alliance. This work has been carried out within a number of UK and European projects, some VO-specific, with other funding from the UK's Science and Technology Facilities Council and from Microsoft Research.

2.2.12 University of Cambridge - Institute of Astronomy (IoA)

The University of Cambridge includes a world class Astronomy research cluster including the IoA. With some 70 research staff, the IoA conducts high impact research across a broad range of topics, from cosmology, active galaxies, Galactic Structure, and star and planet formation. There are some 50 PhD students involved through the postgraduate programmes at the IoA. The IoA hosts the Kavli Institute for Cosmology at Cambridge, and interacts at a close level with the Astrophysics group in the Cavendish Laboratory.

Role: The IoA will participate in WP-100, WP-200 and WP-600. The main focus of IoA activity will be in developing GENIUS systems to increase the scientific impact of the science alerts which will be generated by CU5 from the Cambridge Data Processing Centre during the Gaia operations phase. In particular, preparatory testbed development work in the context of WP-600 to develop the cross federation capability of alerts generated compared to potential existing Gaia knowledge of the alert object. WP-200 activity will focus developing a community refresh of the requirements for data access via a requirements gathering exercise through the GREAT community.

Key competences and facilities: The IoA leads the Photometric processing (CU5) in the Gaia DPAC, and is responsible for the Gaia Photometric Processing Centre. It operates the Cambridge Astronomy Survey Unit (CASU) which develops major science pipelines used to generate the science data products for priority missions such as ESO's VISTA infrared survey telescope (data rates >100TB/year). It will lead the Exoplanet processing for the proposed ESA M3 PLATO planet finding mission. Cambridge is world renowned for both theoretical and applied computer science, with the Computer Lab and the Cambridge eScience Centre. It hosts the High Performance Compute Facility D amongst the most powerful civilian supercomputers in the UK.

Dr. Nicholas Walton, member of the ESA Gaia Science Team, chairs the GREAT (Gaia Research for European Astronomy Training) research network <http://www.great-esf.eu>, the ESF supported GREAT Research Network Programme (<http://www.esf.org/great>) and Coordinator of the EC FP7 Initial Training Network (GREAT-ITN) (<http://www.great-itn.eu>) that brings together 32 European research institutes training a large cohort of Early Stage Researchers focussed on projects where Gaia will have a major impact. Walton supervises a GREAT-ITN PhD student developing improved techniques to classify SNe to be detected by Gaia, and to utilise these as a significant new population of low redshift SNe to probe cosmological models. He is Technical coordinator of the Virtual Atomic and Molecular Data Centre and follow-on SUPP@VAMDC initiative (<http://www.vamdc.eu>). He has been actively involved in the development of the Virtual Observatory both as Euro-VO Technology Centre Project Scientist and secretary of the International Virtual Observatory Alliance. He was co-Chair of the Open Grid Forum's astronomical applications Research Group (2004–2009). **Dr. Floor Van Leeuwen** is the manager for the photometric aspects of the Gaia satellite data processing in Europe, and manager of the UK Gaia data processing activities. He is also co-I on the Planck mission, and has a long running involvement with the Hipparcos mission, for which he did the final re-reduction, resulting in very significant overall improvements in the quality of the astrometric data, as published in November 2007. **Dr. Simon Hodgkin**, leads the Cambridge development of the science alerts system. He is organising the ground based validation campaign of Gaia alerts. He leads the UKIRT Wide field Transit Survey. **Dr. Guy Rixon**, is active within the Gaia DPAC responsible for the quality assurance engineering of the photometric processing system and the science alerts system. He currently leads the Service Deployment activities within VAMDC. He was Chair of the International Virtual Observatory Alliance working group on Grid and Web Services. In this role he had led and developed the VO systems in interfacing to the Grid and Web services realm. He has played a major role in the development of the authentication and authorisation framework for the VO, and the VOspace storage system. Rixon is an acknowledged leader in grid and web service technologies.

2.2.13 Kyoto University (KU)

General description:

The Department of Physics of the Kyoto University (KU) with some 70 researchers covers all areas of theoretical and experimental physics and astrophysics. The Theoretical Astrophysics Group in the department of Physics has the role of coordination for Nano-JASMINE AGIS/ DPAC collaboration in Japan. The KU has a Master's program in Astrophysics, Particle Physics and Cosmology and the current PhD program involves 40 PhD students.

Role:

The KU will act as the coordinator of the Japanese collaboration in GENIUS and will also participate in WP 220. It will be the host of specialized workshops on utilization of astrometric data.

Key competences and facilities:

Membership of JASMINE Working Group (Y. Yamada); membership of Gaia DPAC CU3 as external collaborator; lead for Nano-JASMINE data simulations; responsibility for Nano-JASMINE Initial Data Treatment in collaboration with ESAC, DPAC, Shiga University (SU), Tokyo University of Marine Science and Technology (TUMST), and National Astronomical Observatory of Japan (NAOJ); Nano-JASMINE numerical modeling for analysis in collaboration with IBM Tokyo Research (TRL); Galactic structure in collaboration with NAOJ; kinematics and dynamics of Milky way galaxy in collaboration with NAOJ; scientific output from Nano-JASMINE data in collaboration with Niigata University (NU) and NAOJ; Nano-JASMINE Hipparcos catalogue combination in collaboration with Lund Observatory and ESAC.

Key persons:

Dr. Yoshiyuki Yamada, assistant professor at the Department of Physics, KU, is responsible for Nano-JASMINE data analysis. He is member of DPAC CU3 as an external collaborator and responsible for AGIS/Nano-JASMINE collaboration in data analysis. He was the manager of Nano-JASMINE data utilization, Coordination Funds for Promoting Space Utilization (2009 - 2011, about 0.5M Euro).

Prof. Naoteru Gouda, professor at National Astronomical Observatory of Japan (NAOJ), is team leader of the JASMINE project. He is responsible for the interface to the other astrometry group in Japan (VERA, VLBI astrometry).

Dr. Ryoichi Nishi, associate professor of Niigata University is responsible for Nano-JASMINE science, and so for tracking the user needs.

Prof. Shunsuke Hozumi of Shiga University and **Prof. Satoshi Yoshioka** of Tokyo University of Marine Science and Technology are also members of Nano-JASMINE data analysis team. They are also members of the CU3 collaboration. The Nano-JASMINE reference catalogue data base is managed by S. Yoshioka.

2.3 Consortium as a whole

The core of the consortium, responsible for the core development of GENIUS (see Fig. 2.2) is formed by the four main partners, whose managers constitute the executive board described in Sec. 2.1.

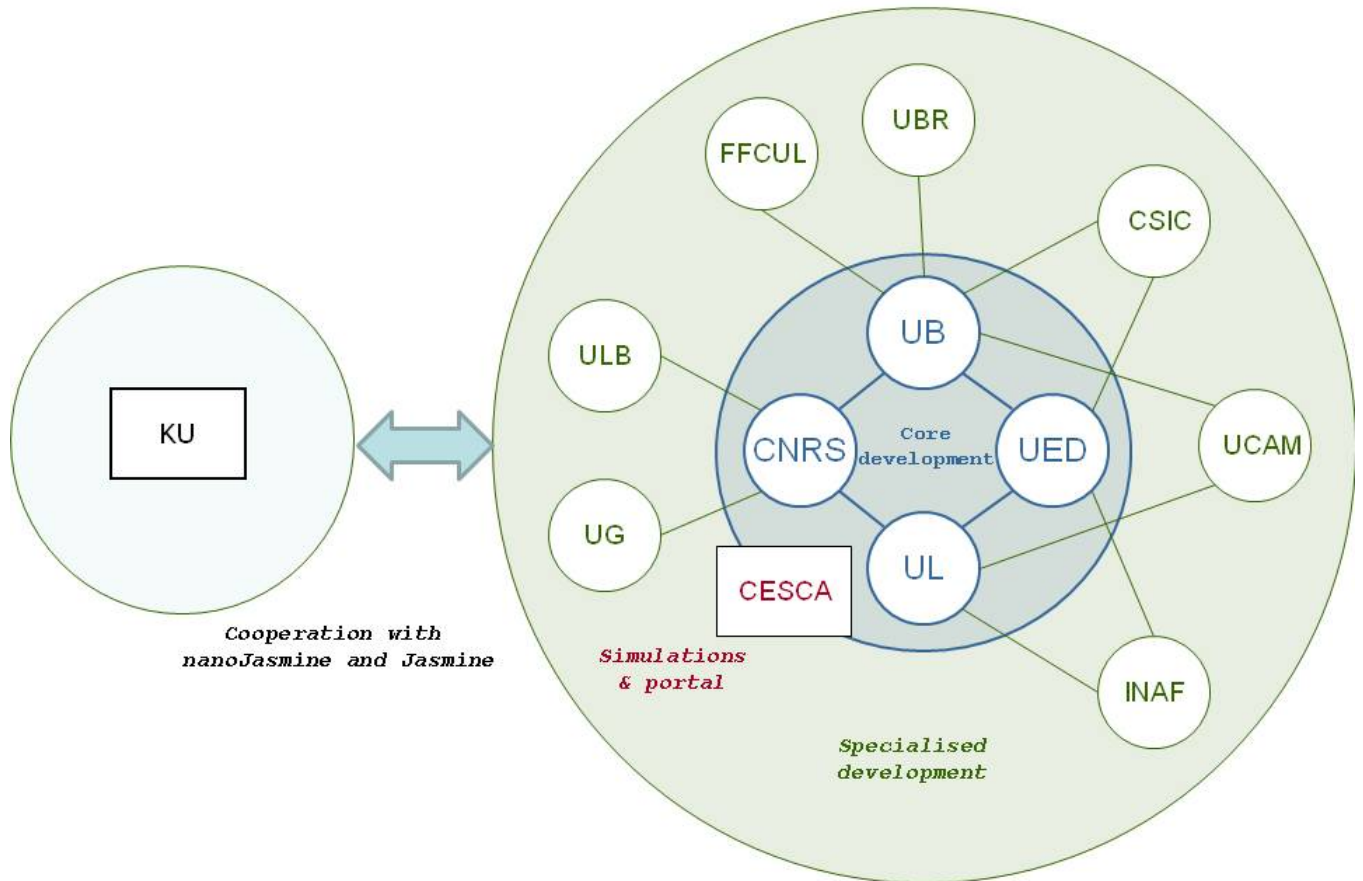


Figure 2.2: The GENIUS collaboration

The four main partners are institutes with a strong involvement in the Gaia data processing, and therefore a good knowledge of the mission and the characteristics of the data it will produce. Furthermore, its managers are all members of the Gaia Archive Preparation group (GAP), and are therefore fully involved in the already ongoing tasks of definition of the Gaia archive in cooperation with ESA and the DPAC.

The core partners will be closely coordinated through the executive board, and will ensure that the GENIUS objectives are fulfilled. Each one of them contributes specific knowledge to the consortium that is relevant to the GENIUS tasks:

Universiteit Leiden: the group at the UL covers many astronomical research areas and provides an especially rich environment, suitable for the task of defining the user needs of the Gaia catalogue in WP200. A. Brown, its coordinator, has been involved during the preparation and implementation of the Gaia mission in numerous aspects of translating the scientific requirements on the Gaia spacecraft and data processing systems into requirements that can be

understood by engineers and IT specialists. In addition he has been involved for a number of years in the discussions on advanced requirements on the future Gaia archive, and therefore provides very relevant expertise for the assigned task

University of Edinburgh: the group at UEDIN has a strong track record in the curation and dissemination of large-scale survey data sets consisting of tens of billions of rows of catalogue data occupying tens of terabytes. It can thus provide the needed expertise for the task of contributing to the design of the Gaia archive system in WP300. In addition, UEDIN has been participating in the Gaia DPAC since 2007.

Universitat de Barcelona: the group at UB will provide overall GENIUS coordination and management of WP400 for the development of exploitation tools. The group at UB has been in charge for more than a decade of the development of the Gaia simulator and has managed the Coordination Unit 2 of the DPAC since its creation. The UB can therefore provide a sound management experience of a pan-European collaboration (the CU2 having more than 80 members distributed all around Europe) a very deep knowledge of the mission and the data it will produce. This experience in managing the production and handling of catalogue-like data is an important asset for GENIUS.

CNRS: the centres at the CNRS participating in GENIUS have been involved from the very beginning in the definition of the Gaia mission, the simulation of its scientific content and the data processing within DPAC. Furthermore, the CNRS group at the Paris observatory was in charge of the validation of the Hipparcos catalogue, and therefore brings an invaluable experience in this field for WP500, devoted to catalogue validation.

The four main nodes provide the basic structure of GENIUS, with personnel (hired through GENIUS funding or permanent staff) that will work full time on the development of the Gaia archive and its tools and will thus ensure the continuity of the development during its duration. The node managers will permanently track the status of the development in their work packages and coordinate the overall status through the executive board.

On the other hand, the Centre de Serveis Científics i Acadèmics de Catalunya (CESCA) node will provide hardware for the running of simulations and support for the development of the community portal. The CESCA is a supercomputing centre providing services to universities and research institutes, recognized as a singular science and technology infrastructure (ICTS) by the Spanish Ministry of Science and Innovation since 2000, and has been involved in Gaia since 2004. It will provide and maintain a permanent computing infrastructure for the running of simulations. Furthermore, it already hosts several data repositories²⁶ open to the public and is directly connected to the European backbone research network GEANT²⁷, providing expertise and a top-quality connectivity that will allow the development and implementation of the community portal.

The rest of the partners will provide expertise in specific areas complementing the above described contributions. Each node has been chosen by its specialised knowledge in an specific task relevant for the GENIUS objectives and their participation may range from time-concentrated tasks to part-time work for the duration of GENIUS.

Istituto Nazionale di Astrofisica Italy: INAF groups at OAR, OATo and OATs have been deeply involved in data processing and analysis from both space- and ground- based experiments. ASDC is a multi-mission science data centre cooperating with INAF, INFN, ESA and

²⁶<http://www.cesca.es/en/cesca/areas-of-activity/description>

²⁷<http://www.geant.net/>

NASA. In ASDC there is much expertise in the development of web based tools for the multi wavelength data analysis and for the interactive on-line analysis of the data stored in the ASDC multi-mission archive. INAF coordinates a national Virtual Observatory initiative (VObs.it). It participates in the Euro-VO initiative and plays an active role in the projects funded by the EU Framework Programmes VO-TECH, EuroVO-DCA, EuroVO-AIDA, and EuroVO-ICE. VObs.it is a member of the worldwide International Virtual Observatory Alliance (IVOA) and is represented in its Executive Committee.

Université de Genève - Observatoire de Genève: the group at the UG is in charge of the variability reduction in DPAC and will provide the relevant expertise for the validation of variability data in the Gaia catalogue in WP500.

Consejo Superior de Investigaciones Científicas: the CSIC/CAB is a major actor of the VO development, leading the Spanish Virtual Observatory Project and working on the VO framework through technical and scientific developments and provision of added-value services (data mining and VO tools). This VO expertise will contribute to WP300 and WP400.

University of Bristol: the UBRI develops a very popular, VO compliant tool for data visualization, TOPCAT²⁸. This tool has already been tailored for integrated use of several large astronomical archives and GENIUS will provide its adaptation to the Gaia catalogue, ensuring an easy entry point for the catalogue visualization.

Université Libre de Bruxelles: the group at the ULB is in charge of the treatment of multiple systems in DPAC and will provide the relevant expertise for the validation of multiple star data in the Gaia catalogue in WP500.

Fundação da Faculdade de Ciências da Universidade de Lisboa: the FFCUL group provides expertise in visualization and, among other examples, has been awarded an ESA contract (ITT) for performing an assessment study of visualization needs and solutions in Climate research, disaster management and Astrophysics. They will contribute this expertise to the visualization tasks in WP400.

University of Cambridge - Institute of Astronomy: the IoA group provides expertise in the Gaia Science Alerts System, that is being developed at the institute. Its contribution will be partly in WP200 for user requirement definition and partly in WP600 for the implementation of a Science Alerts testbed.

Kyoto University: the KU group is involved in the development of the nanoJASMINE and JASMINE missions, and will provide the coordination for the interaction with the wider Japanese community working in these missions.

2.3.1 Third parties

Fundació Bosch i Gimpera (FBG) (Bosch i Gimpera Foundation²⁹) is a non-profit foundation and is considered a third party of Universitat de Barcelona (UB). FBG was created by UB in 1983, and is governed by an long lasting agreement signed between UB and FBG.

According to FP7 Guide to Financial Issues ((v. 28/02/2011, page 45), FBG fall in the Special Cases 1) Foundations, spin-off companies, etc., created in order to manage the administrative tasks

²⁸<http://www.star.bris.ac.uk/~mbt/topcat/>

²⁹<http://www.fbg.ub.es>

of the beneficiary . Thus, FBG handles the financial and administrative aspects of the beneficiary UB involvement in research projects, including all issues relating to the employment and payment of additional personnel, purchase of equipment and consumables, etc. The personnel hired for the project by the foundation works on the premises of the University (beneficiary) and under its responsibility, as the foundation does not have the resources to carry out the work.

This will imply during the project that:

- UB remains the beneficiary of the project and is thus responsible for the scientific and financial aspects of the project.
- UB and FBG together complete the financial forms (Form C) needed for the management report when requested. FBG will prepare this Form C, which is signed by the authorised representative of the University of Barcelona.
- UB and FBG will prepare and submit Certificate on Financial Statements when applicable following the rules set in the EC Financial guidelines.

In the case when UB is the coordinating institution, Special clause 38 applies to the grant agreement.

FBG does not carry out part of the work in the project, and no resources are allocated for FBG in the project, thus FBG should not be covered by Special clause 10 in the ECGA .

2.3.2 Sub-contracting

GENIUS includes a subcontractor for very specific activities and representing a minor fraction of the budget:

Audit: following the FP7 rules some of the partners will have to carry out audits of their expenses. The budget for subcontracting an external auditing has been included in these cases.

Intersystems: this US-based company ³⁰ develops the database engine *Intersystems Caché* that is currently used by the ESAC team for the development of the Gaia prototype archive system. We plan to acquire a license of this DB engine for our own work in the development of the archive system to match the current ESAC developments ³¹ and to hire Intersystems to provide extended support (on top of the one acquired with the licenses) on its product in the form of courses and consultancy. A letter of interest and support from this company has been sent to the GENIUS coordinator.

Intellectual property: we envisage the need for specialised consultancy regarding intellectual property issues (see Sec. 3.2). This may be the hiring of legal advice or support for technology transfer.

Web design: some aspects of the implementation of the community portal are not covered by the in-house expertise of CESCO, specially the ones related to graphical design. A budget for subcontracting these tasks has been included in the CESCO budget.

³⁰<http://www.intersystems.com/aboutus/index.html>

³¹although it can well happen during the development that in agreement with them we switch the choice of DB engine, as has already happened in the past

2.3.3 Additional partners

In addition to the listed formal partners we have identified a possible additional partner interested in the possibility of contributing to the GENIUS tasks in the future:

University of Vienna: the research group at the Department of Astronomy at the University of Vienna led by Prof. Joao Alves ³² has expressed its interest in a possible cooperation with GENIUS regarding the development of visualization tools for Gaia. We intend to explore this possibility to be channelled through the visualization developments planned in WP400, where the Vienna group could contribute its expertise. A letter of interest is included in the appendix of this proposal.

2.4 Resources to be committed

The discussion in this section is divided in two parts: a global overview of GENIUS, regarding budget and staff effort, followed by specific analysis for the partners.

2.4.1 Overview: budget

We have summarised in the following table the budget figures contained in section A3 to provide an overview of the global funding distribution in GENIUS:

Partner	Personnel	Travel	Other	Subtotal	Overheads	Subcont.	Budget	Requested
UB	300,431	20,000	91,000	414,431	248,659	16,000	679,089	558,534
CNRS	357,546	13,000	10,000	380,546	228,327	–	608,873	448,664
UEDIN	261,265	16,153	–	277,418	198,139	–	475,557	360,893
UL	256,655	35,000	–	291,655	225,482	1,500	518,637	395,295
CESCA	84,000	–	–	84,000	50,400	12,000	146,400	109,800
INAF	73,400	5,000	5,000	83,400	50,040	–	133,440	100,080
CSIC	84,802	–	5,000	89,802	44,097	–	133,899	100,424
UNIGE	41,000	667	–	41,667	24,000	–	66,667	50,000
ULB	41,000	650	–	41,650	24,990	–	66,640	49,980
FFCUL	61,667	10,000	–	71,667	43,000	–	114,667	86,000
UBR	21,623	–	3,362	24,985	14,991	–	39,976	29,982
UCAM	115,276	–	7,000	122,276	73,364	–	195,640	153,891
KU	–	41,600	–	41,600	24,960	–	66,560	49,920
Total	1,701,665	142,070	121,362	1,965,097	1,244,363	29,500	3,246,046	2,493,466

We will focus the overall discussion on the column named **Subtotal**. This column sums the budget for personnel, travel and other expenses for each partner, and therefore represents the actual funding available for the GENIUS tasks, leaving the overheads of the institutions aside. The use of the subcontracting funding will be discussed in 2.4.3. Taking this “effective” funding one can see that:

- About 70% corresponds to the four main partners. This is consistent with the GENIUS structure defined in Sec. 2.3, where the main partners take the responsibility of the core developments. Furthermore, the four main partners share roughly similar percentages of the funding (14%-21%) consistent with the similar responsibilities they have in GENIUS

³²<http://www.joaoalves.org/>

- The remaining 30% corresponds to the budget of the specialised partners, also consistent with the GENIUS structure defined in Sec. 2.3. They provide specific expertise for the tasks described for them in 1.3.3 and range from 6% to 1% depending on the extension of these tasks.
- The discussion of the distribution of the costs between the different concepts is left for the next sections, given the different roles and costing approaches of each partner.

2.4.2 Overview: personnel

The overall distribution of effort in GENIUS was first based on the criteria that the core partners should provide the personnel for the global and work package managing and also the personnel continuity for management and technical developments. On the other hand the CESCO, providing hardware for simulations and support for the community portal, should similarly provide personnel under this criteria. Finally, the specialised partners should provide in some cases part-time personnel, for tasks distributed along the duration of GENIUS, or on some other cases personnel devoted to GENIUS for specific periods of time. The resulting distribution of personnel can be obtained from the summary of staff effort table in Sec. 1.3.3, resulting in:

- About 62% of the staff effort corresponds to the four main GENIUS partners. This is again consistent with the GENIUS structure defined in Sec. 2.3, and reflects the responsibility for the core tasks assumed by them.
- The remaining 38% staff effort corresponds to the specialised partners. This amounts to about 1/3 of the effort, which is relatively large compared with the corresponding share of the funding; the reason for this is mostly the different labour costs in the countries of the partners and also the different degrees of specialisation of the provided personnel.

On the other hand, the same table provides also the distribution between work-packages:

- About 81% of the staff effort corresponds to WP200, WP300, WP400 and WP500, reflecting the fact that most of the GENIUS development tasks are concentrated in these RTD work packages. The figure includes not only the development effort but also the management effort of the work packages.
- About 10% of the staff effort is in WP600. This work package includes the provision of simulated catalogue data from the already available simulator GOG using the CESCO hardware, the science alerts testbed and the community portal development. These tasks are naturally less intensive and thus require the quoted smaller fraction of the staff effort.
- The remaining 9% of the staff effort corresponds to the overall management of GENIUS. As described in 2.4.3 it includes the time of the GENIUS coordinator and a part-time project manager. It amounts to a combined dedication of about one FTE, which is not unreasonable for an international cooperation involving 11 partners, organisation of meetings, coordination with the institutions and other consortia involved in Gaia and reporting to the EC and the external advisory board.

2.4.3 Universitat de Barcelona (UB)

The UB funding is divided in four sections:

Personnel: the personnel expenses are invested in three parts. First, the hiring of personnel for the work in WP400 (Sec. 1.3.3), mostly devoted to the hiring of an experienced software engineer. Second, the part time hiring of a project manager for the overall management of GENIUS, included in WP100. Finally, the third part corresponds to permanent UB personnel costs working on GENIUS, including the management of the project.

Travel: about 7,000€ per year have been devoted to travel and subsistence. This will cover the travel needs of the project coordinator, expected to be significant, the participation of UB personnel in the GENIUS meetings and the travel expenses of the external advisory board.

Other expenses: most part of the other expenses correspond to the acquisition of software licenses for Intersystems Cache database systems (see Sec. 2.3.2). The funding allocated has been calculated from the cost estimations provided by the company and amounts to about 78,000€. The rest of this funding is for minor general expenses.

Subcontracting: the remaining budget corresponds to subcontracting of specialised support, as described in Sec. 2.3.2, including the funding allocated to the auditing costs, as required by the EC rules,

2.4.4 Centre National de la Recherche Scientifique (CNRS)

CNRS will be the main provider of WP 500, led by Arenou, devoting 5% together on the management and administration of this Work Package for a total of about 2 months. Due to CNRS interest in VO related activities (such as VO Paris), a part of the CNRS involvement will also take part in the VO tasks of WP 300 and WP 400. Staff positions will provide the scientific guidance while a postdoctoral level on one hand and two engineer levels on the other will contribute to the definition and implementation of the validation tools. Travel funds are needed to cover the meetings within the CNRS partners or foreign (ULB, UG) nodes working in WP 500. Regarding small equipment (e.g. laptops) and consumables, as most CNRS positions are permanent positions for which CNRS already cover the expenses, the requested funding is needed only for hired people.

2.4.5 The University of Edinburgh (UEDIN)

UEDIN leads WP-300 with Hambly as manager at the level of 1 sm per 15 sm of managed staff effort (4 sm, i.e. $\sim 6\%$), consistent with our other partners. UEDIN has the advantage of hosting the Wide Field Astronomy Unit which holds a portfolio of development grants including those for Gaia DPAC, VO initiatives and general wide-field survey astronomy, enabling us to deploy fractions of existing expert staff in relevant development areas, shifting effort as required by the project work undertaken in a proactive way and as collaborative work demands. Our split of effort across the work packages assigns the greater part of our staff resource to server-side VO infrastructure and generic interface work (21 sm) as we perceive these as requiring the most effort. Data centre work is the next largest fraction (10 sm), and finally a smaller fraction (6 sm) to work on research into cloud-based solutions and data mining infrastructure. Given the collaborative nature of the GENIUS project, we request travel and subsistence at an appropriate level ($\sim 4\%$ of our total resource to cover European trips to the tune of ~ 3 trips per staff year). Regarding equipment and consumables etc., these are not required since we have complementary resources from Gaia and other related grants for personal computers/laptops and small-scale testbed hardware.

2.4.6 Universiteit Leiden (UL)

UL will be the lead partner for WP-200 with Brown as manager. It is estimated that he will spend 2 months on the management and coordination for this WP and 2 months on providing scientific guidance. From the funding one or two researchers (postdoctoral level) will be hired at UL to work on the sub-work packages 220, 230, 250, 260. The travel funds requested are needed to cover the rather frequent travel that the WP-200 researchers are expected to undertake in order to effectively liaise with the developers in the other WPs, who depend on WP-200 for providing the science-driven requirements specifications.

The package WP-240 will be carried out at INAF-OATo where a postdoctoral researcher will be hired from the GENIUS funds. The management of this WP and the coordination between INAF and UL will be done by Smart of INAF-OATo (2 staff months), while Spagna of INAF-OATo will contribute his expertise on cross-matching (2 staff months).

2.4.7 Centre de Serveis Científics i Acadèmics de Catalunya (CESCA)

CESCA's participation is based on personnel effort (to set up, configure and manage the hardware infrastructure to run the simulations as well as to contribute to the design and implementation of the community portal by providing advice and knowledge on web and repository infrastructures, mainly on WP500 and WP600, respectively) and on hardware resources in order to provide enough computational power to support the project (simulations and community portal).

2.4.8 Specialised partners

The budget of the specialised partners has been defined in a simplified, way since it is devoted to the provision of specific expertise in concrete tasks and amounts to a relatively small fraction of the total budget.

The starting point was a budget with a ratio personnel/other-costs of 1/3 aiming to provide a reasonable compromise for these smaller contributions between the human resources provided and other development costs (assumed to be travel to participate in meetings and generic expenses), with the exception of the KU partner (see below). No management is included since the task are considered to be purely RTD, with the exception of UCAM where some management effort is included to ensure the coordination with GREAT through N. Walton.

The budgets were then discussed with the partners case by case, adapting them to the specific circumstances of each one. This has led to some variations with respect to the initial proposal, in most cases reducing the fraction devoted to other expenses. The reason for this is that these partners do not expect their generic expenses to be significant and prefer to maximize the funding available for personnel. We have to distinguish here the special case of Kyoto University; for this partner almost all the budget is allocated to travel expenses because the purpose of their participation in GENIUS is to allow the coordination of with the JASMINE missions. In this case they do not intend to hire personnel or carry out development activities but just to be able to travel regularly to Europe (meetings, workshops, etc.).

3 Impact

We consider ‘impact’ within the following categories:

- Societal, e.g. public engagement and public outreach;
- Economic, e.g. industrial, knowledge transfer and commercial exploitation;
- Educational, e.g. undergraduate and postgraduate training; schools and Colleges (under-18s);
- Scientific, e.g. fundamental research and pushing the boundaries of knowledge.

In Section 3.2.1, we show how GENIUS naturally will lead to significant outreach into the public domain. Without a coordinated programme 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 will be developed within GENIUS, not from the Catalogue alone. Furthermore the impact on society goes beyond outreach only with, for example, surveillance activities. We will develop 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 WP300 we expect to work closely with ESAC and their commercial DBMS provider InterSystems. 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). Section 3.2.2 discusses our approach to managing the possibilities in this area.

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 will be developed within WP400 – there are established precedents for Hipparcos which pave the way for Gaia [14]. For postgraduate training, actions are already being coordinated through GREAT.³³

Finally, the scientific impact of Gaia, and therefore GENIUS, is clear. In Section 3.1 we describe the heritage of space astrometry in Europe and how Gaia/GENIUS represents the next giant step on this journey of discovery.

3.1 Expected impacts listed in the work programme

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

³³<http://www.great-esf.eu>

at the end of the nineteenth century” [11]. 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 catalogue³⁴. 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 the as yet unfunded but 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 proposed to be 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 proposal 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.

³⁴http://www.rssd.esa.int/index.php?project=HIPPARCOS&page=science_results

Gaia is a 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, 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 proposal will cement 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.

3.2 Dissemination and/or exploitation of project results, and management of intellectual property

3.2.1 Dissemination of GENIUS results

The very nature of the GENIUS project is the dissemination of the Gaia space data to the astronomical community with the availability of the archive and publication of its Catalogue through DPAC CU9 in coordination with ESA and DPAC. The Gaia archive will be located at ESAC, the European Space Astronomy Centre, where all ESA astronomy and Solar System missions archives are kept and well disseminated (regularly retrieved by more than 3000 registered users). GENIUS will thus ensure that its archive system design will be consistent with ESAC practice. Furthermore, by requirement the Gaia archive will be made to be easily mirrored, so we expect that several such mirrors will be installed around the world. Specifically, during our discussions with the Nano-JASMINE team (see Sec. 1.1.4) the possibility of a Gaia archive mirror located in Japan has been raised. In that sense, the exploitation of the project results is clearly guaranteed.

Beside the scientific community, the educational community and the general public are obvious targets of the GENIUS results. Due to the small size of this project and the presence of specialists of public outreach at ESA, at the national space agencies or universities, the communication to the large public will be done in collaboration, and with targeted funding. Still, the development of the visualisation and analysis tools will always be made with the educational and outreach possibilities in mind. In March 2012 ESA issued a Letter of Interest concerning the preparatory activities for the Gaia archive access. Responses to this Letter have shown that the Gaia community will be fully committed to make the Gaia achievements reach the general public. There is thus no doubt that this community will provide support for outreach and academic activities.

The community portal (see Sec. WP400 and WP600 descriptions) will be the main information source, planning to target each community (science community at large, academic and educational community and general public) the most efficient way, with different access levels. It will thus provide sections devoted to the project documentation, deliverable and reports, the access to the archive, a software area, the mission results, educational tools and associated how-to, and news. Beside, we will contribute to furnishing CU9 with material (images and animations) for press kits to the media. An example of the possibilities of Gaia outreach is the video already produced by the UB partner to promote Gaia in Spain, which can be seen at http://gaiavideo.ub.edu/index_ang.html. This kind of initiative can be enlarged in the framework of GENIUS beyond

Type	Destination	Product	Method
Research	CU9	Status reports	Workshops
Research	Professionals	Gaia Catalogue release	Archive
Research	Professionals	Visualisation of the general properties of the Gaia Catalogue	Project documentation
Research	DPAC	Validation results	Website
Research	Amateurs and professionals	Visualisation software	Website
Research	Amateurs and professionals	Science alerts	Website
Educational	schools	3-D display of the solar neighbourhood	Website
Educational	schools	Minor planets of the solar system	Website
Educational	schools	New worlds: Gaia exoplanets survey	Website
Educational	students	H-R diagrams of clusters	Website
Academic	students	The sky at different epochs	Website
Outreach	General public	A journey through the Galaxy	Film
Outreach	All	Latest news of the mission	Website
Outreach	General public	Visualisation of our Galaxy	Lectures
Outreach	General public	Wikipedia entries related to Gaia archive and its contents	Web
Outreach	General public	Press and news	Website

Table 3.1: Short summary of dissemination methods

the national communities to reach a wider European audience.

A limited list of the dissemination and exploitation methods of GENIUS products is shown in Table 3.1. Thanks to the tools developed, our goal will be to increase both the scientific return and the communication of the exciting science results.

3.2.2 Exploitation and intellectual property

GENIUS will be devoted to the development of software and IT systems for the implementation of the Gaia archive. The correct management of intellectual property rights (IPR) of a software development in a multinational collaboration can be very complex, since the various national regulations and the particularities of the participant institutions regarding the intellectual property of the developed software and systems have to be taken into account. We will therefore take special care to properly handle this issue.

First of all, the basic approach in GENIUS is:

1. The GENIUS products should be freely available for scientific and educational uses, and specifically no limitations should be made for its use in the implementation of the Gaia archive or its scientific exploitation.
2. At the same time, although GENIUS has essentially a scientific goal, we contemplate the possibility that its work on leading edge IT issues may in some cases offer opportunities for technology transfer. Provisions will be made so that if these opportunities arise they are promptly identified and the partners are properly advised on how to take advantage of them.

An appropriate formulation of these principles will be included in the partnership agreement and will govern the responsibilities of the participants.

On the other hand, a fraction of the UB budget has been allocated to IPR management (i.e. subcontracting specialised attorneys) and technology transfer issues.

4 Ethics issues

Note: GENIUS is a project devoted to information technology applied to the data collected by a space astrophysics mission. Hence, no ethics issues are expected to arise in the project.

ETHICS ISSUES TABLE

Research on Human Embryo/ Foetus		YES	Page
	Does the proposed research involve human Embryos?		
	Does the proposed research involve human Foetal Tissues/ Cells?		
	Does the proposed research involve human Embryonic Stem Cells (hESCs)?		
	Does the proposed research on human Embryonic Stem Cells involve cells in culture?		
	Does the proposed research on Human Embryonic Stem Cells involve the derivation of cells from Embryos?		
	I CONFIRM THAT NONE OF THE ABOVE ISSUES APPLY TO MY PROPOSAL	X	

Research on Humans		YES	Page
	Does the proposed research involve children?		
	Does the proposed research involve patients?		
	Does the proposed research involve persons not able to give consent?		
	Does the proposed research involve adult healthy volunteers?		
	Does the proposed research involve Human genetic material?		
	Does the proposed research involve Human biological samples?		
	Does the proposed research involve Human data collection?		
	I CONFIRM THAT NONE OF THE ABOVE ISSUES APPLY TO MY PROPOSAL	X	

Privacy		YES	Page
	Does the proposed research involve processing of genetic information or personal data (e.g. health, sexual lifestyle, ethnicity, political opinion, religious or philosophical conviction)?		
	Does the proposed research involve tracking the location or observation of people?		
	I CONFIRM THAT NONE OF THE ABOVE ISSUES APPLY TO MY PROPOSAL	X	

Research on Animals		YES	Page
	Does the proposed research involve research on animals?		
	Are those animals transgenic small laboratory animals?		
	Are those animals transgenic farm animals?		
	Are those animals non-human primates?		
	Are those animals cloned farm animals?		
	I CONFIRM THAT NONE OF THE ABOVE ISSUES APPLY TO MY PROPOSAL	X	

Research Involving non-EU Countries (ICPC Countries)		YES	Page
Is any material used in the research (e.g. personal data, animal and/or human tissue samples, genetic material, live animals, etc) :			
a) Collected and processed in any of the ICPC countries?			
b) Exported to any other country (including ICPC and EU Member States)?			
I CONFIRM THAT NONE OF THE ABOVE ISSUES APPLY TO MY PROPOSAL		X	

Dual use		YES	Page
Research having direct military use			
Research having the potential for terrorist abuse			
I CONFIRM THAT NONE OF THE ABOVE ISSUES APPLY TO MY PROPOSAL		X	

5 Consideration of gender aspects

The gender dimension will be an explicit component of research organisation and management of this project. The recruitment policy for the GENIUS-FP7 will conform to the principles of the European Charter for Researchers and the Code of Conduct for their recruitment. The recruitment procedure will place an emphasis on individual excellence. However, the principles of equal opportunity and gender balance will be stressed.

Gender bias is still a major problem in computational science, as shown even in the present composition of GENIUS, and it is clear that time will not redress the under-representation of women in an adequate fashion and that specific measures are needed. The number of women entering the sciences and earning degrees is increasing in all fields but one: computer science. Although numbers are encouraging in nearly all fields, including almost all science and engineering fields, a reverse trend in computer sciences is alarming. Particularly in our discipline, there has been a dramatic drop in women earning bachelor's degrees in EEUU. Recent figures from "She Figures 2009" indicate that the proportion of female PhD (ISCED6) graduates in computer science has been almost frozen to 16-18% at EU-27. Thus, it is mandatory to stimulate young woman researchers' careers in computer science.

One essential factor to achieve more equality is to train young female researchers. Here we propose to generate one/two research applications to be filled by young female researchers. In order to achieve this, we propose to have flexible funding be made available for a short period of time in order to stimulate young female researchers to apply for the GENIUS positions. As long as the gender bias exists, such funding should be open specifically to female applicants, because experience shows that otherwise male applicants may obtain a disproportionate share of such funding.

In GENIUS we also propose a transparent access to positions, because insufficient transparency often coincides with downgrading of female candidates. Criteria for open positions and promotion in this GENIUS project is proposed to be transparent and made public. The existing staff proposing to take part in the GENIUS collaboration is composed by 7 females and 36 males in the staff positions involved. This yields 16% female staff which is in the average of the IAU statistics for computer science. An effort to improve the balance will be undertaken.

5.1 Providing for good work-life balance in GENIUS

One of the causes for the present distribution is the relatively more limited availability of women, many of who for whatever reason end up siding with their families and are less mobile, both nationally and internationally. In order to counter this, we propose to allow for more but shorter trips and virtual international contacts. Working conditions are favourable to work-life balance if meetings and extracurricular obligations outside of regular working hours are kept to a minimum and the option of flexible working hours and working spaces is offered under fully transparent conditions.

5.2 GENIUS at university and faculties of computer science

To help increase the pool of female faculty members in computing science, GENIUS proposes an initiative to promote some conferences at the faculties about our project, locating woman as speakers to illustrate that successful women do exist in computing.

6 Bibliography

References

- [1] F. Arenou. *Contribution a la validation statistique des donnees d'Hipparcos: Catalogue d'Entree et donnees preliminaires*. PhD thesis, Observatoire de Paris, CNRS, March 1993.
- [2] J. Binney. Extracting science from surveys of our Galaxy. *Pramana*, 77:39–52, July 2011.
- [3] A. G. A. Brown, F. Arenou, N. Hambly, F. van Leeuwen, X. Luri, J.-C. Malapert, W. O'Mullane, D. Tapiador, and N. A. Walton. Gaia data access scenarios summary. Technical Report GAIA-C9-TN-LEI-AB-026, Leiden Observatory, 2012.
- [4] A. G. A. Brown, F. Arenou, F. van Leeuwen, L. Lindegren, and X. Luri. Considerations in Making Full Use of the HIPPARCOS Catalogue. In R. M. Bonnet, E. Høg, P. L. Bernacca, L. Emiliani, A. Blaauw, C. Turon, J. Kovalevsky, L. Lindegren, H. Hassan, M. Bouffard, B. Strim, D. Heger, M. A. C. Perryman, & L. Woltjer, editor, *Hipparcos - Venice '97*, volume 402 of *ESA Special Publication*, pages 63–68, August 1997.
- [5] N. Hambly, M. Read, R. Mann, E. Sutorius, I. Bond, H. MacGillivray, P. Williams, and A. Lawrence. The SuperCOSMOS Science Archive. In F. Ochsenbein, M. G. Allen, and D. Egret, editors, *Astronomical Data Analysis Software and Systems (ADASS) XIII*, volume 314 of *Astronomical Society of the Pacific Conference Series*, page 137, July 2004.
- [6] N. C. Hambly, R. S. Collins, N. J. G. Cross, R. G. Mann, M. A. Read, E. T. W. Sutorius, I. Bond, J. Bryant, J. P. Emerson, A. Lawrence, L. Rimoldini, J. M. Stewart, P. M. Williams, A. Adamson, P. Hirst, S. Dye, and S. J. Warren. The WFCAM Science Archive. *Monthly Notices of the Royal Astronomical Society*, 384:637–662, February 2008.
- [7] D. W. Hogg and D. Lang. Telescopes don't make catalogues! In *EAS Publications Series*, volume 45 of *EAS Publications Series*, pages 351–358, February 2011.
- [8] A. C. Hume, A. Krause, M. Holliman, R. G. Mann, K. Noddle, and S. Voutsinas. TAP Service Federation Factory. In P. Ballester, D. Egret, and N. P. F. Lorente, editors, *Astronomical Data Analysis Software and Systems XXI*, volume 461 of *Astronomical Society of the Pacific Conference Series*, page 359, September 2012.
- [9] D. Lang, D. W. Hogg, S. Jester, and H.-W. Rix. Measuring the Undetectable: Proper Motions and Parallaxes of Very Faint Sources. *AJ*, 137:4400–4411, May 2009.
- [10] F. Mignard, C. Bailer-Jones, U. Bastian, R. Drimmel, L. Eyer, D. Katz, F. van Leeuwen, X. Luri, W. O'Mullane, X. Passot, D. Pourbaix, and T. Prusti. Gaia: organisation and challenges for the data processing. In W. J. Jin, I. Platais, & M. A. C. Perryman, editor, *IAU Symposium*, volume 248 of *IAU Symposium*, pages 224–230, July 2008.
- [11] A. Mouchez. Letter to Pickering, Huggins, Cruels et Struve. June 1885.
- [12] W. O'Mullane. Blue skies and clouds, archives of the future . Technical Report GAIA-TN-PL-ESAC-WOM-057-01, ESAC, 2010.
- [13] W. O'Mullane and X. Luri. GAIA and Virtual Observatories. In R. J. Brunner, S. G. Djorgovski, & A. S. Szalay, editor, *Virtual Observatories of the Future*, volume 225 of *Astronomical Society of the Pacific Conference Series*, page 201, 2001.

- [14] J. R. Percy. The Educational Potential of the HIPPARCOS Database. In R. M. Bonnet, E. Høg, P. L. Bernacca, L. Emiliani, A. Blaauw, C. Turon, J. Kovalevsky, L. Lindegren, H. Hassan, M. Bouffard, B. Strim, D. Heger, M. A. C. Perryman, and L. Woltjer, editors, *Hipparcos - Venice '97*, volume 402 of *ESA Special Publication*, pages 739–742, August 1997.
- [15] M. Perryman. *Astronomical Applications of Astrometry: Ten Years of Exploitation of the Hipparcos Satellite Data*. Cambridge University Press, 2009.
- [16] M. Perryman, K.S. Boer, G. Gilmore, E. Høg, M. Lattanzi, L. Lindegren, X. Luri, F. Mignard, and P.T. de Zeeuw. *Gaia Concept and Technology Study Report, ESA-SCI(2000)4*. ESA, 2000.
- [17] M. A. C. Perryman, K. S. de Boer, G. Gilmore, E. Høg, M. G. Lattanzi, L. Lindegren, X. Luri, F. Mignard, O. Pace, and P. T. de Zeeuw. Gaia: Composition, formation and evolution of the galaxy. *Astron.Astrophys.*, 369:339-363,2001, 2001.
- [18] F. van Leeuwen. *Hipparcos, the New Reduction of the Raw Data*, volume 350 of *Astrophysics and Space Science Library*. Springer, 2007.

7 PROPOSAL END

**Gaia European Network for Improved data User Services
GENIUS**

PROPOSAL END