Delivering the promise of Gaia

APPENDIX

Response to ESA's Announcement of Opportunity

This document is not formally part of the response to the CU9-AO.

It contains complementary information related to the detailed planning of CU9 tasks which is not definitive and may thus change in the actual implementation.

date of issue 09 January 2013
status Submitted to ESA

Compiled by:

Editors: X. Luri and W. O'Mullane

Cover art design by C. Brown

Contents

A Trace of AO requirements **4**

B Management Team **5**

B.1 Xavier Luri, *CU9 leader, WP970 manager* 5

B.2 William O'Mullane, *CU9 Technical manager* 5

B.3 Floor van Leeuwen, *Documentation, WP920 manager* 6

B.4 Anthony Brown, *Documentation, WP920 co-manager* 6

B.5 Nigel Hambly, *Archive architecture, WP930 manager* 7

B.6 Jesus Salgado, *Archive architecture, WP930 co-manager* 8

B.7 Frédéric Arenou, *Validation, WP940 manager* 8

B.8 Paola Di Matteo, *Validation, WP940 co-manager* 9

B.9 Emmanuel Mercier, *Operations, WP950 manager* 9

B.10 Jose Hernandez, *Operations, WP950 co-manager* 10

B.11 Stefan Jordan, *Outreach, WP960 manager, WP970 co-manager* 11

B.12 Eduard Masana, *Outreach, WP960 co-manager* 11

B.13 André Moitinho, *Visualisation, WP980 manager* 12

B.14 João Alves, *Visualisation, WP980 co-manager* 12

C Detailed WP description **14**

C.1 WP910 14

C.2 WP920 18

C.3 WP930 24
C.4 WP940 36
C.5 WP950 51
C.6 WP960 68
C.7 WP970 78
C.8 WP980 96

D References 111

A Trace of AO requirements

The following table points to the sections replying to the AO requirements in section 3 of (TJP-013)

AO Section	Reference or Section
3.1 Management and Coordination requirements	7, 10.1
3.2 Documentation requirements	10.2
3.3 Architecture and Technical Development requirements	10.3
3.4 Validation requirements	10.4
3.5 Operations requirements	10.5
3.6 Education and Outreach requirements	10.6
3.7 Science enabling applications requirements	10.7
3.8 Visualisation requirements	10.8

B Management Team

A brief vita of each of the workpackage leaders and deputies is given here.

B.1 Xavier Luri, *CU9 leader, WP970 manager*

After obtaining a degree on physics, X. Luri did his Ph.D on astrophysics, developing new statistical methods for the exploitation of the Hipparcos catalogue. This was his first contact with space astrometry, a field in which he has been working ever since. He also has wide experience in software development and computer systems (acquired from academic formation and practical experience in software projects, both in private companies and public research centres) and space systems. Dr. Luri presently holds a tenure position at the Department of Astronomy and Meteorology of the University of Barcelona, where he teaches Astronomy, Mathematics and Statistics. He is member of the board of the Spanish Astronomical Society, where he also chairs the committee in charge of outreach, and has been vice-president of the Spanish Institute of Navigation.

A member of the Gaia Science Advisory Group since its early stages and then later the Gaia Science Team, Dr. Luri led the Simulation Working Group, providing simulation support to the mission design activities, until it was replaced by the Coordination Unit 2 of the DPAC, which he presently manages. As such, he is also a member of the DPACE.

B.2 William O'Mullane, *CU9 Technical manager*

W. O'Mullane has an academic background in computer science and over a decade of experience in space missions and space science. He has worked at ESOC on the SCOS-II system and contributed to the production of the CDROMs for the Hipparcos catalogue. He has created acclaimed visualisation software for the Hipparcos catalogues and contributed to the quality assurance tools for the Guide Star Catalogue. More recently he has worked on image processing and catalogue interrogation software for the Sloan Digital Sky Survey at the Johns Hopkins University. O'Mullane was also a group leader in the US National Virtual Observatory (NVO) contributing to the overall architecture and technology decisions made by NVO. Outside space science he also has industrial experience in high availability dynamic websites with high user loads. He has a PhD in Physics from the University of Barcelona, his area of expertise is in the Gaia astrometric solution.

W. O'Mullane became involved in Gaia as Hipparcos wrapped up in 1997. He was a member of the DACC (DPACE precursor) since its formation and is now a DPACE member. He has been the ESAC Gaia manager since the team was created in August 2005.

As CU1 leader O'Mullane takes responsibility for the overall architecture of the processing system as well as a few core software items such as the MDB. He has been setting down ideas for the architecture of the DPAC systems and the AGIS since 1997. He has been chair of GAP (Gaia Archive Preparation) since it was set up.

B.3 Floor van Leeuwen, *Documentation, WP920 manager*

Floor van Leeuwen was trained in astrometry and photometry at Leiden University, where he was a student of S. Vasilevskis and A. Blaauw. His PhD thesis at Leiden University (1983) was on an astrometric and photometric study of the Pleiades cluster. He was involved in the Hipparcos mission at the Royal Greenwich Observatory, first under Andrew Murray, and from 1985 onwards as team leader, till the publication of the Hipparcos data in 1997. He published several key papers on the Hipparcos mission (a major review in Space Science Reviews in 1997, papers on specific use of the astrometric data, and papers on the derivation of astrometric parameters for open clusters). He completed a re-reduction of the Hipparcos astrometric data in 2007, which, through taking into account dynamical properties and peculiarities of the satellite, has reduced the noise levels on the astrometric data for the brightest stars by a factor 3 to 5, reaching for the first time ever parallax accuracies of 0.1 mas. He has also been data processing manager for the Planck satellite commitments in Cambridge from 1998 till 2006. He has been a member of the Hipparcos Science Team from 1984 till 1997, and of the Gaia Science team 2003 till 2009.

The CU structure for the data processing was first proposed by him at a meeting in Cambridge in February 2005, and then presented in a document which was further developed by Michael Perryman and later by Francois Mignard and Coryn Bailer-Jones. As a member of the DACC he helped structure the DPAC. In the UK he has been leading the (successful) grant application for the Gaia data processing. Since the start of DPAC he has been, and remains as, the CU5 leader. In 2007 he took the intitiative for, and has led since, the Gaia radiation task force, looking at ways to mitigate the effects of radiation damaged CCDs on the astrometric and photometric data.

B.4 Anthony Brown, *Documentation, WP920 co-manager*

Anthony Brown was trained in astronomy at Leiden University. His PhD thesis, directed by P.T. de Zeeuw, at the same university in 1996 concerned a study of the nearby OB associations using ground based data which was collected in preparation for the Hipparcos mission results. Subsequently he worked with M.A.C. Perryman on a detailed study of the Hipparcos data of the Hyades cluster and was closely involved with the Hipparcos census of the nearby OB associations carried out by J.H.J. De Bruijne and R. Hoogerwerf. During his time as a postdoc at UNAM (Mexico, 1997–1999) he worked on the detection of the remnants of disrupted satellite galaxies in the halo of our galaxy from the Gaia catalogue. This was done by creating a large and detailed Monte Carlo simulation of the Gaia catalogue with a realistic number of entries (3.5×10^8 stars). From 2001–2005 he was involved in the development of the SINFONI instrument for ESO's VLT. He worked on detailed simulations of the adaptive optics module for this instrument.

Brown has been involved with Gaia since 1997 and started out by contributing to the science case for the Concept and Technology Study Report. In 1999 he wrote one of the first detailed simulations of Gaia measurements at the detector level. He later contributed to the BP/RP simulation code within the Gaia Instrument and Basic Image Simulator (GIBIS) and was deeply involved in CU5 planning and management, leading the BP/RP preprocessing development unit from 2006 to 2012. Brown was a member of the Gaia Data Access and Analysis System Steering Committee, where he was responsible for overseeing the photometric aspects of the Gaia data processing prototype. In 2005–2006 he was a member of the DACC and at the end of 2005 he joined the Gaia science team. In 2012 he became the DPACE chair.

B.5 **Nigel Hambly,** *Archive architecture, WP930 manager*

N. Hambly is a founder member and science archive architect for the Wide Field Astronomy Unit (WFAU) at the University of Edinburgh, and is an experienced survey astronomer and developer. WFAU has a strong track record in the curation and dissemination of large-scale survey data sets consisting of 10s of billions of rows of catalogue data occupying 10s of terabytes: these include the legacy all-sky digitised Schmidt photographic optical surveys culminating in the terabyte-scale SuperCOSMOS Science Archive[1]; the current generation of infrared sky surveys in both hemispheres, namely the United Kingdom Infrared Deep Sky Survey presented in the WFCAM Science Archive[2] and the VISTA Public Surveys presented in the VISTA Science Archive[3]; and most recently the Gaia–ESO Survey[4], a large–scale ground–based spectroscopic public survey with the VLT to complement the Gaia mission. WFAU serves these data sets through both state-of-the-art interactive user interfaces and VO services[5]. Over the last decade, he has gained much experience in the complete archive project lifecycle, from requirements gathering and analysis, system analysis and design including aspects of hardware design, software engineering to professional standards, and archive systems deployment and maintenance. Throughout this work, he has remained committed to a science-driven philosophy aimed at maximising scientific impact through appropriate use of information technologies. Both the SSA and the WSA archive systems were de-

[1] `http://surveys.roe.ac.uk/ssa/`
[2] `http://surveys.roe.ac.uk/wsa/`
[3] `http://surveys.roe.ac.uk/vsa/`
[4] `http://surveys.roe.ac.uk/ges/`
[5] `http://vo.roe.ac.uk/index.htm`

signed by Hambly (Hambly et al., 2004, 2008). He has much experience in managing small research and development teams for both hardware and software processing and archiving systems.

Hambly is a member of the Gaia Archive Preparation group within DPAC, as well as being a senior developer for the Gaia DPAC Coordination Unit 5 (WFAU plays a key role in the Gaia DPAC collaboration, currently contributing ~ 3 FTEs to critical aspects of the core processing in the ground segment for Gaia data). He has responsibilities for core processing tasks concerned with bias, background and radiation damage calibration and removal.

B.6 Jesus Salgado, *Archive architecture, WP930 co-manager*

After obtaining his degree in Physics in 1992, J. Salgado started investigations in plasma physics simulation, in particular in the area of Plasma Confinement within the Complutense University Plasma Physics Group in Madrid where he published his Masters thesis in 1995 on Simulations of Tokamaks using Genetic Algorithms.

In 1995, he started to work in operations of ESA's ISO satellite and, afterwards, for ESA XMM-Newton operations.

In 2002, he joined the ESA Science Archives and VO Team, where he has been linked to the data access components of the Science Archives, designing, implementing and coordinating the server side of the Science Archives and the machine interfaces.

At the same time, he has been involved in Virtual Observatory projects like the IVOA (International Virtual Observatory Alliance) where he has been Data Access Layer Working Group vice chair, Data Model Working Group vice chair and he is currently the Data Model Working Group Chair that handles the data model standards to be used within the VO. On planetary data, he is member of the Technical Expert Group of the IPDA (International Planetary Data Alliance), leading the efforts of the definition of a universal protocol to access planetary science data (Planetary Data Access Protocol, PDAP)

Salgado has started to work recently within the Gaia CU9 as technical coordinator of the design and implementation of the Gaia Archive Core Systems. He is also the Core Development Team coordinator within SAT, ensuring the proper synergy and reuse of elements between all the Science Archives at ESAC.

B.7 Frédéric Arenou, *Validation, WP940 manager*

Senior research engineer on a permanent position in the french National Center for Scientific Research (CNRS), F. Arenou has been working since 1983 at the Paris-Meudon

Observatory after an academic background both in computer sciences and in astronomy. His Ph.D. which he did in Meudon was devoted to the validation of the preliminary Hipparcos data. During the Hipparcos era, he worked successively on the construction and database management of the Hipparcos Input Catalogue, the validation of the astrometric data, the merging of the astrometric single star data, several parts of the Double and Multiple Systems Annex and finally the scientific exploitation. His scientific interests are related to astrometry, multiple stellar and substellar systems and statistical analysis of astronomical data.

Involved in Gaia since 1997, he was member of the Gaia Science Team from 2001 to 2005 and team leader of two Working Groups, the Gaia On-Board Detection WG, and the Double and Multiple Star WG. In the on-board data handling area, he delivered the prototype on-board software Pyxis during the industrial competitive phase and then supported Astrium. Concerning the data processing within DPAC, he is involved in the simulation activities with the CU2 and CU4 Coordination Units, he is in charge of the CU4 post-processing for double and multiple stars, and he is a member of the CU4 steering committee.

B.8 Paola Di Matteo, *Validation, WP940 co-manager*

Paola Di Matteo was trained in astronomy at the University of Rome La Sapienza, where she defended her thesis in 2005. The subject of her thesis concerned the study, by means of N-body simulations, of the evolution of globular clusters in the galactic tidal field, focusing in particular on the morphology and kinematic properties of tidal tails that are formed around clusters in orbit in the host galaxy.

Subsequently (2005-2011), during her time as a postdoc at the Paris Observatory, she worked with F. Combes and M. Lehnert on the dynamical evolution of galaxies, and in particular on the role galaxy interactions play in shaping the observed properties of galaxies (structure, kinematics, stellar populations). Since 2011, she is associate astronomer at the Paris Observatory, where she continues in developing N-body models to constrain the main mechanisms that rule the evolution of the Milky Way and other disk galaxies, and that shape their components (bulges, disks, halos).

Paola Di Matteo has started to work recently within the Gaia DPAC. She is co-leader of the Validation Work Area in the Gaia Archive Preparation (GAP) and she is co-responsible of the Gaia Spectroscopic Alerts (CU6).

B.9 Emmanuel Mercier, *Operations, WP950 manager*

E. Mercier has an academic background in computer sciences and project management and over a decade of experience in space missions. He started his career in the video game industry in 1995 and was in charge of managing large teams of developers, testers and producers. Then he acted as technical project manager and developed web based applications for an IT company. In 2001 he started to work for the CNES, responsible for a team of Project Management Officers. During more than 8 years, E. Mercier supported the major CNES space missions like ATV-CC, Corot, Curiosity Chemcam, Pleiades, Jason, and Microscope. He also collaborated with the quality assurance directorate and wrote the Scheduling Management Plan of CNES and the Product Management Plans for the microsatellites and altimetry departments.

E. Mercier became involved in Gaia early in 2009 as the Gaia DPAC Project Coordinator. On behalf of DPACE, he is responsible for and empowered to take decisions concerning all aspects of the DPAC implementation, development and operations. That duty implies some strong diplomatic and communication skills. Head of the Project Office, his team is proficient in evaluating the progress of the different European teams and supporting them in the management of their activities. The Project Office ability in deploying the project management practices and tools allows control of the status of the activities described over three hundred work packages, and keeps track of the financial resources. Finally, one of E. Mercier's key roles is to be able to consolidate this detailed management information and summarize it for various committees. In particular he reports on a six months basis to a board representing the European Space Agencies.

B.10 Jose Hernandez, *Operations, WP950 co-manager*

After obtaining his degree in chemistry in 1991, J. Hernandez went to ESTEC with a fellowship where he worked as a SW engineer building a simulator of the ISO satellite. This was used to verify and validate the AOCS, and the AOCS system had to be amended as a result of these tests. After the fellowship, Hernandez joined the ISO SOC team in ESTEC where he contributed to the Camera and Calibration uplink systems, when the ISO operations finished around 1997 he started to work in the development of the ISO data archive where he played a major role in architecture decisions, language choice (Java), and interface look–and–feel. Hernandez was also part of the ESAC archive team working on the adaptation of the archive to new missions (XMM-Newton, Integral, Mars Express, etc). He was in charge of the user front and archive business logic, often interacting with the astronomers and the archive scientist.

Hernandez has worked at the Gaia SOC since 2005. He is in charge of the Gaia Main Database system. This work involves agreeing with the CUs and DPCs the definition of the data model, data flow across DPAC systems and DPCs, data management, tracking and versioning etc. Hernandez has also been heavily involved in the development and scientific validation of several aspects of AGIS. He also contributes to issues related to instrument configuration and calibration data access in DPAC systems. Hernandez is also in charge of the implementation of the Local Plane Coordinates and has been working closely with the CU3 and CU4 leaders on its validation.

B.11 Stefan Jordan, *Outreach, WP960 manager, WP970 co-manager*

Stefan Jordan obtained his diploma in physics in 1985 and his PhD in astronomy in 1988 at the University of Kiel, Germany. In Kiel he also received his habilitation (postdoctoral lecture qualification) in 1997, which was renewed at the Universities of Tübingen, Germany, in 2004, and Heidelberg, Germany, in 2006. Additionally, he had a position at the Louisiana State University in Baton Rouge, USA, in 1988, and in Göttingen, Germany, in 2001.

For many years his main scientific work dealt with stellar atmospheres and radiative transport. His speciality is in the analysis of magnetic white dwarfs and the analysis of their flux and polarisation spectra in order to determine the magnetic field structure of these stars. This work included observations with large observatories on ground and in space as well as theoretical modelling.

Since 2001 he was involved in the preparation of the Diva project, a proposed small German astrometry satellite which should have been launched well before Gaia. After this project was cancelled due to the lack of sufficient financial support, Stefan Jordan joined the Gaia team at the ARI/ZAH in Heidelberg. He has mainly worked on the Gaia First-Look with the special emphasis on the ODAS (One-Day Astrometric Solution) and the First-Look feedback for the radiation damage mitigation for the Gaia CCD measurements. In Heidelberg, he was also involved in the German Grid-computing project (AstroGrid).

Stefan Jordan has given many scientific and public talks on the Gaia mission. In Kiel he was the head of an amateur astronomy group for seven years and has given and organised many public astronomical talks and guided tours with telescopes. Moreover, he had a monthly five-minute radio feature for one year dealing with the sky of the month.

B.12 Eduard Masana, *Outreach, WP960 co-manager*

Eduard Masana obtained his degree in physics in 1992 at the University of Barcelona. He also earned his PhD on the characterisation of Stellar Populations in the Milky Way from a kinematic point of view, developing a new method to determine effective temperatures from IR photometry. He has wide experience as a software developer, in part acquired during his work in the Gaia simulation tasks in CU2. Presently he is a researcher at the Institut d'Estudis Espacials de Catalunya (ICCUB-IEEC).

He became involved in Gaia in 1999, starting to work in the development of the Gaia simulator. Since then, he is in charge of GASS, the simulator of the Gaia telemetry stream, first as a member of the Simulation Working Group and afterwards as the manager of the DU5 of CU2.

E. Masana is also involved in astronomy outreach activities. He has participated in the live webcast of several astronomical events, maintenance of a web portal devoted to astronomy and has given many talks and courses for the general public.

B.13 André Moitinho, *Visualisation, WP980 manager*

A. Moitinho received his PhD in Astrophysics from the University of Granada in 1999. His thesis was on the connection between large scale star formation and the spatial structure of the Galactic disc. Since then, he has observed hundreds of star cluster fields for studying the structure and evolution of the Milky Way as well as the late stages of low mass star formation. Dealing with extensive data sets has led to work on data-mining, development of photometric pipelines and on interactive visualisation of large data sets. Observational work also made him interested in astronomical instrumentation, having been in charge of opto-mechanic and cryogenic systems of instruments designed for the VLT and the GTC. He is currently the head of the *Instrumentation and Computing for Space and Earth Observations* group at the University of Lisbon. He has served on several committees including the ESA Astronomy Working Group and is presently a member of the ESO Scientific and Technical Committee and of the Portuguese delegation to the United Nations Committee on the Peaceful Uses of Outer Space.

Working in the DPAC since 2006, his activity has been mostly in CU7. He manages a WP for assessing the biases in the variability parameters output by the CU7 pipeline. The end product is a statistical description of the Gaia variability data as a function of the time sampling (which depends on the satellite's scanning law), of the uncertainties in the measured fluxes and of the variability characterisation algorithms employed. Within CU7, has also contributed to the assessment of non-supervised algorithms for classification of variability types. On the administrative side, he is the PI of the funding project for the Portuguese participation in the DPAC, which includes five research centres and has members in CU2, CU3, CU4, CU5 and CU7. He is also a member of the steering committee of the GREAT ESF grant.

B.14 João Alves, *Visualisation, WP980 co-manager*

João Alves is a full professor at the University of Vienna and an expert in star formation research in the Milky Way, in particular on the structure of molecular clouds and the initial conditions to star formation. He has been developing for the last decade an alternative Interstellar Medium density tracer technique that is today the most robust way to study the density structure of molecular clouds. Before moving to the University of Vienna, João Alves was ESO Fellow from 1998 to 2001, and ESO staff until 2006 (where was the head of the ESO-VISAS). In 2006 he moved to Southern Spain to take on the directorship of the Max-Planck / CSIC German-Spanish Astronomical Center in Calar Alto. In 2010 he became full professor of the University of Vienna. João Alves has published 94 refereed papers and edited 3 books.

João Alves is a member of the Gaia Archive Preparation group within DPAC. His interests in data visualisation go back to his PhD thesis work where he was interested in the visualisation of the extended sources, such as dark molecular clouds. Today he has a strong interest in the global visualisation of the Gaia archive and in the development of the required data simplifiers. His planned contributions include the development of algorithms from Gaia data that would allow the construction of the most accurate 3D map of the local ISM.

C Detailed WP description

This section presents the detailed description of the WBS at the sub-WP level.

C.1 WP910

Gaia DPAC WP:		GWP-M-911-0000
Title: General management		
Provider(s): Univ. Barcelona		
Manager(s): X. Luri & G. Gracia		
Start: 2013	**End**: 2022	**Total Effort**: 120 SM
Objective: Overall management of CU9		
Tasks: Global management of CU9, including coordination with the rest of DPAC, generation and approval of documentation and software, reporting to DPACE, ESA and the external advisory board. This WP also includes the QA activities.		
Input:		
Output:		
Deliverables: • ECSS documentation • Progress reports • Organisation of meetings, including at least: kick-off meeting and one plenary meeting per development cycle • Development schedule • Formal delivery of archive system at each catalogue release		

Dependencies:
Interfaces:
Remarks:

Gaia DPAC WP:		GWP-M-912-0000
Title: Technical management		
Provider(s): ESA		
Manager(s): W. O'Mullane		
Start: 2013	**End**: 2022	**Total Effort**: 40 SM
Objective: Technical management of CU9		
Tasks: The main technical work of CU9 will be done in WP930 and its sub work packages. There is a deal of coordination to be done between these areas themselves and other work packages. In particular interplay between Visualisation WP980 and the architecture WP930. Also data validation WP940 will ideally use tools developed within WP930 and WP980. But of course technical matters underpin practically all work in CU9. Decisions and trade offs will need to be made at many levels. In addition technical management includes enacting the Product Assurance Plan and approving the tools for milestone tracking and planning as in other CUs. Hence assistance with the SDP etc. Selection of video conferencing for meetings.		
Input:		
Output: Technical guidance in all areas of CU9.		
Deliverables:		
Dependencies:		
Interfaces:		
Remarks:		

Gaia DPAC WP:		GWP-M-913-0000
Title: Science scenarios and requirements mangement		
Provider(s): IoA Cambdridge		
Manager(s): N. Walton		
Start: 2013	**End**: 2022	**Total Effort**: 20 SM
Objective: Ensure a user-driven development of CU9 systems		
Tasks: WP913 will over see science management of CU9, ensuring that all CU9 deliverables are meeting the science requirements for CU9 and all science products released to the community are fully science quality assured.		
Input:		
Output:		
Deliverables: • Progress reports • Organisation of science meetings, including at least sessions within the CU9 kick-off meeting and plenary meeting per development cycles. • Organisation of the 'beta testing team' (see 10.1).		
Dependencies:		
Interfaces:		
Remarks:		

C.2 WP920

Gaia DPAC WP:		GWP-M-921-0000
Title: Documentation: management		
Provider(s): University of Cambridge & Leiden University		
Manager(s): F. van Leeuwen, A.G.A. Brown		
Start: 2013	**End**: 2022	**Total Effort**: 22 SM
Objective: Management of CU9 documentation		
Tasks: 1. Coordinating the documentation effort for CU9 2. Overseeing the editorial team 3. Ensure embedding of documenation in data access facilities 4. Stay abreast of developments in documentation facilities		
Input:		
Output:		
Deliverables: • ECSS documentation • Progress reports • Organisation of meetings and telecons • Documentation production schedule • Formal delivery of documentation at each Gaia data release		
Dependencies:		

Interfaces: Coordinate with WP 930 the linkage between documentation and data at the archive level.
Remarks:

Gaia DPAC WP:		GWP-M-922-0000
Title: Documentation editorial team		
Provider(s): University of Cambridge, Leiden University, ESAC, DPAC PO, ObsPM, University of Barcelona, OCA		
Manager(s): F. van Leeuwen & A.G.A. Brown		
Start: 2013	**End**: 2022	**Total Effort**: 80 SM

Objective:
Synthesizing and editing the CU9 documentation

Tasks:

1. Reviewing CU-level documentation, including the documentation produced at the CU9 work package level. This means checking the documentation for completeness (according to points 1–3 in 10.2), readability, language, clarity, etc.
2. Develop and enforce guidelines for the homogenisation of all CU9 documentation.
3. Process the finished documentation to make it available in various electronic formats (PDF, HTML, etc).

Input:
Documentation contributions from each CU

Output:
Documentation for each Gaia data release

Deliverables:

- Homogenized CU9 documentation in various electronic formats at each Gaia data release
- Integration of the documentation in the data access facilities

Dependencies:
Depends on WP 923 for the CU-provided documentation.

Interfaces:
Remarks:

Gaia DPAC WP:		GWP-M-923-0000
Title: Collect CU documentation		
Provider(s): DPAC PO		
Manager(s): S. Els		
Start: 2013	**End:** 2022	**Total Effort:** 12 SM
Objective: Collecting the documentation contributions from CUs 1–9		
Tasks: 1. Assist with the CU9 documentation production planning and scheduling. 2. Liase with the documentation workpackage managers in CUs 1-8 to collect the CU9 documentation contributions. 3. Monitor the progress of the documentation effort. 4. Collect the documentation contributions from all CUs. 5. Assist in the enforcing of documentation guidelines.		
Input: CU level documentation, guidelines on documentation		
Output: Bundled CU level documentation		
Deliverables: • CU level documentation for input to WP-922.		
Dependencies:		
Interfaces:		

Remarks:

C.3 WP930

Gaia DPAC WP:		GWP-M-931-0000
Title: Archive architecture: management		
Provider(s): Univ. Edinburgh & ESAC		
Manager(s): N.Hambly & J.Salgado		
Start: 2013	**End**: 2022	**Total Effort**: 20 SM
Objective: Management of archive developments		
Tasks: 1. Co-ordinating developments 2. Reporting on developments 3. Updating collaboration resources (planning documents, Wikis etc.) 4. Organising, chairing and minuting meetings		
Input: Procedures as specified in Sect. 10.1		
Output: Coordination of WP930		
Deliverables: • ECSS documentation • Progress reports • Organisation of meetings and telecons • Documentation production schedule • Formal delivery of documentation at each Gaia data release		
Dependencies:		

Interfaces: WP910
Remarks:

Gaia DPAC WP:		GWP-T-932-0000
Title: Archive architecture: core systems		
Provider(s): ESAC		
Manager(s): P. Osuna		
Start: 2013	**End:** 2022	**Total Effort:** 288 SM
Objective: Delivery of Gaia archive core systems		

Tasks:

1. Data management: Data management activities involve all subsystems dealing with the access to the Gaia catalogue (queries and so on), data products associated metadata, pre-computed statistics, etc.

    - TAP: Technology study: Web services (REST) technologies study, SQL grammar frameworks, etc. The effort estimation may vary depending on the outcome of this activity. Software reuse assessment: CADC and other open source implementations study. Architecture and design: Probably the most important activity of the component. It should produce a flexible architecture that can meet the Interrogator requirements without major changes. Implementation: Lower level design and development activities (incremental) of a TAP service. Maintenance: SPR analysis and implementation.
    - DBMS for catalogue: Parallel/Distributed DBMS technology study. PostgreSQL, PL/Proxy, GridSQL, SciDB, InterSystems Caché, Greenplum, MongoDB, Cassandra, Hive, etc. Data Architecture on DBMS. Data layout (partitioning strategies, query analysis, optimisations, etc).
    - Interrogator: Interrogator design and development (including ingestion, integration with other services like VOSpace, security concerns, fast-track data delivery tools development, complex schemas treatment, etc).

2. Ingest: The Ingest involves all subsystems and interfaces required for Gaia data ingestion (catalogue, data products, etc) into the archive from the SOC. It encompasses the data contents familiarisation (understanding of data sets), definition of ICDs, development of ingestion subsystems, etc.

3. Archival storage:

    - VOSpace: Technology study of Web Services technologies (SOAP and REST), data transfer protocols and formats, security, etc. Software reuse assessment of CADC and other open source implementations. Architecture and design. Implementation: Lower level design and incremental development.
    - The service back-end storage infrastructure. Scalability and performance are the key points.

Tasks: 4. Cloud services: • MapReduce for Big Data analysis. Everything related to raw data treatment. Low level data (transits, observations, etc) will probably go to a MapReduce infrastructure (see the Hadoop ecosystem) as it provides flexibility to run different workflows or data analysis workloads and it is the most scalable solution so far. This will also be used for pre-computed statistics generation (like histograms, healpix density maps for 3D visualisation, etc). • IaaS and Grid computing. Technology study of CANFAR services, Univa GridEngine, scalability on local data centre and on Amazon, Eucalyptus, enStratus, RESERVOIR (FP7), StratusLab, etc. Services reuse/collaboration assessment: Mainly an integration of the different services available. Development and integration of services: Any development needed for integrating the different services suitable for the use cases. • SaaS: Technology study: Investigation of the state-of-the-art for this new discipline and its applications to the Gaia Archive. Astro-Application Container: Devise some kind of container that allows users (registered/consortia/paying users or whatever that means) to deploy their applications on the data centre so that they can be used by the community.
Input:
Output:
Deliverables:
Dependencies: WP933
Interfaces: WP920, WP960 and WP980
Remarks:

Gaia DPAC WP:		GWP-T-933-0000
Title: Archive architecture: consortium/ESAC–SAT interface		
Provider(s): Univ. Edinburgh		
Manager(s): N. Hambly		
Start: 2013	**End**: 2017	**Total Effort**: 3 SM
Objective: Definition, documentation and management of agreed interaction between CU9 and ESAC–SAT		
Tasks: 1. Development of a coordination agreement with ESAC–SAT 2. Formal documentation of the interface coordination agreement		
Input: Meeting(s) between CU9 managers and consortium representatives and ESAC–SAT		
Output: Implementation coordination		
Deliverables: Consortium/ESAC–SAT Interface Coordination Document		
Dependencies: WP931, WP932		
Interfaces: WP972, WP973, WP974, WP983		
Remarks:		

Gaia DPAC WP:		GWP-T-934-0000
Title: Archive architecture: database collaboration		
Provider(s): AIP, Univ. Edinburgh, ASDC–INAF		
Manager(s): H. Enke, N. Hambly, P. Marrese		
Start: 2013	**End**: 2017	**Total Effort**: 59 SM
Objective: Delivery of server–side components related to back–end data stores		

Tasks:

1. Although the archival core system is hosted at ESAC, this will not be able to cover the demands of the community working with the Gaia archive. Thus, a number of sites will run either a partial or complete mirror of the archive, and offer specialized services on top of these archives. This has to be coordinated and carefully planned ahead, including ways to keep the various archives in a sufficiently synchronized state. For example, AIP has some experience with running huge databases serving the astronomy community and plans to run at least a partial mirror of the Gaia archive.

2. Development and testing of db engines, SQL and NOSQL DBMS, integration of virtualisation technologies into the archive system, multicolumn indexing (ASDC–INAF)

3. The efficiency of distributed queries can be improved 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 will assess, through quantitative analysis – and, where possible, direct experimentation – the optimal configuration of the multi-wavelength data sets required for the scientific exploitation of Gaia. For example, to determine which external catalogues should be co-located with a copy of the Gaia archive, for which should "cross-neighbour" tables be pre-computed to facilitate queries between data sets that remain geographically separated, and for which can cross-matches be performed on-the-fly with sufficient speed (Univ. Edinburgh)

Input: Requirements via WP910
Output: Server–side subsystems and components for integration at ESAC
Deliverables: Technical reports, SW components and associated documentation
Dependencies: WP933
Interfaces: WP972, WP973, WP974, WP983
Remarks:

Gaia DPAC WP:		GWP-T-935-0000
Title: Archive architecture: interface design (TAP+Server side)		
Provider(s): AIP, ARI/ZAH, Univ. Edinburgh, ASDC–INAF		
Manager(s): S. Jordan, H. Enke, N. Hambly, P. Marrese		
Start: 2013	**End**: 2017	**Total Effort**: 100 SM
Objective: Delivery of server–side interface layers		

Tasks:

1. On top of the databases, several services will run, either using VO protocols, e.g. TAP, or other webservices, e.g. for scripted retrieval of SQL query results. AIP has built query facilities for various surveys (e.g. RAVE) and are building a framework for interfacing databases with large astrophysical data via webservices. Features of AIP's framework include modules such as job management and scheduling, result table transformation (e.g. CSV, binary, VOTable), and data access APIs (contributor: AIP)

2. The VO-Dance suite provides a lightweight method of publishing data to the VO. This package will include a TAP implementation that support the main DBMS (Mysql, Oracle, Postgress). Since TAP is under evolution, Version 1.0 was released on 2010, maintenance, update and development of new version based on experiences on huge amount of data like Gaia is the main task of Trieste group.

3. TAPFactory 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. (contributor: Univ Edinburgh)

4. Integration of documentation (e.g. relational schema browsers etc.) to facilitate interactive usage, and integration of metadata to facilitate VO application functionality

Input:
Requirements via WP910

Output: Server–side subsystems and components for integration at ESAC
Deliverables: Technical reports, SW components and associated documentation
Dependencies: WP933
Interfaces: WP972, WP973, WP974, WP983
Remarks:

Gaia DPAC WP:		GWP-T-936-0000
Title: Archive architecture: correlation function		
Provider(s): ESAC, Univ. Lund		
Manager(s): W. O'Mullane, L. Lindegren		
Start: 2013	**End**: 2017	**Total Effort**: 30 SM
Objective: Provide the means for correct propagation of error correlations		
Tasks: 1. Investigate correlation algorithms, selecting the best according to accuracy/speed trade–off 2. Implement algorithm and any associated interface layer(s) 3. Document the solution		
Input: Algorithm specifications, e.g. as described in Holl & Lindegren (2012) and Holl et al. (2012)		
Output: Server–side subsystems and components for integration at ESAC		
Deliverables: Technical reports, SW components and associated documentation		
Dependencies: WP932, WP933		
Interfaces: WP972, WP973, WP974, WP983		
Remarks:		

C.4 WP940

Gaia DPAC WP:		GWP-M-941-0000
Title: Validation: management		
Provider(s): Observatoire de Paris		
Manager(s): F. Arenou, P. Di Matteo		
Start: 2013	**End**: 2022	**Total Effort**: 20 SM
Objective: Management of validation WP		
Tasks: The objective of this work package is to ensure that WP940 meets its objectives on schedule. Tasks will include: • coordinating and supervising activities to be carried out; • monitoring project progress, quality and timing of deliverables; • Coordinating with other CUs to avoid duplicate verifications; • Implementing the data right policy in practice; • Reporting back to the CU9 executive board		
Input:		
Output:		

Deliverables: • ECSS documentation • Progress reports • Organisation of meetings and telecons • Documentation production schedule • Formal delivery of documentation at each Gaia data release
Dependencies:
Interfaces:
Remarks:

Gaia DPAC WP:		GWP-T-942-0000
Title: Validation: scenarios		
Provider(s): Obs. de Paris, Univ. Groningen, Uni. Barcelona, Obs. de la Côte d'Azur		
Manager(s): C. Fabricius		
Start: 2013	**End**: 2022	**Total Effort**: 85 SM
Objective: Basic tests and elaborate test scenarios		

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

- Blind automated tools for checking that 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).
- Searching for cross correlations between the different variables and for consistency between different data. Verify that astrometric, photometric and spectroscopic data are consistent between each other. For example, parameters from photometry should be consistent with those from spectroscopy; trigonometric parallaxes consistent to photometric ones.
- Quantify the kind of problems that could occur (calibration or instrumental problems; classification errors; data Processing shortcuts or approximate models) and what consequences this would have on observed parameters. To give examples: a annual thermal or calibration effects would introduce a parallax bias; a 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 work areas would then be the following:

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

Input:

Output:

Deliverables: • Definitions of scenarios • Software implementations • Results from database queries • Reports
Dependencies:
Interfaces:
Remarks:

Gaia DPAC WP:		GWP-T-943-0000
Title: Validation: models		
Provider(s): Observatoire de Besançon, Observatoire de Paris, University of Barcelona, ASDC-INAF, Observatoire de la Côte d'Azur		
Manager(s): A. Robin		
Start: 2013	**End**: 2022	**Total Effort**: 156 SM
Objective: Using 'Universe models' to validate the Catalogue		

Tasks:

The Universe Model, initially based on the Besançon Model of our Galaxy, has been complemented in the CU2 DPAC Coordination Unit by an extinction model, a multiple stars model and a variability models, etc., and it now represents the best simulated sky one could hope to test the DPAC algorithms against.
These models can in turn be used to validate the Gaia data, by:

- Extracting the expected statistical properties of the Catalogue from the models: distribution, confidence intervals and correlations between parameters, by object type, by region and by time;
- Comparing these distributions with the output Catalogue. This requires clustering tools and robust implementations (WP 945); tools checking that all Catalogue fields have "reasonable" distributions;
- Retrieving from data already known specific structures;
- Developing tests for special objects.

Input:

Output:

Deliverables: • Definitions of tests • Software implementations • Results from database queries • Reports
Dependencies:
Interfaces:
Remarks: Validation will concern normal or special objects (such as multiple stars or extrasolar planets, extended objects or clusters, and solar system objects), complementing what is already implemented for internal CUs verifications

Gaia DPAC WP:		GWP-T-944-0000
Title: Validation: external archives		
Provider(s): Observatoire de Paris, ASDC-INAF, Observatoire de la Côte d'Azur, INTA-Istituto National de Tecnica Aeroespacial, CDS-Centre de données astronomiques de Strasbourg		
Manager(s): C. Babusiaux		
Start: 2013	**End**: 2022	**Total Effort**: 196 SM
Objective:		

Tasks:
The data validation WP940 aims at ensuring that Gaia data are usable and the errors are well known and within the expected limits. Gaia data validation will thus require the comparison with previously known, ”controlled” data sets. In this context, VO tools can be of great help to discover the catalogues of interest and to manage/compare the information provided by these catalogues with that provided by Gaia. The cross matching will be done with the VO tools elaborated in WP974 (e.g. using CDS tools such as X-match or Aladin, TOPCAT, etc.). Either big catalogs will be built as the result of the cross-match with the most popular catalogues or on-the-fly cross-matches will be done. However this cross-matching will be done knowing that Gaia has no other comparable all-sky survey with the same angular resolution and multiple star discovering power. Therefore 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, in particular the photometric transformations. In the comparison of the Gaia catalogue with external ones, we will face two different cases: • No other large scale catalogue will match the Gaia quality for individual sources (e.g. astrometric measurements). • Large scale catalogues exist for which individual parameters are better characterised (e.g. stellar atmosphere parameters). Both cases will be handled using probability distributions, looking for systematic effects and assessing the reliability of the estimated uncertainty. For this a detailed understanding of both the Gaia data and the external catalogues will be needed to properly account for the uncertainty, selection functions and possible biases on both sides.
Input:
Output:

Deliverables: • Definitions of tests • Software implementations • Results from database queries • Reports
Dependencies:
Interfaces:
Remarks: Validation will concern normal or special objects (such as multiple stars or extrasolar planets, extended objects or clusters, and solar system objects), complementing what is already implemented for internal CUs verifications

Gaia DPAC WP:		GWP-T-945-0000
Title: Validation: statistical tools		
Provider(s): Observatoire de Paris, University of Barcelona, ASDC-INAF, University of Groningen, University of Coruña		
Manager(s): M. Manteiga, A. Helmi		
Start: 2013	**End**: 2022	**Total Effort**: 196 SM
Objective: Statistical tools for the Catalogue validation		
Tasks: This work package will use or adapt tools developed in WP970 and WP980 to enable a statistical analysis of the Gaia data for validation purposes. It will assist in this respect the other Work Packages devoted to validation. The data analysis, whose goal is to detect artifacts, should allow tolerances in order not to select noise realisations instead of actual substructures, so the tools should be adapted to cope with random errors and data with limited precision. At some point special (outlying) sub samples will be detected, thanks to clustering tools and Artificial Intelligence techniques, data mining tools developed in WP973. What is then needed is a characterisation (i.e. statistical analysis and classification) of the properties of these sub samples with a subsequent visualisation. This will claim for adaptation of tools applying dimensionality reduction using statistical or neural network, cluster definition, for subsequent visualisation, such as in WP985. The following tasks are envisioned: • Clustering and sub-population statistical characterisation tools • Derivation of diagnostics from graphics • Study of correlations between observables • Tools using truncated, censored or correlated data, and robust to extreme values.		
Input:		

Output:
Deliverables: • Definitions of tests • Software implementations • Results from database queries • Reports
Dependencies:
Interfaces:
Remarks: Validation will concern normal or special objects (such as multiple stars or extrasolar planets, extended objects or clusters, and solar system objects), complementing what is already implemented for internal CUs verifications

Gaia DPAC WP:		GWP-T-946-0000
Title: Validation: time series and variability		
Provider(s): Geneva Observatory, Royal Observatory Belgium		
Manager(s): L. Eyer		
Start: 2013	**End**: 2022	**Total Effort**: 60 SM
Objective: Validation of time series		

Tasks:
This package will validate special classes of objects with the following purpose:

- Study variable sources to determine if some of their variability behavior is due to the instrument or the reduction,
- 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.

The following list of effect should be studied to see if there are some residual effects in the (spectro)-photometric and spectroscopic time series:

- effect from duplicity of stars (as function of the separation) or from bright contamination sources (possibly variable) coming from the same field of view or from the field other field of view
- search for periodicities linked to the sampling law properties (e.g. period of 6 hours or 63 days), or stars having identical periods.
- detection of variability trends and correlations with astrophysical parameters, color, magnitudes.
- intercorrelation (correlation or anti-correlation) between stars.
- correlation between variability characteristics and duplicity (such as separation).

This will be done in coordination with, and providing expertise to, the other validation work packages:

- scenarios: fraction of variables should be derived as function of position of sky.
- models: comparison with CU2 simulation.
- external catalogues: comparison with CoRoT/Kepler data for some fields.

Input:

Output:

Deliverables: • Definitions of tests • Software implementations • Results from database queries • Reports
Dependencies:
Interfaces:
Remarks:

C.5 WP950

Gaia DPAC WP:		GWP-T-951-0000
Title: Operations: management		
Provider(s): ESAC		
Manager(s): E.Mercier		
Start: 2013	**End**: 2022	**Total Effort**: 10.0 SM
Objective: Coordination of the Operations Systems Support of CU9		
Tasks: 1. in charge of monitoring the activities related to the Operations of CU9 2. update of the management information (SDP, milestones, resources,...) 3. interaction with the ESAC SAT team and the external services providers through monthly teleconferences using webex 4. description and minutes using Gaia DPAC wiki 5. reporting to the CU9 management 6. in charge of the issue of the user feedback reports to the CU9 management		
Input:		

Output: - ECSS documentation   - Progress reports   - Organisation of meetings and telecons   - Documentation production schedule   - Formal delivery of documentation at each Gaia data release
Deliverables: Outputs are generated on a regular basis - for the management information and the user feedback reports, they are delivered at the latest at the same time as the CU1 MDB extract for WP940.
Dependencies: The main dependencies of WP950 come from WP930, WP960, WP970, WP980.
Interfaces:
Remarks:

Gaia DPAC WP:		GWP-T-952-1000
Title: Operations: archive Maintenance		
Provider(s): ESAC		
Manager(s): P. Osuna		
Start: 2013	**End:** 2022	**Total Effort:** 46.0 SM
Objective: Archive Maintenance		

Tasks:

During operations, the maintenance of the Gaia archive will be of crucial importance in order to ensure a proper scientific use of the data archive and its appreciation by the community.
In the case of the Gaia archive, a very complex system is expected with not only source code developed by the ESA science archives team but also including the integration with external applications and services that could require not only deployment time but also extra checks on the system still to be defined.
The archive should be maintained up and running for a high enough percentage of the time and to reach this objective the following tasks need to be performed:

1. Ensure that the routine ingestion of data and the exceptional ingestion of catalogues is done on a daily basis. This includes the proper monitoring of the ingestion process and taking corrective actions in case problems occur which could be related with repository size, formats, hardware, etc.

2. Writing of documentation and operational procedures and keeping the latter updates in order to document corrective actions in case of problems. These procedures are usually distributed to the computer operators team to execute them during non-office hours. Even though proper procedures are put in placed and handed over to computer operators, sometimes the developer team needs to look into the problems themselves, especially if those too complex to be dealt with through a generic procedure.

3. Regular monitoring of system alarms, raised by automatic monitoring checks, and execution of procedures and analysis of problems during office hours.

4. Creation, maintenance and updates of monitoring procedures to ensure that possible problems of the system are identified and corrected.

5. Creation of regular statistics on the use of the archive complemented by their analysis, in order to have a correct estimation of the archive use and its impact on the scientific community and Gaia derived science.

6. Integration with external applications as well as other subsystems developed or used within CU9.

Input:
Output:
Deliverables:
Dependencies:
Interfaces:
Remarks:

Gaia DPAC WP:		GWP-T-953-1000
Title: Operations: technical support		
Provider(s): ESAC		
Manager(s): P. Osuna		
Start: 2013	**End**: 2022	**Total Effort**: 30.0 SM
Objective: Technical Support		

Tasks:

Due to the complexity of the system and the high number of users expected for the Gaia Archive, it is clear that a helpdesk service will be needed by the user community.
Usually, many of the technical questions related to technical support will be forwarded to the group of experts and developers of the Gaia Science Archives team at ESAC.
This group will need to analyse the questions to do a proper investigation of possible problems and create proper responses. It will also be involved in the definition of 'How To' and 'Frequently Asked Questions' pages to provide standard replies to scientists who are interested in addressing special science cases.
It has to be noted that the Gaia science archives will have very complex systems, not only huge catalogues that should be queried efficiently but also analysis systems inside the archive proper following the paradigm of 'send code to the data' typical of Big Data systems.
This will imply a totally new way of working for many scientists and for the community in general which would require extra technical support and an important effort in documentation.
Below is a top-level summary of the tasks to be performed by the ESA Science Archives Team at ESAC to support the helpdesk system:

1. Analyse, reply and correct possible technical problems reported by the users on the Gaia archive interface and/or its integration with other subsystems within CU9.

2. Investigate how to implement possible science analyses requested by the users. This implies the analysis of possible ways to implement the reported use case within the Gaia science archive, identification of possible requirements in case the use case cannot be executed, or the documentation in Frequently Asked Questions or manuals of example use case implementations to support and help the community in the proper use of the archive.

3. As part of the activities related to the development of the Core Systems by SAT, the integration of software developed within the Consortium (e.g. WP970 Science Enabling Applications) will be non-negligible and demand a lot of effort from this team (not only typical Helpdesk questions but also meetings, workshops, investigation of issues affecting the whole system performance, close collaboration with the rest of the consortium, etc).

Input:
Output:
Deliverables:
Dependencies:
Interfaces:
Remarks:

Gaia DPAC WP:		GWP-T-953-2000
Title: Operations: scientific support		
Provider(s): IoA, Cambridge		
Manager(s): N. Walton		
Start: 2013	**End:** 2022	**Total Effort:** 10.0 SM
Objective: Scientific Support Systems		
Tasks: This WP will provide advice to the operational support activity at ESAC concerning queries raised by end users which have a 'science query' content. 1. Analyse, reply and correct possible scientific problems reported by the users on the Gaia archive interface and its integration with other subsystems within CU9. 2. Investigate on how to implement possible science cases reported by the users. 3. Provide input into operational help desk systems required to respond to user queries with strong scientific content.		
Input:		
Output:		
Deliverables:		
Dependencies:		
Interfaces:		
Remarks:		

Gaia DPAC WP:		GWP-T-954-0000
Title: Operations: service monitoring and feedback		
Provider(s): CDS, ASDC INAF Univ		
Manager(s): P. Fernique		
Start: 2013	**End**: 2022	**Total Effort**: 40.0 SM

Objective:
The monitoring of the services evolution. The evolutions can be related to the Services and the Archive.

Tasks:

This WP covers the monitoring of the improvements of the CU9 outcomes.

1. description of the approach to track the issues and improvements to be performed
2. description of a classification scheme for the issues
3. designation of a point of contact and of the issues collection process
4. organisation of specific events to capture the request of the users on a face-to-face basis

Input:
The inputs of this WP will be mainly provided by WP952 and WP954.

Output:
Technical report summarizing the technical and scientific issues reported at ESAC, at the institutes and by the user community. This report will present a list of improvements recommendations to be submitted to the CU9-M.

Deliverables:

Dependencies:

Interfaces:

Remarks:

Gaia DPAC WP:		GWP-T-955-0000
Title: Operations: preservation		
Provider(s): ESA		
Manager(s): P. Osuna		
Start: 2018	**End**: 2022	**Total Effort**: 16.0 SM
Objective: Preservation		
Tasks: WP955 takes care of the data preservation during the final period of time. This usually includes migration (transferring of data to newer systems environments), replication (creation of duplicates of copies and/or backups of the data and meta-data to prevent vulnerabilities), emulation (replication of functionality present in an obsolete system) and metadata attachment that includes information on creation, access rights, restrictions, etc. There are several methodologies that could be used to ensure data preservation and more are expected in a medium term as this area is in fast evolution.		
Input:		
Output:		
Deliverables:		
Dependencies:		
Interfaces:		
Remarks:		

Gaia DPAC WP:		GWP-T-956-0000
Title: Operations: Cloud / Workflow		
Provider(s): IoA, Cambridge		
Manager(s): G. Rixon		
Start: 2013	**End**: 2022	**Total Effort**: 10.0 SM
Objective: Scientific Support Systems		
Tasks: This WP will evaluate use of Cloud storage and compute services for operational deployment through CU9. Use of the cloud will potentially be of use in for instance distribution of advanced Gaia CU9 applications via cloud deployable Virtual Machines. 1. Analysis of cloud solutions for CU9 2. Evaluate standards available to implement the required cloud capabilities 3. Liaise with the EGI Federated Clouds Task Force 4. Evaluate use of EGI Federated Cloud Testbed (see `https://wiki.egi.eu/wiki/Fedcloud-tf:UserCommunities`)		
Input:		
Output:		
Deliverables:		
Dependencies:		
Interfaces:		

Remarks:

Gaia DPAC WP:		GWP-T-957-0000
Title: Operations: auxiliary data		
Provider(s): INAF, Bologna		
Manager(s): E. Pancino		
Start: 2013	**End**: 2022	**Total Effort**: 9.0 SM
Objective: Operations of the CU9 auxiliary data.		
Tasks: WP957 takes care of defining the requirements for the preparation of a web interface to access/browse all DPAC auxiliary data, and similarly defines the requirements on the underlying database when needed. The input of WP957 will come from WP973, where a census of the available data takes place. For all DPAC auxiliary data that are stored in the MDB, or in external databases and archives, WP957 provides the requirements for the integration into a single access-page and, if and when feasible, a single browsing/querying interface.		
Input:		
Output:		
Deliverables:		
Dependencies:		
Interfaces:		
Remarks:		

Gaia DPAC WP:		GWP-T-958-0000
Title: Operations: provision of simulations(GOG)		
Provider(s): UB		
Manager(s): X. Luri		
Start: 2013	**End**: 2022	**Total Effort**: 11.0 SM

Objective:
Provision of simulated data for system testing and validation

Tasks:

Provision of simulated archive data for development, testing and validation purposes. These data will be generated in coordination with DPAC CU2, using mainly the GOG data generator Isasi et al. (YI-003) developed by this unit.
This task includes both the generation of such simulations using the appropriate hardware resources (likely the MareNostrum supercomputer at the DPCB) and the specific developments of GOG (beyond the CU2 implementation) to adapt it to the CU9 needs.

Input:

- Specifications for the contents of the archive to be simulated, preferably based on the description of the MDB contents
- Specifications on the size of the requested data sets and their delivery dates to allow the planning of the GOG executions

Output: Simulated data sets: • Partial (small) data sets for specific testing and validation. Number, dates and sizes to be defined in the global CU9 planning. • Full sky, full density data sets for testing and validation of the complete archive systems. It is almost certain that the simulation of the full archive contents, from raw data to final data (1PB) will not be feasible, but the simulation of the combined and epoch data is feasible (already tested in CU2). One such a full sky, full density simulation is envisaged for each archive release.
Deliverables:
Dependencies:
Interfaces:
Remarks:

C.6 WP960

Gaia DPAC WP:		GWP-M-961-0000
Title: Education & Outreach: Management		
Provider(s): ARI/ZAH, Univ. Barcelona		
Manager(s): S. Jordan, E. Masana		
Start: 2013	**End**: 2022	**Total Effort**: 20 SM
Objective: Coordination of the Education & Outreach Tasks		
Tasks: 1. Assuring that the objectives and schedule of WP960 are met. 2. Definition of requirements. 3. Regular video conferences. 4. Coordination by Wiki, svn, and Email. 5. Coordination with other WPs (visualisation tools, documentations). 6. Coordination with ESA Gaia Portal and supporting web sites		
Input:		
Output: Coordination of WP960, see deliverables.		

Deliverables: • ECSS documentation • Progress reports • Organisation of meetings and telecons • Documentation production schedule • Formal delivery of documentation at each Gaia data release
Dependencies:
Interfaces: WP910, WP920, WP930, WP940, WP950, WP970, WP980
Remarks:

Gaia DPAC WP:		GWP-T-962-0000
Title: Outreach: general resources		
Provider(s): ESAC, University of Coruña		
Manager(s): S. Els, M. Manteiga		
Start: 2013	**End**: 2022	**Total Effort**: 104 SM

Objective:
General Support for Education & Outreach Task

Tasks:

1. Editorial support.
2. Support of early technical and science demonstration cases.
3. Collecting and monitoring of Gaia science and technology impact.
4. Collection and consolidation of contents for the Gaia web portal and supporting web sites.
5. Inventory of useful astronomical resources in Internet for use in Gaia outreach.
6. Support in translation to different European languages of information and materials.
7. Support in development of visualisation computer tools.

Input:

Output:
See deliverables.

Deliverables:
1. A list of astronomy outreach open resources available in Internet to serve as inspiration or for direct use in Gaia outreach program. 2. Documents with early technical and science demonstration cases. 3. Documents on Gaia science and technology impact. 4. Material for the Gaia web portal and supporting web sites. 5. Support in the development of resources in the framework of the most common social networks environments (Facebook, Twitter, blogs, forums, etc). 6. Support in general maintenance of contents and resources available on the internet Gaia outreach server 7. Support in the translation to different languages of Gaia resources and materials.
Dependencies: WP983, WP985, WP986, WP970
Interfaces: WP961, WP963, WP964, WP965
Remarks:

Gaia DPAC WP:		GWP-T-963-0000
Title: Outreach: academic outreach		
Provider(s): University of Coruña, ESAC		
Manager(s): A. Ulla, S. Els		
Start: 2013	**End**: 2022	**Total Effort**: 48 SM
Objective: Providing Information and Support for the Academic Community		

Tasks:

1. Providing Web pages with well structured information on how to obtain information about the Gaia science capabilities and how to access to the Gaia data itself.
2. Providing FAQs on Gaia (technique, usage of Gaia data, science applications).
3. Providing a help desk which would answer questions on the technique and science applications of Gaia.
4. Producing tutorials for the usage of the Gaia data covering the defined use cases for CU9.
5. The latter includes feedback from the users in order to update the tutorials.

Input:

Output:
See deliverables.

Deliverables: 1. A Help-Desk to assist scientist and academic users on technical questions of Gaia data, products and tools. 2. An efficient search tool trough CU9 resources. 3. Web pages contents with information about the access to Gaia data, reduction procedures and science exploitation. 4. FAQs covering the main academic applications on Gaia (techniques, usage of Gaia data, science applications). 5. Tutorials (including videos) containing clearly detailed step-by-step cases. 6. Distribution of news and alerts over the internet (mailing lists, social networks, etc.) to keeps users up to date on recent developments or improvements, related do Gaia data and science applications. 7. A feedback procedure of interaction with interested users, in order to keep up to date and improve all the above.
Dependencies:
Interfaces: WP961, WP962, WP964, WP965
Remarks:

Gaia DPAC WP:		GWP-T-964-0000
Title: Outreach: education and general public outreach		
Provider(s): ARI/ZAH, University of Barcelona		
Manager(s): S. Jordan, E. Masana		
Start: 2013	**End**: 2022	**Total Effort**: 77 SM

Objective:
Providing Support and Material for the General Public

Tasks:

1. Support in preparation of lectures and courses about Gaia technical and scientific aspects, for a variety of audiences (school level, elder people, general public) and in different formats (written, on-line, in presence, etc).
2. Support in preparation of exhibition material
3. Support in the development of computer games/tools for android operative system
4. Preparation of booklets, comics, etc. about Gaia and Gaia science.
5. Providing material for school children and teachers, including practical exercises with Gaia data (e.g. with Celestia).
6. Extensive testing of such material and including the feedback into the production of newer versions.
7. Providing a help desk for general questions on Gaia.

Input:

Output:
See deliverables.

Deliverables:

1. Elaborate educational material useful for different groups (children, amateur astronomers, elder people, etc.):
   - Booklets,
   - Comics,
   - Movies.
2. Development of scientific data-mining projects Gaia@home, in the manner of SDSS galaxy zoo classification project attractive for the public and useful for the archive exploitation
3. Support for a discussion forum.
4. Support in the preparation of presentation material in the form of exhibition posters, a traveling show, scale models, lectures, demos, etc

Dependencies:

Interfaces:
WP961, WP962, WP963, WP965

Remarks:

Gaia DPAC WP:		GWP-T-965-0000
Title: Outreach: media		
Provider(s): University of Barcelona, ARI/ZAH		
Manager(s): X. Luri, S. Jordan		
Start: 2013	**End**: 2022	**Total Effort**: 44 SM
Objective: Providing Information and Tools for the media		
Tasks: 1. Preparation of diverse material for the press and media: short articles, review articles, graphical material, audiovisual material 2. Pay attention to press request of information, attend media interviews 3. Serve as contact between the media and Gaia team specialists in technical/scientific specific subjects.		
Input:		
Output: See deliverables.		
Deliverables: 1. Material for the press and media: • short articles, • review articles, • graphical material, • audiovisual material.		
Dependencies:		

Interfaces: WP961, WP962, WP963, WP964
Remarks:

C.7 WP970

Gaia DPAC WP:		GWP-M-971-0000
Title: Science enabling applications: management		
Provider(s): Universitat de Barcelona		
Manager(s): X. Luri & S. Jordan		
Start: 2013	**End**: 2022	**Total Effort**: 20 SM
Objective: Management of Science Enabling Applications WP		
Tasks: The objective of this work package is to ensure that WP970 meets its objectives on schedule. Tasks will include: • coordinating and supervising activities to be carried out • monitoring project progress, quality and timing of deliverables • coordination with the rest of packages to ensure proper interfacing with the archive (WP930) and suitability of the tools for the CU9 tasks (WP940 & WP980) • Implementing the data right policy • Reporting back to the CU9 executive board		
Input:		
Output:		

Deliverables: • ECSS documentation • Progress reports • Organisation of meetings and telecons • Documentation production schedule • Formal delivery of documentation at each Gaia data release	
Dependencies:	
Interfaces: WP910, WP920, WP930, WP940, WP970, WP980	
Remarks:	

Gaia DPAC WP:		GWP-T-972-0000
Title: Science enabling applications: advanced data access tools		
Provider(s): ARI/ZAH, ESAC, Univ. Coruna, Royal Obs. Belgium, CDS, Obs. Paris-Meudon, IFA Edinburgh, INTA, Univ. Bristol, AIP		
Manager(s): S. Jordan & H. Enke		
Start: 2013	**End**: 2022	**Total Effort**: 263 SM
Objective: Delivery of client–side applications to facilitate Gaia exploitation		

Tasks:

This category of applications addresses need for advanced tools to access the massive amount of data that will be contained in the Gaia archive.
These applications fall into three categories:

A Interactive fitting of spectra:

Development of tools for interactively fitting spectra using model spectra or templates, building on VO tools such as SPLAT. In particular, the VO SPLAT (Spectral Analysis Tool) needs extensions to use not only SSAP (Simpole Spectral Access Protocol), but also the TAP.

B Integration of Gaia data cataloges into services like Aladin + VizieR

VizieR catalogue service - allowing to select and download excerpts of Gaia catalogue

Aladin sky atlas - allowing to display the Gaia catalogue as individual sources over a large palette of reference surveys (DSS, SDSS, 2MASS, GALEX...) and catalogues

C Adaptation of VO-Tools to service Gaia Data

Adaptation of VO-tools already available (VOSED, VOSA, Filter Profile Service,...) For TOPCAT, a GBIN input handler for STILTS is required to deliver direct MDB access.

D Extending query functionalities in RDBMS for astrometric and astronomical purposes Fast access to the 6-dimensional phase space of Gaia data is currently not properly supported, as well as HEALPix or other indexing of data. This also enables better support for Visualisation of large data volumes.

E Non-interactive tools for accessing and using Gaia data

Support for non-interactive access to the data can take various forms. Aside from plain non-interactive querying, a facility e.g. to call plotting services with suitable parameters, once the interactive usage of a service has provided the desired data range.

Input:

Output:

Deliverables:
1. Tools for interactive fitting of spectra. 2. Integration of Gaia data cataloges into services like Aladin + VizieR. 3. Adaptation of VO-Tools to service Gaia Data. 4. Extending query functionalities in RDBMS for astrometric and astronomical purposes. 5. Non-interactive tools for accessing and using Gaia data
Dependencies: WP932, WP933
Interfaces: WP963, WP971, WP984
Remarks:

Gaia DPAC WP:		GWP-T-973-0000
Title: Science enabling applications: data mining		
Provider(s): Univ. Barcelona, ASDC-INAF, AIP, Univ. Coruna, KU Leuven, INTA		
Manager(s): X. Luri & L. Sarro		
Start: 2013	**End**: 2022	**Total Effort**: 216 SM
Objective: Delivery of data mining tools for the analysis of the Gaia catalogue		

Tasks:

This sub-work package covers the 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.

A: analysis of the infrastructure for data mining

Applying data mining algorithms to the Gaia data poses a challenge. Data mining algorithms require the whole data set to be scanned multiple times, depending on the computational order of the employed algorithm. In order to allow fast full table scans, different data access strategies need to be assessed to find the optimal solution for each task. Some algorithms might work well employing a brute force strategy using technologies such as Hadoop while other do better with optimised tree structures and specialised indexes. The first step in the implementation of the tasks in this work package will be to **define the infrastructure required for the data mining tasks and its implementation into the Gaia archive, in coordination with WP930.**

This implies an infrastructural challenge, since we have to ensure that the Gaia archive will be able to handle all computationally demanding jobs. From the architecture point of view, the data mining 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.

Tasks: **B: development of a standardised data access interface for data mining** Based on the analysis carried out in Task A we will work on the development of a standardised data access service interface for the various data mining algorithms. This interface should allow an easy implementation of data mining tasks running on the Gaia archive, hiding as much as possible the particularities of its specific implementation. In what form such an access has to be organised needs to be closely coordinated with the archive development itself as defined in WP930.
Tasks: **C : provision of general data mining tools** Using the interface defined in B we will implement a set of general-purpose data mining tools. The aim of this tool set will be to cover the most common data mining needs in order to facilitate the exploitation of the Gaia archive for the typical user. The definition of what these general purpose tools should be will be based on the compilation of requirements and use cases from the actual Gaia archive users, as it has been done through GREAT (http://great.ast.cam.ac.uk/Greatwiki/GaiaDataAccess). Examples of typical use cases are to reveal patterns and relationships within the astronomical data that can lead to the detection of new types of objects or isolated, exotic objects, the capability of automated dimensionality reduction (feature extraction, feature selection) or the development of key learning algorithms (clustering, outlier analysis, swarm intelligence,...).

Tasks: **D: provision of specialized data mining tools for key Gaia topics** Aside from the general purpose tools described in Task C we also plan to implement some specialized data mining tools that will be needed for key CU9 tasks or, more generally, needed to fulfill some of the key issues of the Gaia scientific case. The selection of these specialized tools will be performed after a careful analysis and evaluation and priorisation and could include • A tool allowing a global comparison of a galaxy model (e.g. the one used for the catalogue validation tasks described in WP940) with the actual Gaia data. • Data mining tools enabling selection and exploration in the Fourier domain for variability analysis. • Tools to cross-correlate spatial position and variability information, to easily explore the structure of our galaxy using variable stars. • Tools to query per-CCD photometry to mine and explore short-term variability.
Input:
Output:
Deliverables:
Dependencies:
Interfaces:
Remarks:

Gaia DPAC WP:		GWP-T-974-0000
Title: Science enabling applications: auxiliary and external data		
Provider(s): ASCD-INAF, INAF-OABo, Laboratoire d'Astrophysique de Bordeaux, MPIA, INTA, Univ. Coruna, CDS		
Manager(s): P.M. Marrese		
Start: 2013	**End**: 2022	**Total Effort**: 326 SM
Objective: Integration and matching of external and auxiliary data		
Tasks: **A: Integration of auxiliary data** We will first make a census of all the auxiliary data used in DPAC to calibrate Gaia data or to train Gaia algorithms, and make provisions to maintain the census updated if and when new data come in. We will consequently evaluate whether these data are adequately published and available elsewhere for the needs of Gaia data quality, availability, and repeatability. In case some data are not available (they are not in public archives or databases, are not in standard formats, or the teams providing them request help to preserve the data), we will help the teams preparing their data for publication. Strong interaction with WP 950, where the actual requirements for preparing web interfaces for auxiliary data are prepared, is necessary. Also, cross-match of those auxiliary data which are not already in the MDB need to take place in Task B. Ultimately, every Gaia data release is accompanied, when appropriate, by the relevant auxiliary data.		

Tasks:

Task B: Cross-match with external catalogues

The Gaia astrometric, spectrophotometric and radial velocity data promise to enable excellent science in a wide range of fields. Yet these data will not exist in isolation: Infrared data and higher resolution spectroscopy already exists or will exist for many Gaia targets from other surveys. Many scientific questions will be best addressed through the combination of data sets. The DPAC processing does not take into account such external data: calibration aside, the Gaia catalogue is based on Gaia-only data.

We think that data retrieval across archives should be made available to scientists and we thus propose that the pre-computed cross-match should be integrated in the Gaia query system (see for example GDAS-GA, GDAS-ST-15, GDAS-GA-22, GDAS-BR-03, GDAS-ST-23, from Brown et al. (AB-026)).

In addition, the validation (WP 940, in particular WP 943) will require an intensive cross-matching and inter-operation of Gaia data with existing and future astronomical archives.

B1 Multi-wavelength cross-match

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 possibilities for astronomical research. While considering a cross-match of the Gaia Archive with external catalogues, one faces two very different problems. On one side, the matching of large data sets (usually in the optical range) where problems are mainly related to the amount of data. On the other side, the inter-matching of data sets from largely different wavelength ranges, where the problems arise because of the heterogeneity of the data.

These two problems need quite different techniques, that is a cross-match tool that is flexible enough or more probably a set of different tools or even different HW (for example GPU in the case of large data sets).

Tasks:

B1 continues

Our proposal would be to:

1. add to the Gaia archive pre-computed cross-match (list of neighbours, nearest neighbour) for the largest (number of objects > 100 million) and more interesting public available catalogues;

2. supply the developed cross-match tools to the final user, to allow them to compute cross-match on the fly for smaller and/or not public data sets;

3. pre-compute the cross-match (list of neighbours, nearest neighbour) between Gaia data and the auxiliary data when needed (see Task A).

The algorithms made available to the final user should provide not only an association between sources, but also a figure of merit to describe the probability (or goodness) of the association.

B2 Classification and astrophysical parameter estimates

We plan to produce extended catalogues, and to apply CU8 work to give improved classification and astrophysical parameter estimates on the corresponding subset of Gaia sources. The first step is to cross match the Gaia sources with other major catalogues, such as 2MASS, UKIDSS, VISTA, SDSS, Pan-STARRS, LSST, LAMOST (depending on what becomes publicly available when), thus giving extended information on many Gaia sources. The DPAC classification models are retrained to take into account the expanded data sets. This is valuable for a number of investigations, such as better estimations of the three-dimensional variation of interstellar extinction. For example, in estimating the metallicity, extinction and effective temperature of a set of stars, we should, where possible, use not only the Gaia BP/RP spectrum and the parallax, but also broad band infrared data, perhaps also UV photometry. Alternatively, the user may want to find all low metallicity stars according to some set of stellar models, using not just the Gaia data but also, say, LAMOST spectra.

Tasks:

B3 VO-tools to cross-match

We plan to participate in the design and development of workflows as well as use of VO-tools to cross-match Gaia data with external catalogues.

B4 CDS integration of Gaia data

The CDS contribution would be the integration of GAIA catalogs (various releases) into the well known CDS tools in order to make them available to the astronomical community, notably in:

Xmatch service - allowing to cross-match GAIA catalogue with any other catalogue (10.000 already available at CDS), including very large ones, $> 10^9$ objects

Input:

Task A Through the GBOG working group representatives, we will gather the auxiliary data and related information which is needed for the census and the evaluation of the auxiliary data publication needs.

Task B We will gather the relevant and publicly available external catalogues, together with their documentation.

Input:

Task A Through the GBOG working group representatives, we will gather the auxiliary data and related information which is needed for the census and the evaluation of the auxiliary data publication needs.

Task B We will gather the relevant and publicly available external catalogues, together with their documentation.

Output:

Task A

- Evaluation, done in conjunction with the teams who own the data and the DPACE, of what data are adequately published elsewhere (public databases and archives), or in the other CUs by means of the MDB, and what data need to be preserved in some better way and published through CU9;
- Evaluation, done in conjunction with the teams who own the data, about the best way to make the data VO compliant and to store the data in the GBOG ESAC disk backup space in Madrid;
- Recommendations and description to WP 957 about the data that needs to be published, and interaction with that group to help them define the best requirements for the database and web interface.

Task B

- Integration (cross-match and classification) of auxiliary and external catalogues into the Gaia archive through some different methods (precomputed cross-match, cross-match tools available to final user, use of VO cross-match tools, integration of Gaia data into CDS services, use of CU8 classification tools).

Deliverables:
1. **Task A** Census of all the auxiliary data programs in DPAC including both the observational data programs and the theoretical libraries, both those used to calibrate Gaia data, and those used to train Gaia algorithms. 2. **Task A** List of auxiliary data objects that need to be cross-matched, to be given to Task B; which is expected to be at most a few thousand objects (conservative estimate). 3. **Task A** Data delivery of auxiliary data associated with each Gaia data delivery. 4. **Task B1** A pre-computed cross-match (list of neighbours, nearest neighbour) added to the Gaia archive for the largest (number of objects $>$ 100 million) and more interesting public available catalogues (e.g. 2MASS, UKIDSS, VISTA, SDSS, Pan-STARRS, LSST, LAMOST). 5. **Task B1** Cross-match tools for the final user, to allow them to compute cross-match on the fly for smaller and/or not public data sets. 6. **Task B1** A distributed list with pre-compute the cross-match (list of neighbours, nearest neighbour) between Gaia data and the auxiliary data. 7. **Task B2** Resulting custom catalogue(s) including improved classification and astrophysical parameter estimates. 8. **Task B4** Integration of Gaia data into CDS services, in particular into the Xmatch service.
Dependencies:

Interfaces: **Task A** We will interface with GBOG members which are the main contact point with the CUs and the auxiliary data teams; with WP 957 members which are the main contact-point with the team which will design the actual auxiliary database (if needed) and web pages; and with the scientific teams who own the data, who at the time of writing are part of DPAC, but when some of the CUs task will be completed after launch might not be part of DPAC anymore. **Task B** We will need interfaces with WP 930 and WP 950 teams, to make sure that the combined catalogues delivered by Task B team will be well integrated into the Gaia archive and made available by the Operation & Services team.
Remarks:

Gaia DPAC WP:		GWP-T-975-0000
Title: Science enabling applications: science alerts		
Provider(s): IoA Cambridge		
Manager(s): N. Walton		
Start: 2013	**End**: 2022	**Total Effort**: 27 SM
Objective: This WP will ensure the integration of the high throughput daily Gaia science alerts stream into the archive, including suitable end user access and visualisation tools.		

Tasks:

A: The Gaia flux-based science alert stream will be issued to the community through the science alert processing carried out at the Cam-bridge Photometric Data Processing Centre (DPCI).

B: The science alerts processing will issue basic information for each flux alert via the VOEvent system to the community in a timely fashion (with alerts being produced 12 days after observation by Gaia).

C: The alert packet will contain basic characterisation information for each event, including parameters such as estimated alert object type, and more advanced classification for certain objects such as supernovae (SNe). For these, inherent Gaia photometric data will be used to provide additional information concerning SNe alerts including class, epoch, redshift, reddening.

D: The work to be carried out in this WP will develop the interfaces required to connect the real time science alerts classification processing to the main Gaia data products in the archive. Thus, as the mission evolves, and more knowledge is accumulated about objects measured by Gaia as it successively scans the sky, there will be opportunity to cross reference new alerts against previous knowledge of that sky point as well as previous alerts against new information. Thus for instance, irregular outburst events may show up multiple times during the Gaia mission. Identification will be improved through correlation with earlier Gaia knowledge.

E: The testbed will in addition provide linkages to external data resources provided through CU9, in particular via interfaces to the archive development in WP930. In particular some staging of accumulated alerts may be provided via cloud storage WP931-4 - to enable higher throughput access to client visualisation tools.

F: Finally the alerts visualisation system will plug into the data portal - allowing simple visualisation of the daily and accumulated alerts stream with end user tools such as Worldwide Telescope, Aladin, DS9.

Input:

Output:
Deliverables: 1. Technical requirements document input specific to Alerts stream. 2. Alert packet stream integration into compliant existing astronomical all sky data visualisers (e.g. WWT, Aladin, DS9). 3. Deployment of interactive collaborative visualisation - to facilitate community annotation of newly released alerts. 4. Requirements for mobile device based alert visualisation. 5. Workflow linking science alert stream to visualisation tools (e.g. deployment of WWT alert visualiser currently under development Cambridge/ MSR). 6. WP975 software test reports. 7. Software integration.
Dependencies: Requirements will be provided via WP913
Interfaces: This WP will provide status reports to WP971. Documentation will be provided to WP920
Remarks:

C.8 WP980

Gaia DPAC WP:		GWP-M-981-0000
Title: Visualisation: Management		
Provider(s): Univ. Lisbon, Univ. Vienna, ARI/ZAH		
Manager(s): A. Moitinho, J. Alves		
Start: 2013	**End**: 2022	**Total Effort**: 31 SM
Objective: Coordination of the Visualisation Work Package		
Tasks: The main role of WP981 is to assure that the objectives and schedule of WP980 are met. It will also be in charge of setting requirements transversal to the other WP980 sub-WPs. The specific tasks are: 1. Overall coordination of WP980 and sub-WPs via meetings, video conferences, wiki, svn, and email. 2. Coordination of interface with the other CU9 WPs. 3. Gather the requirements for WP980. 4. Definition of visualisation templates (pre-built visualisation modes). The templates themselves will be developed within the pertinent sub-WPs. 5. Coordination of documentation and tutorials.		
Input:		
Output: Coordination of WP980		

Deliverables: 1. Minutes of meetings. 2. Document with visualisation, functional, interface, architecture, software engineering, operational, performance and security requirements. 3. List and description of pre-defined visualisation modes.
Dependencies:
Interfaces: WP910, WP930, WP940, WP960, WP970
Remarks:

Gaia DPAC WP:		GWP-T-982-0000
Title: Visualisation: data and services interface		
Provider(s): ARI/ZAH		
Manager(s): S. Jordan		
Start: 2013	**End**: 2022	**Total Effort**: 30 SM
Objective: Provision of an efficient interface to Gaia data bases.		
Tasks: This sub-workpackage will provide the necessary tools and infrastructure to efficiently connect the visualisation software to the Gaia data bases. While the VO developments in WP970 are more related to publishing the archive to the VO and providing VO services the VO development in WP982 is aimed at providing a VO layer that feeds data and services into the visualisation infrastructure.		
Input:		
Output:		

Deliverables: 1. Contribute modules providing interoperability with other VO applications via SAMP. 2. Contribute modules providing interoperability with VO archives (e.g. TAP/Obscore and SIAP). 3. Provide a Command Language interface (possibly Python or R) for interacting with the Visualisation Framework. 4. Transform data selection through interaction with the Visualisation Framework into queries to the Gaia data base. 5. Documentation (including user manuals and tutorials) of the software and the interfaces to the Visualisation framework.
Dependencies: WP935, WP972, WP973
Interfaces: WP981, WP983, WP984, WP985, WP986
Remarks:

Gaia DPAC WP:		GWP-T-983-0000
Title: Visualisation: visualisation infrastructure		
Provider(s): Univ. Lisbon, UNINOVA, Vienna, ARI/ZAH, Univ. Bristol, Geneve		
Manager(s): A. Krone-Martins, A. Moitinho		
Start: 2013	**End**: 2022	**Total Effort**: 200 SM
Objective: Delivery of a visualisation infrastructure for the Gaia archive		

Tasks:

Development of the visualisation infrastructure

This sub-workpackage will provide an interactive visualisation infrastructure, integrated with the Gaia archive, and capable of coping with the volume and nature of the data in the archive. Interactivity will be a central part of its design and implementation. A compilation, evaluation and adoption studies of technologies adequate for the interactive 3D visualisation of the archive will be done. Whenever possible, off-the-shelf solutions will be adopted. Collaborative visualisation functionalities will also be assessed during the design of the infrastructure. This functionality is foreseen for the final release.

Definition and implementation of the requirements and integration

- Supported facilities: The infrastructure will be mainly aimed at particle visualisation, with support for extended structures through WP984 products. It will also provide adequate text and glyph annotation, as well as data overlaying capabilities.
- Interactivity: Besides internally supported particle and volume data-selection strategies, the infrastructure will support algorithmically assisted data-selection – this will be provided by WP985.
- Interfaces: Depending on the availability, adequacy and adoption by the community of existing specialised visualisation solutions, the infrastructure may interface with them. This integration with external tools will be assured through virtual-observatory protocols layers defined by WP982 and WP970.
- Scalability: The infrastructure must be able to scale up to the entire data set depending on the underlying hardware it is deployed, and this should be performed while keeping interactive capabilities – this will be performed via off-loading strategies integrated within the infrastructure.
- Assisting visualisation: Pre-defined visualisation modes will be defined depending on usage of the data release scenarios – two special such cases have dedicated WPs: time-domain (WP986) and transients (WP987).
- Cosmetics and outreach: Visualisation on domes and outreach-aimed cosmetic filters will be assessed in cooperation with WP964.

Input: Gaia Archive or Gaia-like simulated data
Output: Integrated visualisation framework
Deliverables: 1. Technical Reports on definitions, visualisation scenarios and visualisation technology adoption adequacy. 2. Software documentation (SRS, SDD, STS, STR). 3. Software implementation: visualisation infrastructure adapted for each archive release. 4. Software integration.
Dependencies: WP982
Interfaces: WP931, WP960, WP972, WP981, WP984, WP985, WP986, WP987
Remarks:

Gaia DPAC WP:		GWP-T-984-0000
Title: Visualisation: volume/isosurface rendering for extended structures		
Provider(s): U. of Vienna, Univ. Innsbruck, Univ. Milan		
Manager(s): J. Alves		
Start: 2013	**End**: 2022	**Total Effort**: 135 SM

Objective:
Visualisation of extended structures

Tasks:

The goal of this sub-workpackage is the development of robust algorithms to provide astrophysically meaningful surfaces inside the Gaia database. These surfaces will be implemented in the visualisation infrastructure.

1. Compilation and evaluation of available, of the shelf, solutions for Volume/isosurface rendering in the context of the Gaia archive.
2. Development of algorithms for the construction of statistically robust volumes/isosurfaces for the most relevant astrophysical quantities, based on Gaia data only (e.g., Av, metallicity, age, etc.).
3. Development and construction of software tools for the interactive visualisation of the volumes/isosurfaces, together with other Gaia observables.
4. Coordination of documentation and tutorials.

Input:
Gaia Archive or Gaia-like simulated data

Output:

Deliverables:
1. Report on of the shelf, solutions for volume/isosurface rendering in the context of the Gaia archive. 2. Algorithms for the construction of robust volumes/isosurfaces for the most relevant astrophysical quantities, based on Gaia data only (e.g., Av, metallicity, age, etc.). 3. Volumes/isosurfaces of relevant astrophysical quantities. 4. Software tools, based on existing solutions, for the interactive visualisation of the Volumes/isosurfaces. 5. Documentation and tutorials.
Dependencies: WP983
Interfaces: WP981, WP985
Remarks:

Gaia DPAC WP:		GWP-T-985-0000
Title: Visualisation: clustering and advanced data selection for multi-D visualisation		
Provider(s): U. of Groningen, U. Coruna, U. Vigo		
Manager(s): A. Helmi		
Start: 2013	**End**: 2022	**Total Effort**: 74 SM
Objective: Tools for multi-D visualisation		

Tasks:

The focus of WP985 is to enable the exploitation of the Gaia data set through the exploration of its multi-dimensional character. Gaia is unique in the sense that it combines full phase-space information with astrophysical information of each of the $\sim 10^9$ stars, and this multi-dimensional nature is what guarantees that we will be able to fulfil the primary goal of the mission, namely to unravel the formation and evolution of the Galaxy.
However, the visualisation of a multi-D data set when the dimensionality is $D > 3$, and for a large data set like Gaia's, may be very cumbersome. Furthermore, often we do no want to visualise all dimensions, but only those that are "rich in information". For example, in the case of searches of moving groups, the velocities and spatial location of the stars are useful, but their temperature or log g less so (or might only at a later stage). In the most general case, the interesting dimensions will not be known a priori.
This is why we propose here to take a pragmatic approach and implement dimensionality reduction methods. We will do this in two ways:

- Subspace clustering and ranking of subspaces The idea is to explore all possible subspaces (1D, 2D pair-wise, 3D combinations, etc of the different dimensions), apply a clustering algorithm to quantify the degree of "information" (e.g. clusters or correlations between dimensions/variables), and rank these subspaces on this basis (e.g. quality/significance of the clusters).

- Self-Organising Maps SOMs are suitable for big data sets because firstly, they cluster the data, hereby producing a good approximation of the data set with a reduced amount of "prototypes", and secondly, they project the data in a 2D or 3D map which preserves the input topology.

Once this steps have been implemented, we proceed to the *visualisation stage*. Here we propose to use standard visualisation toolkits linked within a graphical user interface, such as those discussed in WP983 and 984 that can deal with large data sets (scatter plots, density fields, parallel coordinates plots, etc). Since interactivity is an important aspect of visualisation, it is critical to integrate the dimensionality reduction method with the visualisation in a flexible manner to allow refinement/selection of the clustering steps described above.

Input:
Gaia Archive or Gaia-like simulated data

Output:
Deliverables: 1. Data and dimensionality reduction (compression) by means of clustering algorithms, both statistical (kmeans, fuzzy k-means, PCA, etc) and neural network based (Self Organising Maps, SOM). 2. Projection of multidimensional data sets to 2D or 3D manifolds, using prior clustering analysis, SOM, and multidimensional Scaling 3. Advanced cluster selection coupled to visualisation. 4. Preprocessing steps previous to clustering like signal processing and noise filtering. 5. Optimised methods for fast computation and supercomputing enabled software.
Dependencies: WP981
Interfaces: WP981, WP983, WP984
Remarks:

Gaia DPAC WP:		GWP-T-986-0000
Title: Visualisation: time-domain visualisation		
Provider(s): U. Geneva		
Manager(s): L. Eyer		
Start: 2013	**End**: 2022	**Total Effort**: 27 SM
Objective: Provision of tools to visualise the variability related data of the Gaia archive.		

Tasks:

Historically, the variability related interaction of scientists with astronomical archives was focused on a single or a few observational targets. Information access on the basis of an individual star or position in the sky was the only path required. For many astronomers this access path will remain a vital element. It is an invaluable probe to assess the quality of the information derived for the whole ensemble in a similar way that samples play in statistical process control. In addition to accessing the data specific to one observed variable star, astronomers today and in the future require visualising data aggregated over the large number of members of one or several classes of variables stars.

Visualisation of the variability data for individual stars.

For individual stars a detailed description of the variability related attributes needs to be provided. Besides a textual display of attributes like star type, variability type, period etc. several graphical displays are needed:

1. Visualisation of the star's light curve in all available bands G, BP and RP.
2. Visualisation of the star's radial velocity time series.
3. Visualisation of time series folded with the recovered period of the signal.
4. Visualisation of the harmonic model fit to the observational data.

The selection of the star for which the data is displayed is made in several ways:

- By entering the Gaia sourceId, name or position of a star
- By selecting a point in an aggregated plot, e.g. the HR-diagram of an ensemble of stars.

Visualisation of data aggregated over large ensembles of variable stars for a pre-defined set of objects and attributes.

1. Visualisation of light curves for a set of stars.
2. Visualisation of 1 and 2-dimensional histograms.
3. Visualisation of 2-dimensional and 3-dimensional point clouds.

Tasks: **Visualisation of data aggregated over large ensembles of variable stars for a set of objects and attributes selected in an ad-hoc manner.** In addition to the pre-defined diagrams described above, the same diagrams have to be available for data selected by the user in an interactive way. This data selection can be for example: • All stars of a given variability type in a given region of the sky. • The data set is defined by interactively selecting a shape in a 2- or 3-dimensional diagram described above.
Input: Gaia Archive or Gaia-like simulated time series
Output:
Deliverables: 1. TN on existing tools for time domain related visualisation and the applicability for CU9. 2. Time domain visualisation related input to the CU9 documentation like SRS, SDD, STS and STR. 3. Releases of the time domain related visualisation software adequate for the data contents of the intermediate and final releases of the Gaia data.
Dependencies: WP981, WP982, WP983, WP985
Interfaces:
Remarks:

D References

[AB-026], Brown, A., Arenou, F., Hambly, N., et al., 2012, *Gaia data access scenarios summary*,
GAIA-C9-TN-LEI-AB-026,
URL http://www.rssd.esa.int/llink/livelink/open/3125400

Hambly, N., Read, M., Mann, R., et al., 2004, In: Ochsenbein, F., Allen, M.G., Egret, D. (eds.) Astronomical Data Analysis Software and Systems (ADASS) XIII, vol. 314 of Astronomical Society of the Pacific Conference Series, 137, ADS Link

Hambly, N.C., Collins, R.S., Cross, N.J.G., et al., 2008, MNRAS, 384, 637, ADS Link

Holl, B., Lindegren, L., 2012, A&A, 543, A14, ADS Link

Holl, B., Lindegren, L., Hobbs, D., 2012, A&A, 543, A15, ADS Link

[YI-003], Isasi, Y., Zaldua, I., Sartoretti, P., et al., 2012, *GOG User Guide*,
GAIA-C2-UG-UB-YI-003,
URL http://www.rssd.esa.int/llink/livelink/open/2775659

[TJP-013], Prusti, T., 2012, *Announcement of Opportunity for the Gaia Data Processing Archive Access Co-Ordination Unit*,
GAIA-CG-PL-ESA-TJP-013,
URL http://sci.esa.int/science-e/www/object/index.cfm?fobjectid=51116