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

Grant agreement for: Collaborative project*

Annex I - "Description of Work"

Project acronym: GENIUS

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

Grant agreement no: 606740

Version date: 2013-06-21

Table of Contents

Part A

A.1 Project summary....3
A.2 List of beneficiaries....4
A.3 Overall budget breakdown for the project....5

Workplan Tables

WT1 List of work packages....1
WT2 List of deliverables....2
WT3 Work package descriptions....7
Work package 1....7
Work package 2....10
Work package 3....16
Work package 4....22
Work package 5....28
Work package 6....35
Work package 7....39
WT4 List of milestones....43
WT5 Tentative schedule of project reviews....45
WT6 Project effort by beneficiaries and work package....46
WT7 Project effort by activity type per beneficiary....47
WT8 Project efforts and costs....48

A1:
Project summary

Project Number [1]	606740	Project Acronym [2]	GENIUS

One form per project

General information

Project title [3]	Gaia European Network for Improved data User Services
Starting date[4]	The first day of the month after the signature by the Commission
Duration in months[5]	42
Call (part) identifier[6]	FP7-SPACE-2013-1
Activity code(s) most relevant to your topic[7]	SPA.2013.2.1-01:Exploitation of space science and exploration data

Abstract [9]

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

A2:
List of Beneficiaries

Project Number [1]	606740	Project Acronym [2]	GENIUS

List of Beneficiaries

No	Name	Short name	Country	Project entry month[10]	Project exit month
1	UNIVERSITAT DE BARCELONA	UB	Spain	1	42
2	CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE	CNRS	France	1	42
3	THE UNIVERSITY OF EDINBURGH	UEDIN	United Kingdom	1	42
4	UNIVERSITEIT LEIDEN	UL	Netherlands	1	42
5	CONSORCI CENTRE DE SERVEIS CIENTIFICS I ACADEMICS DE CATALUNYA	CESCA	Spain	1	42
6	ISTITUTO NAZIONALE DI ASTROFISICA	INAF	Italy	1	42
7	AGENCIA ESTATAL CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS	CSIC	Spain	1	42
8	UNIVERSITE DE GENEVE	UNIGE	Switzerland	1	42
9	UNIVERSITE LIBRE DE BRUXELLES	ULB	Belgium	1	42
10	FUNDACAO DA FACULDADE DE CIENCIAS DA UNIVERSIDADE DE LISBOA	FFCUL	Portugal	1	42
11	UNIVERSITY OF BRISTOL	UBR	United Kingdom	1	42
12	THE CHANCELLOR, MASTERS AND SCHOLARS OF THE UNIVERSITY OF CAMBRIDGE	UCAM	United Kingdom	1	42
13	National University Corporation, Kyoto University	KU	Japan	1	42

A3:
Budget Breakdown

Project Number [1]	606740	Project Acronym [2]	GENIUS

One Form per Project

Participant number in this project[11]	Participant short name	Fund. %[12]	Ind. costs[13]	Estimated eligible costs (whole duration of the project)					Total Receipts	Requested EU contribution
				RTD / Innovation (A)	Demonstration (B)	Management (C)	Other (D)	Total A+B+C+D		
1	UB	75.0	T	471,893.20	0.00	133,049.51	79,347.44	684,290.15	0.00	566,316.85
2	CNRS	75.0	T	567,809.60	0.00	0.00	23,379.20	591,188.80	0.00	449,236.40
3	UEDIN	75.0	S	481,191.00	0.00	0.00	0.00	481,191.00	0.00	360,893.00
4	UL	75.0	S	517,137.00	0.00	1,500.00	0.00	518,637.00	0.00	389,352.75
5	CESCA	75.0	T	57,600.00	0.00	0.00	66,600.00	124,200.00	0.00	109,800.00
6	INAF	75.0	T	133,440.00	0.00	0.00	0.00	133,440.00	0.00	100,080.00
7	CSIC	75.0	A	133,899.00	0.00	0.00	0.00	133,899.00	0.00	100,424.00
8	UNIGE	75.0	T	66,667.20	0.00	0.00	0.00	66,667.20	0.00	50,000.00
9	ULB	75.0	T	66,640.00	0.00	0.00	0.00	66,640.00	0.00	49,980.00
10	FFCUL	75.0	T	114,667.20	0.00	0.00	0.00	114,667.20	0.00	86,000.00
11	UBR	75.0	T	39,976.00	0.00	0.00	0.00	39,976.00	0.00	29,982.00
12	UCAM	75.0	T	203,641.60	0.00	0.00	0.00	203,641.60	0.00	152,198.00
13	KU	75.0	F	38,400.00	0.00	0.00	20,400.00	58,800.00	0.00	49,200.00
Total				2,892,961.80	0.00	134,549.51	189,726.64	3,217,237.95	0.00	2,493,463.00

Note that the budget mentioned in this table is the total budget requested by the Beneficiary and associated Third Parties.

*** The following funding schemes are distinguished**

Collaborative Project (if a distinction is made in the call please state which type of Collaborative project is referred to: (i) Small of medium-scale focused research project, (ii) Large-scale integrating project, (iii) Project targeted to special groups such as SMEs and other smaller actors), Network of Excellence, Coordination Action, Support Action.

1. Project number

The project number has been assigned by the Commission as the unique identifier for your project, and it cannot be changed. The project number **should appear on each page of the grant agreement preparation documents** to prevent errors during its handling.

2. Project acronym

Use the project acronym as indicated in the submitted proposal. It cannot be changed, unless agreed during the negotiations. The same acronym **should appear on each page of the grant agreement preparation documents** to prevent errors during its handling.

3. Project title

Use the title (preferably no longer than 200 characters) as indicated in the submitted proposal. Minor corrections are possible if agreed during the preparation of the grant agreement.

4. Starting date

Unless a specific (fixed) starting date is duly justified and agreed upon during the preparation of the Grant Agreement, the project will start on the first day of the month following the entry info force of the Grant Agreement (NB : entry into force = signature by the Commission). Please note that if a fixed starting date is used, you will be required to provide a detailed justification on a separate note.

5. Duration

Insert the duration of the project in full months.

6. Call (part) identifier

The Call (part) identifier is the reference number given in the call or part of the call you were addressing, as indicated in the publication of the call in the Official Journal of the European Union. You have to use the identifier given by the Commission in the letter inviting to prepare the grant agreement.

7. Activity code

Select the activity code from the drop-down menu.

8. Free keywords

Use the free keywords from your original proposal; changes and additions are possible.

9. Abstract

10. The month at which the participant joined the consortium, month 1 marking the start date of the project, and all other start dates being relative to this start date.

11. The number allocated by the Consortium to the participant for this project.

12. Include the funding % for RTD/Innovation - either 50% or 75%

13. Indirect cost model

A: Actual Costs
S: Actual Costs Simplified Method
T: Transitional Flat rate
F :Flat Rate

Workplan Tables

Project number

606740

Project title

GENIUS - Gaia European Network for Improved data User Services

Call (part) identifier

FP7-SPACE-2013-1

Funding scheme

Collaborative project

WT1
List of work packages

Project Number [1]	606740	Project Acronym [2]	GENIUS

LIST OF WORK PACKAGES (WP)

WP Number[53]	WP Title	Type of activity[54]	Lead beneficiary number[55]	Person-months[56]	Start month[57]	End month[58]
WP1	Management	MGT	1	18.00	1	42
WP2	Tailoring to the end user community	RTD	4	73.00	1	42
WP3	Aspects of archive system design	RTD	3	68.80	1	42
WP4	Tools for data exploitation	RTD	1	83.00	1	42
WP5	Tools for data validation and analysis	RTD	2	98.10	1	42
WP6	Support activities	RTD	1	30.00	1	42
WP7	Disemination	OTHER	5	20.80	1	42
			Total	391.70		

WT2:
List of Deliverables

Project Number [1]	606740	Project Acronym [2]	GENIUS

List of Deliverables - to be submitted for review to EC

Delive-rable Number[61]	Deliverable Title	WP number[5]	Lead benefi-ciary number	Estimated indicative person-months	Nature[62]	Dissemi-nation level[63]	Delivery date[64]
D1.1	Kick-off meeting (plenary)	WP1	1	3.00	Other	PU	1
D1.2	Semestral report 1	WP1	1	1.00	Report	PP	6
D1.3	Semestral report 2	WP1	1	1.00	Report	PP	12
D1.4	Semestral report 3	WP1	1	1.00	Report	PP	18
D1.5	Mid-term meeting (plenary)	WP1	1	2.50	Other	PU	21
D1.6	Semestral report 4	WP1	1	1.00	Report	PP	24
D1.7	Semestral report 5	WP1	1	1.00	Report	PP	30
D1.8	Semestral report 6	WP1	1	1.00	Report	PP	36
D1.9	Completion meeting (plenary)	WP1	1	3.00	Other	PU	41
D1.10	Final report to the External Advisory board	WP1	1	3.00	Report	PP	42
D1.11	Report on dynamics of the gender balance	WP1	1	0.50	Report	PU	42
D2.1	Requirements specification for catalogue and data archive	WP2	4	8.00	Report	PU	12
D2.2	Requirements specification for outreach facilities built into the archive system	WP2	4	3.00	Report	PU	12
D2.3	Requirements specification for generic projection module	WP2	4	6.00	Report	PU	18

WT2:
List of Deliverables

Delive-rable Number[61]	Deliverable Title	WP number[5]	Lead benefi-ciary number	Estimated indicative person-months	Nature[62]	Dissemi-nation level[63]	Delivery date[64]
D2.4	Requirements specification for data retrieval across archives	WP2	4	16.00	Report	PU	18
D2.5	Conclusion of requirements update gathering exercise.	WP2	4	8.00	Report	PU	24
D2.6	Requirements specification for incorporating new information into an existing archive.	WP2	4	12.00	Report	PU	30
D2.7	Requirements specification for the archiving of raw and intermediate data.	WP2	4	9.00	Report	PU	36
D2.8	Requirements specification for the archiving of the original software with which the archive was produced.	WP2	4	4.00	Report	PU	36
D2.9	Requirements specification for model comparison and optimization tools.	WP2	4	7.00	Report	PU	40
D3.1	GENIUS/ ESAC–SAT Co-ordination and Interface Control document	WP3	3	6.00	Report	PU	3
D3.2	Web2.0 user interface demonstration prototype deployment	WP3	3	12.00	Prototype	PU	18
D3.3	Gaia Data IVOA compliance document	WP3	3	6.00	Report	PU	42

WT2:
List of Deliverables

Delive-rable Number[61]	Deliverable Title	WP number[5]	Lead benefi-ciary number	Estimated indicative person-months	Nature[62]	Dissemi-nation level[63]	Delivery date[64]
D3.4	Deployed web services, code and documentation	WP3	3	28.80	Other	PU	42
D3.5	Data centre report and analysis document	WP3	3	3.00	Report	PU	42
D3.6	TAP+ code and documentation	WP3	3	7.00	Other	PU	42
D3.7	Deployed CANFAR–style VM research environment and produced reports and documentation	WP3	3	6.00	Prototype	PU	42
D4.1	Requirement specification document for the exploitation tools	WP4	1	5.00	Report	PU	6
D4.2	Delivery of first prototype of exploitation tools	WP4	1	23.00	Prototype	PP	12
D4.3	Delivery of second prototype of exploitation tools	WP4	1	20.00	Prototype	PP	24
D4.4	Delivery of third prototype of exploitation tools	WP4	1	20.00	Prototype	PP	36
D4.5	Delivery of exploitation tools user manuals and technical documentation	WP4	1	10.00	Report	PU	36
D4.6	Deployment of exploitation tools on the first actual Gaia archive	WP4	1	5.00	Other	PU	42
D5.1	Delivery of prototype of internal checking tools (WP 520)	WP5	2	4.00	Prototype	PP	12
D5.2	Delivery of prototype of	WP5	2	4.00	Prototype	PP	18

Delive-rable Number[61]	Deliverable Title	WP number[5]	Lead benefi-ciary number	Estimated indicative person-months	Nature[62]	Dissemi-nation level[63]	Delivery date[64]
	statistical tools (WP 550)						
D5.3	Delivery of internal consistency checking tools (WP 520)	WP5	2	6.00	Other	PP	24
D5.4	Delivery of statistical tools (WP 550)	WP5	2	6.00	Other	PP	36
D5.5	Delivery of model-based validation tools (WP 530)	WP5	2	14.00	Other	PP	36
D5.6	Delivery of prototype of external validation tools (WP 540)	WP5	2	11.00	Prototype	PP	24
D5.7	Delivery of external validation tools (WP 540)	WP5	2	12.00	Other	PP	36
D5.8	Delivery of special object tools (WP 560)	WP5	2	37.10	Other	PP	36
D5.9	Deployment of validation tools on the Gaia archive	WP5	2	4.00	Other	PU	42
D6.1	Delivery of first simulated catalogue data	WP6	1	6.00	Other	PU	6
D6.2	Deployment of first public science alerts prototype	WP6	12	4.00	Prototype	PU	12
D6.3	Delivery of second simulated catalogue data	WP6	1	6.00	Other	PU	18
D6.4	Deployment of second public science alerts prototype	WP6	12	4.00	Prototype	PU	24

WT2:
List of Deliverables

Delive-rable Number[61]	Deliverable Title	WP number[5]	Lead benefi-ciary number	Estimated indicative person-months	Nature[62]	Dissemi-nation level[63]	Delivery date[64]
D6.5	Delivery of third simulated catalogue data	WP6	1	6.00	Other	PU	30
D6.6	Deployment of third public science alerts prototype	WP6	12	4.00	Prototype	PU	36
D7.1	Basic setup for the community portal internally available for working	WP7	5	5.00	Prototype	PP	3
D7.2	First public version of the community portal	WP7	5	5.00	Other	PU	6
D7.3	Upgraded public version of the community portal	WP7	5	5.00	Other	PU	24
D7.4	Final (complete) version of the community portal. Handover to the Gaia community	WP7	5	5.80	Other	PU	42
			Total	**391.70**			

WT3:
Work package description

Project Number [1]	606740	Project Acronym [2]	GENIUS

One form per Work Package

Work package number [53]	WP1	**Type of activity[54]**	MGT
Work package title	Management		
Start month	1		
End month	42		
Lead beneficiary number [55]	1		

Objectives

This package provides the overall administrative management of GENIUS, as described in Sec. 2.1.

Description of work and role of partners

WP1 - Management [Months: 1-42]
UB
This work package includes the administrative tasks to fulfill the EC requirements and rules as well as the global administrative tasks inside the consortium, including financial management, intelectual property management and project documentation. These tasks will be carried out by the GENIUS coordinator assited by a hired project manager (to be devoted part time to the GENIUS tasks).

Person-Months per Participant

Participant number and short name[10]	**WP1 additional effort**	**WP1 TOTAL**
1 - UB	18.00	18.00
2 - CNRS	0.00	0.00
UFC	0.00	0.00
3 - UEDIN	0.00	0.00
4 - UL	0.00	0.00
5 - CESCA	0.00	0.00
6 - INAF	0.00	0.00
7 - CSIC	0.00	0.00
8 - UNIGE	0.00	0.00
9 - ULB	0.00	0.00
10 - FFCUL	0.00	0.00
11 - UBR	0.00	0.00
12 - UCAM	0.00	0.00
13 - KU	0.00	0.00
Total	18.00	18.00

WT3:
Work package description

List of deliverables

Delive-rable Number [61]	Deliverable Title	Lead benefi-ciary number	Estimated indicative person-months	Nature [62]	Dissemi-nation level [63]	Delivery date [64]
D1.1	Kick-off meeting (plenary)	1	3.00	Other	PU	1
D1.2	Semestral report 1	1	1.00	Report	PP	6
D1.3	Semestral report 2	1	1.00	Report	PP	12
D1.4	Semestral report 3	1	1.00	Report	PP	18
D1.5	Mid-term meeting (plenary)	1	2.50	Other	PU	21
D1.6	Semestral report 4	1	1.00	Report	PP	24
D1.7	Semestral report 5	1	1.00	Report	PP	30
D1.8	Semestral report 6	1	1.00	Report	PP	36
D1.9	Completion meeting (plenary)	1	3.00	Other	PU	41
D1.10	Final report to the External Advisory board	1	3.00	Report	PP	42
D1.11	Report on dynamics of the gender balance	1	0.50	Report	PU	42
		Total	**18.00**			

Description of deliverables

D1.1 : Kick-off meeting (plenary) [month 1]

D1.2 : Semestral report 1 [month 6]

D1.3 : Semestral report 2 [month 12]

D1.4 : Semestral report 3 [month 18]

D1.5 : Mid-term meeting (plenary) [month 21]

D1.6 : Semestral report 4 [month 24]

D1.7 : Semestral report 5 [month 30]

D1.8 : Semestral report 6. The effort of this deliverable includes the collection of information from the partners, the attendance to meetings (global and local for partners) and the writing of the report. [month 36]

D1.9 : Completion meeting (plenary) [month 41]

D1.10 : Final report to the Advisory board [month 42]

D1.11 : Report on dynamics of the gender balance [month 42]

Schedule of relevant Milestones

Milestone number [59]	Milestone name	Lead benefi-ciary number	Delivery date from Annex I [60]	Comments
MS1	Kick-off meeting (plenary)	1	1	
MS3	Hiring of main developers	1	4	

Schedule of relevant Milestones

Milestone number [59]	Milestone name	Lead beneficiary number	Delivery date from Annex I [60]	Comments
MS8	Mid-term review	1	20	
MS12	Prototype archive tools open to community	1	24	User testing will allow to give the necessary feedback on the archive tools which have been developed. Most Gaia data will not be available at the date of the milestone, but the testing will nevertheless be possible by providing access to simulated data. In order to get a feedback as large as possible on the archive tools, the access will be given to a large panel of users within and outside the Gaia community. They will e.g. be selected among the users having made requirements proposals (cf. GAIA-C9-TN-LEI-AB-026-1).
MS15	Completion meeting & final external review	1	41	
MS16	GENIUS products availability	1	42	GENIUS products availability

WT3:
Work package description

Project Number [1]	606740	Project Acronym [2]	GENIUS

One form per Work Package

Work package number [53]	WP2	**Type of activity**[54]	RTD
Work package title	Tailoring to the end user community		
Start month	1		
End month	42		
Lead beneficiary number [55]	4		

Objectives

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

Description of work and role of partners

WP2 - Tailoring to the end user community [Months: 1-42]
UL
• 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)

T2.1 - Technical coordination [Months: 1-42]
UL
This work package oversees the work conducted within 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.

T2.2 - Analysis and working out of requirements gathered by GAP [Months: 1-24]
UL, FFCUL, UCAM, KU
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 scenarios (the requirements gathering process and the collected data access scenarios are summarized in [3], available online at http://www.read.esa.int/1link/livelink/open/3125400) 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 ‘science alerts’.

T2.3 - Confronting complex models with complex catalogues [Months: 1-42]
UL
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 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

T2.4 - Seamless data retrieval across archives and wavelength domains [Months: 1-24]
INAF
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-Starss, 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).

T2.5 - The living archive [Months: 1-42]
UL
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.

T2.6 - Re-processing of archived (raw) data [Months: 1-42]
UL
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.

Person-Months per Participant

WT3:
Work package description

Participant number and short name[10]	Task1 specific effort	Task2 specific effort	Task3 specific effort	Task4 specific effort	Task5 specific effort	Task6 specific effort	WP2 additional effort	WP2 TOTAL
1 - UB	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2 - CNRS	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
UFC	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3 - UEDIN	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4 - UL	4.00	9.00	12.00	0.00	12.00	12.00	0.00	49.00
5 - CESCA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6 - INAF	0.00	0.00	0.00	16.00	0.00	0.00	0.00	16.00
7 - CSIC	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
8 - UNIGE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
9 - ULB	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10 - FFCUL	0.00	2.00	0.00	0.00	0.00	0.00	0.00	2.00
11 - UBR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
12 - UCAM	0.00	4.00	0.00	0.00	0.00	0.00	0.00	4.00
13 - KU	0.00	2.00	0.00	0.00	0.00	0.00	0.00	2.00
Total	4.00	17.00	12.00	16.00	12.00	12.00	0.00	73.00

List of deliverables

Deliverable Number [61]	Deliverable Title	Lead beneficiary number	Estimated indicative person-months	Nature [62]	Dissemination level [63]	Delivery date [64]
D2.1	Requirements specification for catalogue and data archive	4	8.00	Report	PU	12
D2.2	Requirements specification for outreach facilities built into the archive system	4	3.00	Report	PU	12
D2.3	Requirements specification for generic projection module	4	6.00	Report	PU	18
D2.4	Requirements specification for data retrieval across archives	4	16.00	Report	PU	18
D2.5	Conclusion of requirements update gathering exercise.	4	8.00	Report	PU	24
D2.6	Requirements specification for incorporating new information into an existing archive.	4	12.00	Report	PU	30
D2.7	Requirements specification for the archiving of raw and intermediate data.	4	9.00	Report	PU	36

WT3:
Work package description

List of deliverables

Delive-rable Number [61]	Deliverable Title	Lead benefi-ciary number	Estimated indicative person-months	Nature [62]	Dissemi-nation level [63]	Delivery date [64]
D2.8	Requirements specification for the archiving of the original software with which the archive was produced.	4	4.00	Report	PU	36
D2.9	Requirements specification for model comparison and optimization tools.	4	7.00	Report	PU	40
		Total	**73.00**			

Description of deliverables

D2.1 : Requirements specification for catalogue and data archive [month 12]

D2.2 : Requirements specification for outreach facilities built into the archive system [month 12]

D2.3 : Requirements specification for generic projection module [month 18]

D2.4 : Requirements specification for data retrieval across archives [month 18]

D2.5 : Conclusion of requirements update gathering exercise. [month 24]

D2.6 : Requirements specification for incorporating new information into an existing archive. [month 30]

D2.7 : Requirements specification for the archiving of raw and intermediate data. [month 36]

D2.8 : Requirements specification for the archiving of the original software with which the archive was produced. [month 36]

D2.9 : Requirements specification for model comparison and optimization tools. [month 40]

Schedule of relevant Milestones

Milestone number [59]	Milestone name	Lead benefi-ciary number	Delivery date from Annex I [60]	Comments
MS3	Hiring of main developers	1	4	
MS5	Archive user requirements document	4	12	
MS6	Requirements document for each subsystem	1	12	
MS8	Mid-term review	1	20	
MS9	User prototype archive review	3	24	
MS10	Exploitation tools review	1	24	
MS11	Validation tools review	2	24	Validation tools review
MS12	Prototype archive tools open to community	1	24	User testing will allow to give the necessary feedback on the archive tools which have been developed. Most Gaia data will not be available at the date of the milestone, but the testing

Schedule of relevant Milestones

Milestone number [59]	Milestone name	Lead beneficiary number	Delivery date from Annex I [60]	Comments
				will nevertheless be possible by providing access to simulated data. In order to get a feedback as large as possible on the archive tools, the access will be given to a large panel of users within and outside the Gaia community. They will e.g. be selected among the users having made requirements proposals (cf. GAIA-C9-TN-LEI-AB-026-1).
MS16	GENIUS products availability	1	42	GENIUS products availability

WT3:
Work package description

Project Number [1]	606740	Project Acronym [2]	GENIUS

One form per Work Package

Work package number [53]	WP3	**Type of activity**[54]	RTD
Work package title	Aspects of archive system design		
Start month	1		
End month	42		
Lead beneficiary number [55]	3		

Objectives

The objective of this workpackage is to design, prototype and develop aspects of the archive infrastructure needed for the scientific exploitation of Gaia data.

Description of work and role of partners

WP3 - Aspects of archive system design [Months: 1-42]
UEDIN
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.

T3.1 - Technical coordination [Months: 1-42]
UEDIN
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.

T3.2 - Aspects of archive interface design [Months: 1-42]
UEDIN
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)

T3.3 - VO infrastructure [Months: 1-42]
INAF, CNRS, UEDIN, CSIC
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 place (http://www.ivoa.net/Documents/) 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 SkyBOT (http://vo.imcce.fr/webservices/skybot/) service suite will provide VO-compliant tools for the treatment of solar system bodies within Gaia data, while Miriade (http://vo.imcce.fr/webservices/miriade/) 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 functionality 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

T3.4 - Data Centre Collaboration [Months: 1-42]
UEDIN
With the Table Access Protocol (TAP http://www.ivoa.net/Documents/TAP/) 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 Infrastructue (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 precomputed 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 undertaken by Read and Voutsinas of UEDIN.

T3.5 - Cloud-based research and data mining environments [Months: 1-42]
UEDIN
Research environments such as that provided by CADC with CANFAR (http://canfar.phys.uvic.ca/) 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 work (https://sites.google.com/site/nickballastronomer/research/canfar_skytree) 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.

WT3:
Work package description

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 date centre will be undertaken and conclusions reported.
This work will be undertaken by Read at UEDIN.

Person-Months per Participant

Participant number and short name[10]	Task1 specific effort	Task2 specific effort	Task3 specific effort	Task4 specific effort	Task5 specific effort	WP3 additional effort	WP3 TOTAL
1 - UB	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2 - CNRS	0.00	0.00	1.80	0.00	0.00	0.00	1.80
UFC	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3 - UEDIN	6.00	12.00	9.00	10.00	6.00	0.00	43.00
4 - UL	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5 - CESCA	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6 - INAF	0.00	0.00	18.00	0.00	0.00	0.00	18.00
7 - CSIC	0.00	0.00	6.00	0.00	0.00	0.00	6.00
8 - UNIGE	0.00	0.00	0.00	0.00	0.00	0.00	0.00
9 - ULB	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10 - FFCUL	0.00	0.00	0.00	0.00	0.00	0.00	0.00
11 - UBR	0.00	0.00	0.00	0.00	0.00	0.00	0.00
12 - UCAM	0.00	0.00	0.00	0.00	0.00	0.00	0.00
13 - KU	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Total	6.00	12.00	34.80	10.00	6.00	0.00	68.80

List of deliverables

Delive-rable Number [61]	Deliverable Title	Lead benefi-ciary number	Estimated indicative person-months	Nature [62]	Dissemi-nation level [63]	Delivery date [64]
D3.1	GENIUS/ESAC–SAT Co-ordination and Interface Control document	3	6.00	Report	PU	3
D3.2	Web2.0 user interface demonstration prototype deployment	3	12.00	Prototype	PU	18
D3.3	Gaia Data IVOA compliance document	3	6.00	Report	PU	42
D3.4	Deployed web services, code and documentation	3	28.80	Other	PU	42
D3.5	Data centre report and analysis document	3	3.00	Report	PU	42
D3.6	TAP+ code and documentation	3	7.00	Other	PU	42

WT3:
Work package description

List of deliverables

Delive-rable Number [61]	Deliverable Title	Lead benefi-ciary number	Estimated indicative person-months	Nature [62]	Dissemi-nation level [63]	Delivery date [64]
D3.7	Deployed CANFAR–style VM research environment and produced reports and documentation	3	6.00	Prototype	PU	42
		Total	**68.80**			

Description of deliverables

D3.1 : GENIUS/ESAC–SAT Co-ordination and Interface Control document [month 3]

D3.2 : Web2.0 user interface demonstration prototype deployment [month 18]

D3.3 : Gaia Data IVOA compliance document [month 42]

D3.4 : Deployed web services, code and documentation [month 42]

D3.5 : Data centre report and analysis document [month 42]

D3.6 : TAP+ code and documentation [month 42]

D3.7 : Deployed CANFAR–style VM research environment and produced reports and documentation [month 42]

Schedule of relevant Milestones

Milestone number [59]	Milestone name	Lead benefi-ciary number	Delivery date from Annex I [60]	Comments
MS2	Agreed testbed environment with ESAC-SAT & CU9	3	4	
MS3	Hiring of main developers	1	4	
MS6	Requirements document for each subsystem	1	12	
MS8	Mid-term review	1	20	
MS9	User prototype archive review	3	24	
MS12	Prototype archive tools open to community	1	24	User testing will allow to give the necessary feedback on the archive tools which have been developed. Most Gaia data will not be available at the date of the milestone, but the testing will nevertheless be possible by providing access to simulated data. In order to get a feedback as large as possible on the archive tools, the access will be given to a large panel of users within and outside

Schedule of relevant Milestones

Milestone number [59]	Milestone name	Lead beneficiary number	Delivery date from Annex I [60]	Comments
				the Gaia community. They will e.g. be selected among the users having made requirements proposals (cf. GAIA-C9-TN-LEI-AB-026-1).
MS13	Stress test	3	35	
MS14	Load of actual Gaia data	2	38	Load of actual Gaia data
MS16	GENIUS products availability	1	42	GENIUS products availability

WT3:
Work package description

Project Number [1]	606740	Project Acronym [2]	GENIUS

One form per Work Package

Work package number [53]	WP4	**Type of activity**[54]	RTD
Work package title	Tools for data exploitation		
Start month	1		
End month	42		
Lead beneficiary number [55]	1		

Objectives

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:

• 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.

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.

Description of work and role of partners

WP4 - Tools for data exploitation [Months: 1-42]

UB

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.

T4.1 - Technical coordination [Months: 1-42]

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.

T4.2 - Visualization tools ([Months: 1-42]

FFCUL, UB

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-730).

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.

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).

T4.3 - Data mining [Months: 1-42]
UB, CSIC
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

T4.4 - VO tools and services [Months: 1-42]
CSIC, UBR
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 http://www.star.bris.ac.uk/~mbt/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; 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 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.

VOSA (http://svo.cab.inta-csic.es/theory/vosa/): 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:

a. 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.

b. Design and Implement VO services and tools for the Gaia data.

c. Test and optimise, and validate of the VO tools and services providing performance monitoring.

d. Define/implement the query extensions necessary to query the catalogue to fulfil the specifications.

e. Obtain user feedback and update the tools and services if necessary

f. Write documentation

Person-Months per Participant

Participant number and short name[10]	Task1 specific effort	Task2 specific effort	Task3 specific effort	Task4 specific effort	WP4 additional effort	WP4 TOTAL
1 - UB	2.00	8.00	30.00	0.00	0.00	40.00
2 - CNRS	0.00	0.00	0.00	0.00	0.00	0.00
UFC	0.00	0.00	0.00	0.00	0.00	0.00
3 - UEDIN	0.00	0.00	0.00	0.00	0.00	0.00
4 - UL	0.00	0.00	0.00	0.00	0.00	0.00
5 - CESCA	0.00	0.00	0.00	0.00	0.00	0.00
6 - INAF	0.00	0.00	0.00	0.00	0.00	0.00
7 - CSIC	0.00	0.00	2.00	10.00	0.00	12.00
8 - UNIGE	0.00	0.00	0.00	0.00	0.00	0.00
9 - ULB	0.00	0.00	0.00	0.00	0.00	0.00

WT3:
Work package description

Participant number and short name[10]	Task1 specific effort	Task2 specific effort	Task3 specific effort	Task4 specific effort	WP4 additional effort	WP4 TOTAL
10 - FFCUL	0.00	28.00	0.00	0.00	0.00	28.00
11 - UBR	0.00	0.00	0.00	3.00	0.00	3.00
12 - UCAM	0.00	0.00	0.00	0.00	0.00	0.00
13 - KU	0.00	0.00	0.00	0.00	0.00	0.00
Total	2.00	36.00	32.00	13.00	0.00	83.00

List of deliverables

Delive-rable Number [61]	Deliverable Title	Lead benefi-ciary number	Estimated indicative person-months	Nature [62]	Dissemi-nation level [63]	Delivery date [64]
D4.1	Requirement specification document for the exploitation tools	1	5.00	Report	PU	6
D4.2	Delivery of first prototype of exploitation tools	1	23.00	Prototype	PP	12
D4.3	Delivery of second prototype of exploitation tools	1	20.00	Prototype	PP	24
D4.4	Delivery of third prototype of exploitation tools	1	20.00	Prototype	PP	36
D4.5	Delivery of exploitation tools user manuals and technical documentation	1	10.00	Report	PU	36
D4.6	Deployment of exploitation tools on the first actual Gaia archive	1	5.00	Other	PU	42
		Total	**83.00**			

Description of deliverables

D4.1 : Requirement specification document for the exploitation tools [month 6]

D4.2 : Delivery of first prototype of exploitation tools [month 12]

D4.3 : Delivery of second prototype of exploitation tools [month 24]

D4.4 : Delivery of third prototype of exploitation tools [month 36]

D4.5 : Delivery of exploitation tools user manuals and technical documentation [month 36]

D4.6 : Deployment of exploitation tools on the first actual Gaia archive [month 42]

Schedule of relevant Milestones

Milestone number [59]	Milestone name	Lead benefi-ciary number	Delivery date from Annex I [60]	Comments
MS3	Hiring of main developers	1	4	
MS6	Requirements document for each subsystem	1	12	

WT3:
Work package description

Schedule of relevant Milestones

Milestone number [59]	Milestone name	Lead beneficiary number	Delivery date from Annex I [60]	Comments
MS8	Mid-term review	1	20	
MS10	Exploitation tools review	1	24	
MS12	Prototype archive tools open to community	1	24	User testing will allow to give the necessary feedback on the archive tools which have been developed. Most Gaia data will not be available at the date of the milestone, but the testing will nevertheless be possible by providing access to simulated data. In order to get a feedback as large as possible on the archive tools, the access will be given to a large panel of users within and outside the Gaia community. They will e.g. be selected among the users having made requirements proposals (cf. GAIA-C9-TN-LEI-AB-026-1).
MS16	GENIUS products availability	1	42	GENIUS products availability

WT3:
Work package description

Project Number [1]	606740	Project Acronym [2]	GENIUS

One form per Work Package

Work package number [53]	WP5	**Type of activity**[54]	RTD
Work package title	Tools for data validation and analysis		
Start month	1		
End month	42		
Lead beneficiary number [55]	2		

Objectives

The preparation of the Gaia archive before its publication requires a careful, detailed and indepth 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:
520. Looking for trouble: definition of problem cases, validation scenarios and tools
530. Simulation versus reality: from models to observables
540. Confronting Gaia to external archives
550. Data demining: outlier analysis
560. Transversal tools for special objects

Description of work and role of partners

WP5 - Tools for data validation and analysis [Months: 1-42]
CNRS
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.

T5.1 - Technical coordination [Months: 1-42]
CNRS
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.

T5.2 - Looking for trouble: definition of problem cases, validation scenarios and tools [Months: 1-42]
CNRS

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

T5.3 - Simulation versus reality: from models to observables [Months: 1-42]
CNRS
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.

T5.4 - Confronting Gaia to external archives [Months: 1-42]
CNRS, CSIC, KU
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 behaviour 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.

T5.5 - Data demining: outlier analysis [Months: 1-42]
CNRS, FFCUL
Outliers being by definition objects which deviate from an assumed model, it would be surprising that a mission such as Gaia planned for deciphering the complex structure of the Galaxy would exhibit no outliers departing from our current knowledge.

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

T5.6 - Transversal tools for special objects [Months: 1-42]
CNRS, UNIGE, ULB
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 107 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 behaviour 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 behaviour 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.

Person-Months per Participant

Participant number and short name[10]	Task1 specific effort	Task2 specific effort	Task3 specific effort	Task4 specific effort	Task5 specific effort	Task6 specific effort	WP5 additional effort	WP5 TOTAL
1 - UB	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2 - CNRS	4.00	10.00	13.45	17.00	8.00	24.60	0.00	77.05
UFC	0.00	0.00	0.55	0.00	0.00	0.00	0.00	0.55
3 - UEDIN	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4 - UL	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5 - CESCA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6 - INAF	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
7 - CSIC	0.00	0.00	0.00	4.00	0.00	0.00	0.00	4.00
8 - UNIGE	0.00	0.00	0.00	0.00	0.00	6.50	0.00	6.50
9 - ULB	0.00	0.00	0.00	0.00	0.00	6.00	0.00	6.00
10 - FFCUL	0.00	0.00	0.00	0.00	2.00	0.00	0.00	2.00
11 - UBR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
12 - UCAM	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
13 - KU	0.00	0.00	0.00	2.00	0.00	0.00	0.00	2.00
Total	4.00	10.00	14.00	23.00	10.00	37.10	0.00	98.10

WT3:
Work package description

List of deliverables

Delive-rable Number [61]	Deliverable Title	Lead benefi-ciary number	Estimated indicative person-months	Nature [62]	Dissemi-nation level [63]	Delivery date [64]
D5.1	Delivery of prototype of internal checking tools (WP 520)	2	4.00	Prototype	PP	12
D5.2	Delivery of prototype of statistical tools (WP 550)	2	4.00	Prototype	PP	18
D5.3	Delivery of internal consistency checking tools (WP 520)	2	6.00	Other	PP	24
D5.4	Delivery of statistical tools (WP 550)	2	6.00	Other	PP	36
D5.5	Delivery of model-based validation tools (WP 530)	2	14.00	Other	PP	36
D5.6	Delivery of prototype of external validation tools (WP 540)	2	11.00	Prototype	PP	24
D5.7	Delivery of external validation tools (WP 540)	2	12.00	Other	PP	36
D5.8	Delivery of special object tools (WP 560)	2	37.10	Other	PP	36
D5.9	Deployment of validation tools on the Gaia archive	2	4.00	Other	PU	42
		Total	**98.10**			

Description of deliverables

D5.1 : Delivery of prototype of internal checking tools (WP 520) [month 12]

D5.2 : Delivery of prototype of statistical tools (WP 550) [month 18]

D5.3 : Delivery of internal consistency checking tools (WP 520) [month 24]

D5.4 : Delivery of statistical tools (WP 550) [month 36]

D5.5 : Delivery of model-based validation tools (WP 530) [month 36]

D5.6 : Delivery of prototype of external validation tools (WP 540) [month 24]

D5.7 : Delivery of external validation tools (WP 540) [month 36]

D5.8 : Delivery of special object tools (WP 560) [month 36]

D5.9 : Deployment of validation tools on the Gaia archive [month 42]

Schedule of relevant Milestones

Milestone number [59]	Milestone name	Lead benefi-ciary number	Delivery date from Annex I [60]	Comments
MS3	Hiring of main developers	1	4	
MS6	Requirements document for each subsystem	1	12	
MS8	Mid-term review	1	20	

WT3:
Work package description

Schedule of relevant Milestones

Milestone number [59]	Milestone name	Lead benefi-ciary number	Delivery date from Annex I [60]	Comments
MS11	Validation tools review	2	24	Validation tools review
MS12	Prototype archive tools open to community	1	24	User testing will allow to give the necessary feedback on the archive tools which have been developed. Most Gaia data will not be available at the date of the milestone, but the testing will nevertheless be possible by providing access to simulated data. In order to get a feedback as large as possible on the archive tools, the access will be given to a large panel of users within and outside the Gaia community. They will e.g. be selected among the users having made requirements proposals (cf. GAIA-C9-TN-LEI-AB-026-1).
MS14	Load of actual Gaia data	2	38	Load of actual Gaia data
MS16	GENIUS products availability	1	42	GENIUS products availability

WT3:
Work package description

Project Number [1]	606740	Project Acronym [2]	GENIUS

One form per Work Package

Work package number [53]	WP6	**Type of activity**[54]	RTD
Work package title	Support activities		
Start month	1		
End month	42		
Lead beneficiary number [55]	1		

Objectives

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

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

2. The provision of an testbed for science alerts; the prototypes of the science alerts system will be installed in it for testing and validation and made accessible to the test users.

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

Description of work and role of partners

WP6 - Support activities [Months: 1-42]
UB
The work is divided into three sub-work packages:
610 - Technical coordination (1 staff months): 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 (17 staff months): the products of the project should be avail- able 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-workpackage is devoted to the provision of sim- ulated 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.
630 - Science alerts testbed (12 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 ba- sic information for each flux alert via the VOEvent system to the community in a timely fashion (with alerts being produced 1–2 days after observation by Gaia). The alert packet will contain basic characterisation information for each event, including parameters such as estimated alert object type, and more advanced classification for certain objects such as su- pernovae (SNe). For these, inherent Gaia photometric data will be used to provide additional information concerning SNe alerts including class, epoch, redshift, reddening.
640 - Community portal infrastructure (12 staff months): The development and implementa- tion 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. 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, also the consequent maintenance of the needed software and hardware.

T6.1 - Technical coordination [Months: 1-42]
UB

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.

T6.2 - Simulated catalogue data [Months: 1-42]
CESCA , UB
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 CESCA. For this purpose CESCA has several supercomputers with different architectures which will allow running the software needed to carry out these simulations. As an example CESCA 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. CESCA 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

T6.3 - Science alerts testbed [Months: 1-42]
UCAM
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 ba- sic information for each flux alert via the VOEvent system to the community in a timely fashion (with alerts being produced 1–2 days after observation by Gaia). The alert packet will contain basic characterisation information for each event, including parameters such as estimated alert object type, and more advanced classification for certain objects such as su- pernovae (SNe). For these, inherent Gaia photometric data will be used to provide additional information concerning SNe alerts including class, epoch, redshift, reddening.
The testbed work to be carried out in WP-630 will develop the interfaces required to connect the real time science alerts classification processing to the main Gaia data products. Thus, as the mission evolves, and more knowledge is accumulated about objects measured by Gaia as it successively scans the sky, there will be opportunity to cross reference new alerts against previous knowledge of that sky point as well as previous alerts against new information. Thus for instance, irregular outburst events may show multiple times during the Gaia mis- sion. Identification will be improved through correlation with earlier Gaia knowledge. The testbed will in addition provide linkages to external data resources provided through GE- NIUS, in particular via interfaces to the archive development through WP300. Finally the alerts testbed will plugin to the portal testbed developed in WP720.
With the termination of the GENIUS WP630 testbed activity, the full functionality will be deployed for community use - providing enhanced access to science alert data from 2015 onwards.

Person-Months per Participant

WT3:
Work package description

Participant number and short name[10]	Task1 specific effort	Task2 specific effort	Task3 specific effort	WP6 additional effort	WP6 TOTAL
1 - UB	1.00	5.00	0.00	0.00	6.00
2 - CNRS	0.00	0.00	0.00	0.00	0.00
UFC	0.00	0.00	0.00	0.00	0.00
3 - UEDIN	0.00	0.00	0.00	0.00	0.00
4 - UL	0.00	0.00	0.00	0.00	0.00
5 - CESCA	0.00	12.00	0.00	0.00	12.00
6 - INAF	0.00	0.00	0.00	0.00	0.00
7 - CSIC	0.00	0.00	0.00	0.00	0.00
8 - UNIGE	0.00	0.00	0.00	0.00	0.00
9 - ULB	0.00	0.00	0.00	0.00	0.00
10 - FFCUL	0.00	0.00	0.00	0.00	0.00
11 - UBR	0.00	0.00	0.00	0.00	0.00
12 - UCAM	0.00	0.00	12.00	0.00	12.00
13 - KU	0.00	0.00	0.00	0.00	0.00
Total	1.00	17.00	12.00	0.00	30.00

List of deliverables

Delive-rable Number [61]	Deliverable Title	Lead benefi-ciary number	Estimated indicative person-months	Nature [62]	Dissemi-nation level [63]	Delivery date [64]
D6.1	Delivery of first simulated catalogue data	1	6.00	Other	PU	6
D6.2	Deployment of first public science alerts prototype	12	4.00	Prototype	PU	12
D6.3	Delivery of second simulated catalogue data	1	6.00	Other	PU	18
D6.4	Deployment of second public science alerts prototype	12	4.00	Prototype	PU	24
D6.5	Delivery of third simulated catalogue data	1	6.00	Other	PU	30
D6.6	Deployment of third public science alerts prototype	12	4.00	Prototype	PU	36
		Total	**30.00**			

Description of deliverables

D6.1 : Delivery of first simulated catalogue data [month 6]

D6.2 : Deployment of first public science alerts prototype [month 12]

D6.3 : Delivery of second simulated catalogue data [month 18]

D6.4 : Deployment of second public science alerts prototype [month 24]

D6.5 : Delivery of third simulated catalogue data [month 30]

D6.6 : Deployment of third public science alerts prototype [month 36]

Schedule of relevant Milestones

Milestone number [59]	Milestone name	Lead beneficiary number	Delivery date from Annex I [60]	Comments
MS3	Hiring of main developers	1	4	
MS4	GENIUS portal available at CESCA for internal use	5	6	GENIUS portal available at CESCA for internal use
MS7	Public version of GENIUS portal	5	12	Public version of GENIUS portal
MS8	Mid-term review	1	20	
MS12	Prototype archive tools open to community	1	24	User testing will allow to give the necessary feedback on the archive tools which have been developed. Most Gaia data will not be available at the date of the milestone, but the testing will nevertheless be possible by providing access to simulated data. In order to get a feedback as large as possible on the archive tools, the access will be given to a large panel of users within and outside the Gaia community. They will e.g. be selected among the users having made requirements proposals (cf. GAIA-C9-TN-LEI-AB-026-1).
MS16	GENIUS products availability	1	42	GENIUS products availability

WT3:
Work package description

Project Number [1]	606740	Project Acronym [2]	GENIUS

One form per Work Package

Work package number [53]	WP7	**Type of activity**[54]	OTHER
Work package title	Disemination		
Start month	1		
End month	42		
Lead beneficiary number [55]	5		

Objectives

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

Description of work and role of partners

WP7 - Disemination [Months: 1-42]
CESCA
Covering Community portal, outreach and academic activities. 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.

T7.1 - Coordination of dissemination activities [Months: 1-42]
UB
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.

T7.2 - Community portal infrastructure [Months: 1-42]
CESCA
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.

T7.3 - Community portal, outreach and academic activities [Months: 1-42]
UB, CNRS
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 CESCA 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.

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) 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.

Person-Months per Participant

Participant number and short name[10]	Task1 specific effort	Task2 specific effort	Task3 specific effort	WP7 additional effort	WP7 TOTAL
1 - UB	4.00	0.00	3.00	0.00	7.00
2 - CNRS	0.00	0.00	1.80	0.00	1.80
UFC	0.00	0.00	0.00	0.00	0.00
3 - UEDIN	0.00	0.00	0.00	0.00	0.00
4 - UL	0.00	0.00	0.00	0.00	0.00
5 - CESCA	0.00	12.00	0.00	0.00	12.00
6 - INAF	0.00	0.00	0.00	0.00	0.00

WT3:
Work package description

Participant number and short name[10]	Task1 specific effort	Task2 specific effort	Task3 specific effort	WP7 additional effort	WP7 TOTAL
7 - CSIC	0.00	0.00	0.00	0.00	0.00
8 - UNIGE	0.00	0.00	0.00	0.00	0.00
9 - ULB	0.00	0.00	0.00	0.00	0.00
10 - FFCUL	0.00	0.00	0.00	0.00	0.00
11 - UBR	0.00	0.00	0.00	0.00	0.00
12 - UCAM	0.00	0.00	0.00	0.00	0.00
13 - KU	0.00	0.00	0.00	0.00	0.00
Total	4.00	12.00	4.80	0.00	20.80

List of deliverables

Deliverable Number [61]	Deliverable Title	Lead beneficiary number	Estimated indicative person-months	Nature [62]	Dissemination level [63]	Delivery date [64]
D7.1	Basic setup for the community portal internally available for working	5	5.00	Prototype	PP	3
D7.2	First public version of the community portal	5	5.00	Other	PU	6
D7.3	Upgraded public version of the community portal	5	5.00	Other	PU	24
D7.4	Final (complete) version of the community portal. Handover to the Gaia community	5	5.80	Other	PU	42
		Total	**20.80**			

Description of deliverables

D7.1 : Basic setup of the community portal internally available for working [month 3]

D7.2 : First public version of the community portal [month 6]

D7.3 : Upgraded public version of the community portal [month 24]

D7.4 : Final (complete) version of the community portal. Handover to the Gaia community. [month 42]

Schedule of relevant Milestones

Milestone number [59]	Milestone name	Lead beneficiary number	Delivery date from Annex I [60]	Comments
MS3	Hiring of main developers	1	4	
MS4	GENIUS portal available at CESCA for internal use	5	6	GENIUS portal available at CESCA for internal use

WT3:
Work package description

Schedule of relevant Milestones

Milestone number [59]	Milestone name	Lead beneficiary number	Delivery date from Annex I [60]	Comments
MS7	Public version of GENIUS portal	5	12	Public version of GENIUS portal
MS8	Mid-term review	1	20	
MS12	Prototype archive tools open to community	1	24	User testing will allow to give the necessary feedback on the archive tools which have been developed. Most Gaia data will not be available at the date of the milestone, but the testing will nevertheless be possible by providing access to simulated data. In order to get a feedback as large as possible on the archive tools, the access will be given to a large panel of users within and outside the Gaia community. They will e.g. be selected among the users having made requirements proposals (cf. GAIA-C9-TN-LEI-AB-026-1).
MS16	GENIUS products availability	1	42	GENIUS products availability
MS17	Handover of GENIUS portal to Gaia community	5	42	Handover of GENIUS portal to Gaia community

WT4:
List of Milestones

Project Number [1]	606740	Project Acronym [2]	GENIUS

List and Schedule of Milestones

Milestone number[59]	Milestone name	WP number[53]	Lead beneficiary number	Delivery date from Annex I[60]	Comments
MS1	Kick-off meeting (plenary)	WP1	1	1	
MS2	Agreed testbed environment with ESAC-SAT & CU9	WP3	3	4	
MS3	Hiring of main developers	WP1, WP2, WP3, WP4, WP5, WP6, WP7	1	4	
MS4	GENIUS portal available at CESCA for internal use	WP6, WP7	5	6	GENIUS portal available at CESCA for internal use
MS5	Archive user requirements document	WP2	4	12	
MS6	Requirements document for each subsystem	WP2, WP3, WP4, WP5	1	12	
MS7	Public version of GENIUS portal	WP6, WP7	5	12	Public version of GENIUS portal
MS8	Mid-term review	WP1, WP2, WP3, WP4, WP5, WP6, WP7	1	20	
MS9	User prototype archive review	WP2, WP3	3	24	
MS10	Exploitation tools review	WP2, WP4	1	24	
MS11	Validation tools review	WP2, WP5	2	24	Validation tools review
MS12	Prototype archive tools open to community	WP1, WP2, WP3, WP4, WP5, WP6, WP7	1	24	User testing will allow to give the necessary feedback on the archive tools which have been developed. Most Gaia data will not be available at the date of the milestone, but the testing will nevertheless be possible by providing access to simulated data. In order to get a feedback as large as possible on the archive tools, the access will be given to a large panel of

WT4:
List of Milestones

Milestone number[59]	Milestone name	WP number[53]	Lead beneficiary number	Delivery date from Annex I[60]	Comments
					users within and outside the Gaia community. They will e.g. be selected among the users having made requirements proposals (cf. GAIA-C9-TN-LEI-AB-026-1).
MS13	Stress test	WP3	3	35	
MS14	Load of actual Gaia data	WP3, WP5	2	38	Load of actual Gaia data
MS15	Completion meeting & final external review	WP1	1	41	
MS16	GENIUS products availability	WP1, WP2, WP3, WP4, WP5, WP6, WP7	1	42	GENIUS products availability
MS17	Handover of GENIUS portal to Gaia community	WP7	5	42	Handover of GENIUS portal to Gaia community

WT5:
Tentative schedule of Project Reviews

Project Number [1]	606740	Project Acronym [2]	GENIUS

Tentative schedule of Project Reviews

Review number[65]	Tentative timing	Planned venue of review	Comments, if any
RV1	12	Leiden	First review after first year
RV2	27	Paris	Mid term review. First actual Gaia data available.
RV3	42	Barcelona	Final review. First full Gaia catalogue release available.

WT6:
Project Effort by Beneficiary and Work Package

Project Number [1]	606740	Project Acronym [2]	GENIUS

Indicative efforts (Person-months) per Beneficiary per Work Package

Beneficiary number and short-name	WP1	WP2	WP3	WP4	WP5	WP6	WP7	Total per Beneficiary
1 - UB	18.00	0.00	0.00	40.00	0.00	6.00	7.00	71.00
2 - CNRS	0.00	0.00	1.80	0.00	77.05	0.00	1.80	80.65
UFC	0.00	0.00	0.00	0.00	0.55	0.00	0.00	0.55
3 - UEDIN	0.00	0.00	43.00	0.00	0.00	0.00	0.00	43.00
4 - UL	0.00	49.00	0.00	0.00	0.00	0.00	0.00	49.00
5 - CESCA	0.00	0.00	0.00	0.00	0.00	12.00	12.00	24.00
6 - INAF	0.00	16.00	18.00	0.00	0.00	0.00	0.00	34.00
7 - CSIC	0.00	0.00	6.00	12.00	4.00	0.00	0.00	22.00
8 - UNIGE	0.00	0.00	0.00	0.00	6.50	0.00	0.00	6.50
9 - ULB	0.00	0.00	0.00	0.00	6.00	0.00	0.00	6.00
10 - FFCUL	0.00	2.00	0.00	28.00	2.00	0.00	0.00	32.00
11 - UBR	0.00	0.00	0.00	3.00	0.00	0.00	0.00	3.00
12 - UCAM	0.00	4.00	0.00	0.00	0.00	12.00	0.00	16.00
13 - KU	0.00	2.00	0.00	0.00	2.00	0.00	0.00	4.00
Total	18.00	73.00	68.80	83.00	98.10	30.00	20.80	391.70

WT7:
Project Effort by Activity type per Beneficiary

Project Number [1]	606740	Project Acronym [2]	GENIUS

Indicative efforts per Activity Type per Beneficiary

Activity type	Part. 1 UB	Part. 2 CNRS	UFC	Part. 3 UEDIN	Part. 4 UL	Part. 5 CESCA	Part. 6 INAF	Part. 7 CSIC	Part. 8 UNIGE	Part. 9 ULB	Part. 10 FFCUL	Part. 11 UBR	Part. 12 UCAM	Part. 13 KU	Total
1. RTD/Innovation activities															
WP2	0.00	0.00	0.00	0.00	49.00	0.00	16.00	0.00	0.00	0.00	2.00	0.00	4.00	2.00	73.00
WP3	0.00	1.80	0.00	43.00	0.00	0.00	18.00	6.00	0.00	0.00	0.00	0.00	0.00	0.00	68.80
WP4	40.00	0.00	0.00	0.00	0.00	0.00	0.00	12.00	0.00	0.00	28.00	3.00	0.00	0.00	83.00
WP5	0.00	77.05	0.55	0.00	0.00	0.00	0.00	4.00	6.50	6.00	2.00	0.00	0.00	2.00	98.10
WP6	6.00	0.00	0.00	0.00	0.00	12.00	0.00	0.00	0.00	0.00	0.00	0.00	12.00	0.00	30.00
Total Research	**46.00**	**78.85**	**0.55**	**43.00**	**49.00**	**12.00**	**34.00**	**22.00**	**6.50**	**6.00**	**32.00**	**3.00**	**16.00**	**4.00**	**352.90**
2. Demonstration activities															
Total Demo	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
3. Consortium Management activities															
WP1	18.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	18.00
Total Management	**18.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**18.00**
4. Other activities															
WP7	7.00	1.80	0.00	0.00	0.00	12.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	20.80
Total other	**7.00**	**1.80**	**0.00**	**0.00**	**0.00**	**12.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**20.80**
Total	**71.00**	**80.65**	**0.55**	**43.00**	**49.00**	**24.00**	**34.00**	**22.00**	**6.50**	**6.00**	**32.00**	**3.00**	**16.00**	**4.00**	**391.70**

WT8:
Project Effort and costs

Project Number [1]	606740	Project Acronym [2]	GENIUS

Project efforts and costs

Beneficiary number	Beneficiary short name	Estimated eligible costs (whole duration of the project)						Requested EU contribution (€)
		Effort (PM)	Personnel costs (€)	Subcontracting (€)	Other Direct costs (€)	Indirect costs OR lump sum, flat-rate or scale-of-unit (€)	Total costs	
1	UB	71.00	314,953.40	16,000.01	102,727.94	250,608.80	684,290.15	566,316.85
2	CNRS	80.65	351,035.00	0.00	15,000.00	219,621.00	585,656.00	445,086.80
	UFC	0.55	3,458.00	0.00	0.00	2,074.80	5,532.80	4,149.60
3	UEDIN	43.00	261,265.00	0.00	18,320.00	201,606.00	481,191.00	360,893.00
4	UL	49.00	256,655.00	1,500.00	35,000.00	225,482.00	518,637.00	389,352.75
5	CESCA	24.00	68,000.00	9,000.00	4,000.00	43,200.00	124,200.00	109,800.00
6	INAF	34.00	73,400.00	0.00	10,000.00	50,040.00	133,440.00	100,080.00
7	CSIC	22.00	84,802.00	0.00	5,000.00	44,097.00	133,899.00	100,424.00
8	UNIGE	6.50	41,000.00	0.00	667.00	25,000.20	66,667.20	50,000.00
9	ULB	6.00	41,000.00	0.00	650.00	24,990.00	66,640.00	49,980.00
10	FFCUL	32.00	61,667.00	0.00	10,000.00	43,000.20	114,667.20	86,000.00
11	UBR	3.00	21,623.00	0.00	3,362.00	14,991.00	39,976.00	29,982.00
12	UCAM	16.00	115,276.00	0.00	12,000.00	76,365.60	203,641.60	152,198.00
13	KU	4.00	0.00	0.00	49,000.00	9,800.00	58,800.00	49,200.00
Total		391.70	1,694,134.40	26,500.01	265,726.94	1,230,876.60	3,217,237.95	2,493,463.00

1. Project number

The project number has been assigned by the Commission as the unique identifier for your project. It cannot be changed. The project number **should appear on each page of the grant agreement preparation documents (part A and part B)** to prevent errors during its handling.

2. Project acronym

Use the project acronym as given in the submitted proposal. It cannot be changed unless agreed so during the negotiations. The same acronym **should appear on each page of the grant agreement preparation documents (part A and part B)** to prevent errors during its handling.

53. Work Package number

Work package number: WP1, WP2, WP3, ..., WPn

54. Type of activity

For all FP7 projects each work package must relate to one (and only one) of the following possible types of activity (only if applicable for the chosen funding scheme # must correspond to the GPF Form Ax.v):

• **RTD/INNO =** Research and technological development including scientific coordination - applicable for Collaborative Projects and Networks of Excellence

• **DEM =** Demonstration - applicable for collaborative projects and Research for the Benefit of Specific Groups

• **MGT =** Management of the consortium - applicable for all funding schemes

• **OTHER =** Other specific activities, applicable for all funding schemes

• **COORD =** Coordination activities - applicable only for CAs

• **SUPP =** Support activities - applicable only for SAs

55. Lead beneficiary number

Number of the beneficiary leading the work in this work package.

56. Person-months per work package

The total number of person-months allocated to each work package.

57. Start month

Relative start date for the work in the specific work packages, month 1 marking the start date of the project, and all other start dates being relative to this start date.

58. End month

Relative end date, month 1 marking the start date of the project, and all end dates being relative to this start date.

59. Milestone number

Milestone number:MS1, MS2, ..., MSn

60. Delivery date for Milestone

Month in which the milestone will be achieved. Month 1 marking the start date of the project, and all delivery dates being relative to this start date.

61. Deliverable number

Deliverable numbers in order of delivery dates: D1 - Dn

62. Nature

Please indicate the nature of the deliverable using one of the following codes

R = Report, **P** = Prototype, **D** = Demonstrator, **O** = Other

63. Dissemination level

Please indicate the dissemination level using one of the following codes:

• **PU =** Public

• **PP =** Restricted to other programme participants (including the Commission Services)

• **RE =** Restricted to a group specified by the consortium (including the Commission Services)

• **CO =** Confidential, only for members of the consortium (including the Commission Services)

• **Restreint UE =** Classified with the classification level "Restreint UE" according to Commission Decision 2001/844 and amendments

• **Confidentiel UE =** Classified with the mention of the classification level "Confidentiel UE" according to Commission Decision 2001/844 and amendments

• **Secret UE =** Classified with the mention of the classification level "Secret UE" according to Commission Decision 2001/844 and amendments

64. Delivery date for Deliverable

Month in which the deliverables will be available. Month 1 marking the start date of the project, and all delivery dates being relative to this start date

65. Review number

Review number: RV1, RV2, ..., RVn

66. Tentative timing of reviews

Month after which the review will take place. Month 1 marking the start date of the project, and all delivery dates being relative to this start date.

67. Person-months per Deliverable

The total number of person-month allocated to each deliverable.