Announcement of Opportunity for the Gaia Data Processing Archive Access Co-Ordination Unit

Author: Timo Prusti
Document code: GAIA-CG-PL-ESA-TJP-013-01
Issue date: 19 November 2012

1 Introduction

The Science Management Plan (SMP, ESA/SPC(2006)45) defines the top-level scientific management of the Gaia mission and describes the respective roles of all parties involved. In the SMP the data processing task was entrusted to the scientists in the ESA member states. In November 2006, ESA issued an Announcement of Opportunity (AO) for the data processing and analysis. For this call ESA received one proposal. The proposal was by the Gaia Data Processing and Analysis Consortium (DPAC) which was selected for the task by the Science Programme Committee (SPC, ESA/SPC(2007)20). In the AO which led to the selection of DPAC there was one element which was explicitly excluded from the call. The activities required for the Gaia archive access were not solicited. This was done in order to achieve a more mature view of the actual Gaia data products and required access mechanisms. The AO for DPAC stated explicitly that the Archive Access Co-Ordination Unit will be solicited in a later, separate AO. This call is for the Archive Access Co-Ordination Unit.

The AO which led to the selection of DPAC stated that the later addition of the Archive Access unit must be done with the full (proven) agreement of the selected consortium. This is necessary as the Archive Access unit cannot work independently of the selected data processing consortium, DPAC. This was already taken into account in the management structure of DPAC with numbered Co-Ordination Units (CUs) from CU1 to CU9, with CU9 reserved for the future Archive Access CU. The purpose of this call is to complete DPAC by selecting CU9 through this AO. The SMP is an applicable document and as such has been included in the AO data package. Where and when relevant, the SMP takes precedence over what is described in this AO.

In March 2011 ESA issued a call for Letters of Interest for the community to participate in the Gaia Archive Preparation (GAP) group. The call resulted into 21 groups being included in GAP, allowing also scientists outside DPAC to participate in CU9 preparations. The preparatory work in GAP, in conjunction with the rest of the DPAC, has also been used by ESA to prepare the CU9 requirements for this AO. In addition to the scientific and technical requirements stemming from the GAP work, the schedule of establishing CU9 is determined by the schedule on which the Gaia data is planned to be released. The SMP determines that the Gaia Science Team (GST) defines a more precise schedule and the contents of the Gaia intermediate releases after consultation of the DPAC Executive (DPACE) and Astronomy Working Group (AWG). After the DPACE and AWG consultations, GST issued the Intermediate Data Release Scenario (GAIA-CG-PL-ESA-TJP-011-01), which is an applicable document and has been included in the AO data package. The issue of this AO also formally marks the end of the GAP group. As was stated in the AO for Letters of Interest to GAP, the activity will be terminated at the issue of the AO for CU9.

2 Gaia mission

This summary aims to describe concisely the top level scientific objectives of the Gaia mission, the satellite, and the mission operations and data processing ground segment. It further details the top level work breakdown within DPAC and the miscellaneous ESA contributions to the Gaia data processing as far as CU9 is concerned.

2.1 Scientific Objectives

Gaia is a space astrometry mission, a broad survey project following the measurement and operational principles of Hipparcos. It will help solving one of the most difficult yet deeply fundamental challenges in modern astronomy: to create an extraordinarily precise three-dimensional map of about one billion stars throughout our Galaxy and beyond. In the process, it will map their three-dimensional motions, which encode the origin and subsequent evolution of the Galaxy. Through comprehensive photometric and spectroscopic classification, it will provide the detailed physical properties of each star observed: characterising their luminosity, temperature, gravity, and elemental composition. This massive stellar census will provide the basic observational data to tackle an enormous range of important problems related to the origin, structure, and evolutionary history of our Galaxy.

Using on-board object detection, Gaia will survey more than one billion stars at the few micro arc-seconds level accuracy, complete to 20th magnitude, with extensive multi-colour multi-epoch photometry and spectroscopic measurements.

2.2 Mission and Satellite

The space segment is characterised by a 3-axis stabilized satellite, operated at the Lagrange point L2. Gaia will be systematically scanning the sky at a constant angular velocity according to a well-defined scanning law. The high-stability payload will systematically and repeatedly measure the astrometric, photometric and spectroscopic properties for some 1 billion celestial objects. On-board detection of each object brighter than a given signal-to-noise ratio at the observation epoch will ensure completeness of the sky to 20th magnitude in brightness. The telemetry rate of some 15 Mbps will ensure collection of all scientific data from the satellite.

2.3 Mission Operations Ground Segment

The mission operations ground segment is composed of the Mission Operations Centre (MOC) located at ESOC, and the associated ground stations and communication links. The Mission Operations Centre is responsible for all spacecraft mission operations, including: spacecraft status monitoring, payload health monitoring, spacecraft control, mission planning of spacecraft and payload activities, orbit determination and control, attitude determination and control, on-board software maintenance, and transmission of the raw satellite data to the data processing ground segment.

2.4 Data Processing Ground Segment

The data processing ground segment is formed by DPAC, which is a collaboration between a consortium of scientists in ESA member states and the Gaia Science Operations Centre (SOC). The data processing consortium as a whole (i.e. including the ESA provided elements and called DPAC hereafter) will be responsible for the development of the computer systems required for the processing of Gaia science data in order to fully support the scientific goals of the mission. This will include development and testing of the computer systems according to agreed standards and guidelines, management of the transition from development to operations of its systems, and maintenance of the systems until the end of the post-operational phase. DPAC will also be responsible for the operation of these processing systems. These activities are to: (a)

provide the operational environment for the processing systems; (b) operate the processing systems during the routine operations and post-operational phases; (c) deliver data products in a timely manner, consistent with the overall data processing schedule; (d) report anomalous system behaviour of the processing systems to the developers of the processing systems, such that it may be corrected without unnecessary delay.

The DPAC CUs are as follows (NOTE: CU9 is the CU to be established with this AO):

CU1 System Architecture: This CU is responsible for the system architecture to be used within the DPAC, including: (a) definition of the system architecture; (b) definition and coordination of the development and operational environments, including software, hardware and database technology; (c) definition and execution of the end-to-end system testing; (d) tailoring of ECSS standards and definition of Quality Assurance procedures; (e) provision of the central Gaia database, dissemination of data to the DPCs and integration of results.

CU2 Data Simulations: Validation and verification of the DPAC system(s) requires detailed simulated data for all aspects of the Gaia mission, and at various levels of detail, including the astronomical, spacecraft and instrument contributions to: (a) low-volume high-fidelity instrument simulations; (b) high-volume telemetry stream simulations for all instruments; (c) higher level data relevant for validation by each CU. CU2 will develop and maintain the simulator throughout the project lifetime.

CU3 Core Processing: The core processing tasks provided include to: (a) unpack, decompress, and process the science data retrieved from the MOC to provide rapid (approximately 1 day) monitoring (and feedback to ESOC) of the spacecraft and payload performances at the ultra-precise accuracy levels targeted by the mission; (b) provide the astrometric global iterative solution; (c) contribute to key parts of the reduction of the science data and the generation of the final products (pre-processing and core-processing tasks); (d) periodically re-process raw telemetry, taking advantage of improved calibration data and reduction systems; (e) assist with instrument characterization and calibration.

CU4 Object Processing: The object processing tasks include processing the astrometric and photometric data for more complex objects not handled by the astrometric core processing, specifically: (a) non-single stars (binary and multiple stars); (b) Solar System objects (asteroids, near-Earth objects, etc); (c) extended objects.

CU5 Photometric Processing: Photometric processing tasks include: (a) basic calibrations, photometric parameters, calibration model; (b) accumulation of mean flux information and epoch photometry; (c) pass-band characterisations and absolute flux calibrations; (d) variability detection; (e) flux-based and classification-based science alerts; (f) image restoration in 2d.

CU6 Spectroscopic Processing: Spectroscopic processing tasks include: (a) spectral extraction and background correction; (b) data sanity check and quick calibrations; (c) nominal calibrations and meta-processing; (d) single star radial and rotational velocities determination.

CU7 Variability Processing: Variability processing tasks include: (a) variability classification, period search and variability models; (b) feedback of variability analysis

on the calibration models; (c) catalogue exploration and checks of the variability database; (d) statistical analysis of the data; (e) external observation coordination; (f) variability announcements.

CU8 Astrophysical Parameters: Astrophysical parameter determination tasks include: (a) classification of discrete sources; (b) photometric and spectroscopic stellar parameterization; (c) treatment of interstellar extinction; (d) object cluster analysis; (e) luminosity, age and mass estimation.

CU9 Archive Access: The main elements for the archive access co-ordination unit are detailed below. CU9 is set up in this call to provide: (a) archive and interrogation systems for Gaia data products; (b) systems to make Gaia data products publicly available; (c) visualization tools for Gaia data products; (d) tools and documents describing the Gaia data products and their interrelation; (e) science enabling applications to ensure availability of tools necessary for Gaia catalogue utilisation; (f) education and outreach opportunities and activities.

The DPAC Data Processing Centres (DPCs) are responsible for running the code developed by the various CUs. Six DPCs are organized in a hub and spokes architectural structure. The central hub is supplied by ESA and located at ESAC, Madrid. The central hub maintains the mission database which forms the interface for all data exchanges between various processing elements in DPAC. The five other data processing centers are: Barcelona, Cambridge, Geneva, Toulouse, and Turin.

2.5 ESA Commitments

ESA has committed a substantial contribution to the ground data processing. It consists of the ESA Gaia Project Scientist, his deputy and the Gaia Science Operations Centre (SOC), which is responsible for a number of tasks critical to the timely reduction of all Gaia data. The ESAC Gaia Archive Core System will be a collaboration between Gaia SOC and the ESAC Archive group. The ESA elements of relevance to this CU9 AO are:

- the overall processing system architecture
- hosting, operation and maintenance of the main data base
- definition and execution of end-to-end system testing
- the development, integration, testing and operation of the data reception and archiving system
- technical support in definition and production of the final catalogue
- hosting the final catalogue archive, documentation and user access
- hosting the Gaia web-portal and the Gaia mission helpdesk for questions by Gaia data users
- acquisition, maintenance and operation of all the hardware and infrastructure required for the above tasks

Taking due account of the above defined undertakings, the ESA SOC at ESAC will form part of DPAC.

3 Archive Access requirements

The implementation plan of the DPAC Archive Access Co-Ordination Unit (CU9) needs to be detailed in the response to this AO. The schedule must be compatible with the overall schedule and budgetary constraint of the Gaia Programme. CU9 will be the only responsible for making all Gaia data available, except the Science Alerts which are released directly by the respective processing entities. For Science Alerts the CU9 responsibility is limited to recording a full set of all Gaia alerts and therefore coordination with the Science Alerts team is required to be able to start compiling the alerts as soon as they start. CU9 must be ready to issue the first intermediate data release, currently estimated for launch+22 months. CU9 must also assume 6 months for commissioning and performance verification activities and 5 years of operations. After the nominal 5 year operational period CU9 must assume a three year post operational period for the final data processing and catalogue issue. As defined in the SMP, the Gaia mission follows a 'no data rights' policy. This is a condition imposed on all DPAC work and is of special importance to CU9. As stated in the SMP, the GST will define a policy with regard to data access in DPAC.

3.1 Management and Coordination requirements

The AO issued in 2006 for the Gaia data processing resulted in DPAC which forms a complete Gaia science data processing entity except for the Archive Access element. This AO is for the Archive Access Co-Ordination unit (CU9) to complete DPAC. The 2006 AO already noted that CU9 will be added to an existing consortium and therefore must demonstrate full proven agreement with DPAC. The management and coordination tasks within CU9 should be explained as well as the interface to the ESA Archive group. Furthermore, the management and coordination tasks with respect to the rest of DPAC and toward DPACE must be defined and any deviations from practices followed by other CUs must be fully explained. The response must outline the reporting approach CU9 intends to use. Any deviations from practices done by other CUs must be detailed. The planning for the operational phase schedule must assume the Gaia Intermediate Data Release Scenario (GAIA-CG-PL-ESA-TJP-011-01) as the baseline.

3.2 Documentation requirements

The Gaia catalogue will be the ultimate mission product for the astronomical community. In order to fully utilize the Gaia catalogue it must be properly documented. The response to this AO must demonstrate that adequate documentation will be available for the Gaia catalogue users covering the needs of all user categories, from novices to experts. Proposers must explain the relations between documentation written within the existing DPAC CUs and provided to the community through CU9. The proposal must also contain details of the documentation provided by CU9 to the Gaia catalogue users.

3.3 Architecture and Technical Development requirements

The Gaia archive architecture and technical developments plans must be given in the proposal. The response must detail the connections of the proposed solutions to the overall DPAC architecture and technical development practices. The response must outline the performance and availability requirements, product and quality assurance

practices, testing requirements, and configuration control approach. Deviations from practices followed across the rest of DPAC must be detailed.

3.4 Validation requirements

Data validation is an essential activity required prior to data release. For Gaia data validation is a multi-stage process where part of the exercise is conducted in CUs producing the data and part is done in CU9. The proposal must clarify the division of validation tasks within DPAC and detail CU9 activities in this task. The CU9 data validation activities can be conducted in various ways. Some methods are very close to performing a scientific analysis of a data set. Given the principle of 'no data rights' having been adopted very early on for the Gaia mission, it is mandatory to indicate that the policy is respected by all validation methods. The SMP identifies the GST as the entity defining procedures for any early access to the data. Therefore the proposal should take into account that data access for the groups involved should be approved by GST and should respect the 'no data rights' policy.

3.5 Operations requirements

The proposal must explain the operational plans for CU9 activities. The plans must indicate the interfacing toward already existing DPAC entities, CUs and DPCs. The response must outline the performance and availability requirements, product and quality assurance practices, testing requirements, and configuration control approach. Deviations from practices followed across the rest of DPAC must be detailed. The operational coordination between the ESA Archive group and the rest of CU9 must also be detailed.

3.6 Education and Outreach requirements

The potential of Gaia for education and outreach activities has been recognized early on in the mission and has been assumed as an integral part of the Gaia programme. Therefore it is necessary to plan education and outreach activities as part of CU9. Gaia related education and outreach activities will be subject to the overall ESA Communications Plan and detailed Gaia Communication Plan at a later stage. The response should assume coordination of all education and outreach activities with ESA in order to achieve maximum impact and coherence of output.

3.7 Science enabling applications requirements

The extent and accuracy of Gaia data has been early on recognized as a challenge to many existing tools used in scientific analysis. In order to handle the Gaia data and to exploit fully its potential it is necessary to evaluate suitability of existing tools and develop new ones as needed. The response should explain the reasoning for specific tools needing to be developed in the CU9 context. The proposal should also outline a strategy to cope with newly emerging requirements and to benefit from new tools which are suitable for Gaia data analysis.

3.8 Visualization requirements

The three dimensional visualization needs have always been considered necessary for Gaia due to its capability to provide distances. In a more general sense the scientific power of Gaia is in the multi-dimensionality of the catalogue. This multi-dimension aspect together with a large number of catalogue entries requires novel visualization tools to enable the science available in the catalogue. The response must explain the planned visualization tools for Gaia data. The proposal must also outline a strategy to benefit from the likely new developments both in hardware and software from the visualization perspective.

4 Proposal submission details and evaluation timeline

Proposers from ESA Member States are invited to respond to this Announcement of Opportunity. The proposal length is limited to 70 pages. The proposal should contain as appendices signed letters of initial commitment from the relevant funding authorities. Additional material may be included as further appendices, but they should not contain any material essential to the proposal evaluation.

Proposals should be submitted electronically no later than 10 January 2013 to:
Timo.Prusti@rssd.esa.int

In addition a copy of the complete proposal must be sent to each of the funding agencies which will later be requested to provide financial support to the CU9 activities.

The submitted proposal(s) shall be reviewed by an evaluation committee appointed by ESA. They shall evaluate the proposal(s) primarily based on:

- Scientific quality and completeness of the analysis and proposed processing
- Background and expertise of the proposed participants
- Adequacy of resources assigned to development, processing and management
- Compliance with the SMP and AO

It should be noted that following the proposal selection by the SPC, a negotiation process between ESA and the (national) funding agencies is foreseen. These negotiations should result in a multilateral agreement (MLA), identifying the willingness of all parties involved to fund the proposed activities such as to allow achieving the goals of CU9 as specified in the proposal. In the case of Gaia with an existing MLA for the already approved DPAC elements it is likely that the existing MLA will be amended to include CU9 activities.

The AO package is available through http://sci.esa.int/Gaia. Further questions on the AO itself, or on the associated approval process, should be addressed by E-mail to:
Dr. Timo Prusti
Gaia Project Scientist
Timo.Prusti@rssd.esa.int