Gaia Catalogue and Archive Software Requirements and Specification

prepared by: William O'Mullane and the DPACE
approved by: Approved by DPACE and GST
reference: GAIA-C9-SD-ESAC-WOM-033-01
issue: 01
revision: 0
date: 2009-06-20
status: Issued

Abstract

This is the requirements specification for CU9. Other CUs made SRSs as part of their start up activities. However CU9, starting later, needs to have several specific requirements laid out. There may be later "product" SRSs within CU9.

Document History

Issue	Revision	Date	Author	Comment
1	0	2009-06-17	WOM	Approved
D	15	2009-05-20	WOM	Finalise RD comments
D	14	2009-04-17	LE	Addition for announcements
D	13	2009-03-31	WOM	Comments from RD
D	12	2009-03-27	WOM	Post DPACE/GST meet - some open points adde, discussion of release removed.
D	11	2009-02-22	WOM	NW,AB comments
D	10	2009-02-02	WOM	spell check - send to DPACE/GST as input for joint meeting
D	9	2009-01-23	WOM	More from CJ on science alerts
D	8	2009-01-19	WOM	Comments from SR. Elaborate intermediate releases
D	7	2009-01-14	WOM	Comments from CJ.
D	6	2008-12-17	WOM	Comments from SK, contents of archive added .
D	5	2008-11-30	WOM	Comments from CA included.
D	4	2008-11-04	WOM	removed some questions comments - send to GST before GST25
D	3	2008-10-06	WOM	XL comments from wiki added, outreach in diagram
D	2	2008-09-30	BAS	corrections and clarifications from UB, addition of outreach aspects & field explanations as pf DPACE-07
D	1	2008-09-24	WOM	Comments from LA MB
D	0	2008-07-21	WOM	First draft

Contents

1 Introduction

This document lays out the initial requirements for the Gaia Catalogue and Archive (Sect. 2). This is a CU9 task and will be taken over by CU9 when it is formed. This document represents the requirements and views of the DPACE (Data Processing and Analysis Consortium) and the GST (Gaia Science Team) and will give some guidance and limits to CU9 as it is being set up. The exact specification of the archive contents, i.e. the Catalogue, will be defined at a later stage.

CU9 will probably produce more Software Requirement Specifications for sub systems or products within CU9. This may be seen as the top-level requirements from which these documents may trace. As such, CU9 requirements will not be added to the SSS (WOM-018) which contains the processing and operation requirements for DPAC.

1.1 Objectives

The archive must deliver the Gaia data products to the science community. This mainly means making a suite of search and data mining tools available to the community to filter and interpret the catalogue.

Each CU will have to contribute to CU9 in their area of expertise. This is analogous to the way each CU contributes simulation information to CU2.

1.2 Scope

This document represents the top-level requirements for CU9. It is superseded only by the SMP. However, note should be taken of general requirements on all CUs in the DPAC SSS (WOM-018).

Furthermore a science requirements document for the Archive and Catalogue will be produced by the Project Scientist, such a document will be applicable and may override some requirements in this document.

1.3 Assumptions

It is assumed that CU9 will be an integral part of DPAC and that indeed many existing members of DPAC will form part of CU9. It should remain clear however that no funding for CU9 activities was included in the DPAC proposal as specified in the AO.

In the Cost At Completion (CAC) for Gaia some technical effort for the archive is foreseen — CU9 should assume that technical support from ESAC will be included in the planning. The

likely area for this support would be in the back end database (interrogation and storage systems below) and virtual observatory interfaces. ESA will certainly provide hardware and hosting for the catalogue. The CU-T for CU9 could be provided by ESAC.

1.4 Applicable Documents

Whenever applicable documents change, a change may be required in this document. The applicable documents are listed here for clarity - the full references are in Sect. 1.5.

ESA/SPC(2006	Gaia Science Management Plan (SMP)
WOM-017	DPAC Project implementation Plan (PIP)
WOM-018	DPAC Software System Specification
WOM-011	DPAC Software Engineering Guidelines
TL-001	DPAC Product Assurance Plan

1.5 Reference Documents

[WOM-018], CU1 and DPAC CU leaders, 2008, *DPAC Software and System Specification*, GAIA-C1-SP-DPAC-WOM-018, URL http://www.rssd.esa.int/llink/livelink/open/2786798

[ESA/SPC(2006)45], Gaia Project Scientist, 2006, *Revised Gaia Science Management Plan (SMP)*, ESA/SPC(2006)45, URL http://www.rssd.esa.int/llink/livelink/open/2720576

[JH-001], Hernandez, J., 2009, *Main Database Interface Control Document*, GAIA-C1-SP-ESAC-JH-001, URL http://www.rssd.esa.int/llink/livelink/open/2786145

[JH-004], Hernandez, J., ter Linden, M., 2009, *Main Database Software Requirements Specification*, GAIA-C1-SP-ESAC-JH-004, URL http://www.rssd.esa.int/llink/livelink/open/2698216

Hogg, D.W., Lang, D., 2008, In: American Institute of Physics Conference Series, vol. 1082 of American Institute of Physics Conference Series, 331–338, ADS Link

[TL-001], Levoir, T., Damery, J., Hoar, J., et al., 2007, *DPAC Software Product Assurance Plan*, GAIA-C1-PL-CNES-TL-001, URL http://www.rssd.esa.int/llink/livelink/open/2439085

O'Mullane, W., Li, N., Nieto-Santisteban, M., et al., 2005, In: International Conference on Web Services,
Also MS technote `http://arxiv.org/pdf/cs.DC/0502072`

[WOM-017], O'Mullane, W., Drimmel, R., Mignard, F., et al., 2008, *Project Implementation Plan (PIP)*,
`GAIA-CD-PL-ESAC-WOM-017`,
URL `http://www.rssd.esa.int/llink/livelink/open/2812481`

[WOM-011], O'Mullane, W., Hoar, J., Levoir, T., et al., 2008, *Software Engineering Guidelines for DPAC*,
`GAIA-C1-UG-ESAC-WOM-011`,
URL `http://www.rssd.esa.int/llink/livelink/open/2760364`

1.6 Definitions, acronyms, and abbreviations

The requirements set out in this SRS follow the labeling scheme:

CU9-*ARC*-*X*-*SCOPE*-*xxx*

where *ARC* is the WP number or Software Product Label (e.g. ITG for "Interrogator") or the Module in the case that there are multiple modules in the product , , *X* is either `S` (for **S**cientific), `T` (for **T**echnical), `Q` (for **Q**uality Assurance), or `M` (for **M**anagement). *SCOPE* is a three or four letter scope specification of the requirement following the identified list of possible values of TL-001 whenever feasible. *xxx* is a monotonically increasing counter for every unique combination of *X-SCOPE*.

Each requirement is presented with its unique label and a number of attributes in accordance with TL-001 in the following form:

*CU9-ARC-X-SCOPE-***020**	$C.v$	Verific.	Status
Description			
Parent: Parent			

with (see TL-001 for lists of allowed values, meanings and valid ranges):

CU9-ARC-X-SCOPE-xxx	The unique identifier of the requirement (see above)
$C.v$	Version number of the requirement composed of major (C) and minor (v) part
Verification	Envisaged validation method of requirement - this will be either AUT for automated or MAN for Manual. The Software Test Plan defines precisely how the requirement is verified.

Status	Status identifier
Parent	Higher level requirement or requirements, comma separated list

SCOPE is a three character scope specification of the requirement one of :

REQ: concerning requirements,
SPC: concerning specification activities,
DES: concerning design activities,
IMP: concerning implementation activities,
INT: concerning integration activities,
VAL: concerning validation activities,
ACP: concerning acceptance activities,
MNT: concerning maintenance activities,
OPS: concerning operation activities,
FUN: concerning functional requirements,
PLN: concerning organisational requirements (including planning, phases, reviews, etc.),
SUP: concerning support requirements,
PRF: concerning performance requirements (including dimensioning, resource estimation, etc.),
RSK: concerning risk management requirements,
CTL: concerning verification and control activities,
CNF: concerning configuration management activities,
DIM: concerning dimensioning activities.

The following is a complete list of acronyms used in this document. The following table has been generated from the on-line Gaia acronym list:

Acronym	**Description**
ADQL	Astronomical Data Query Language
AGIS	Astrometric Global Iterative Solution
AO	Announcement of Opportunity
API	Application Programming Interface
ASCII	American Standard Code for Information Interchange
AUT	AUTomated
BTI	Bad Time Interval
CAC	Cost at Completion
CCD	Charge-Coupled Device
CIF	Community InterFace
CSV	Comma-Separated Value (database output format, e.g., for MS Excel)

CU	Coordination Unit (in DPAC)
DBMS	DataBase Management System
DOC	DOCumentation
DPAC	Data Processing and Analysis Consortium
DPACE	Data Processing and Analysis Consortium Executive
DPC	Data Processing Centre
ESA	European Space Agency
ESAC	European Space Astronomy Centre (VilSpa)
FITS	Flexible Image Transport System
FPA	Focal Plane Assembly (Focal Plane Array)
GREAT	Gaia Research for European Astronomy Training
GST	Gaia Science Team
GTI	Good Time Interval
HEALPix	Hierarchical Equal-Area iso-Latitude Pixelisation
HK	Housekeeping (also denoted H/K)
HLP	Help Desk
HTM	Hierarchical Triangular Mesh
HTML	HyperText Markup Language
IAU	International Astronomical Union
ICD	Interface Control Document
IDU	Intermediate Data Update
ING	Ingestor
ITG	Interrogator
LDAP	Lightweight Directory Access Protocol
MAN	MANual
MB	MegaByte
MDB	Main DataBase
MPC	Minor Planet Centre
PBO	Public Outreach
PIP	Project Implementation Plan
RD	Reference Document
SAA	Science Anomaly Alerts
SDSS	Sloan Digital Sky Survey
SK	Spare Kit
SMP	Science Management Plan
SQL	Structured Query Language
SR	Status Report
SRS	Software Requirements Specification
SSAP	Simple Spectral Access Protocol
SSS	Software and System Specification
TBC	To Be Confirmed

UB	University of Barcelona (Spain)
URL	Uniform Resource Locator
UWE	Unit-Weight Error
VO	Virtual Observatory
WP	Work Package
XML	eXtensible Markup Language

2 General description and requirements

The Gaia Catalogue is the collection of information made publicly available. The Gaia Archive is the storage and delivery mechanism for that data. At time of writing this is considered to be a web-based systems such as the Sloan Digital Skyserver (see `http://skyserver.sdss.org`).

2.1 Context

Essentially the archive containing the Gaia Catalogue may be considered an MDB (Main Database) client. Hence the DPC where the catalogue is produced (or extracted from the MDB) needs a complete copy of the MDB. There may, however, be mirrors of the archive which do not have MDB access. The context is presented in the context diagram Fig. 1.

Hence the input to the archive comes from the DPAC MDB (JH-004). A good indication of the contents and access are given in the MDB ICD (JH-001). The contents of the archive at any point in time, there will be at least one intermediate release, must be agreed by CU9 with the GST and DPACE. The audience for the archive is certainly the broadest science community possible. One must also consider outreach and educational possibilities.

2.2 Decomposition

Although CU9 may decompose the system differently we define a few logical blocks for the purposes of defining requirements. These are detailed starting in Sect. 4 but are described briefly here to give an overall view of the archive and catalogue.

The main components considered in this document appear in Fig. 2 and are discussed further below. The components are :

- **Ingestor (ING)** to populate the archive Sect. 4
- **Storage System** Physical disk, machines and DBMS, these implementation details

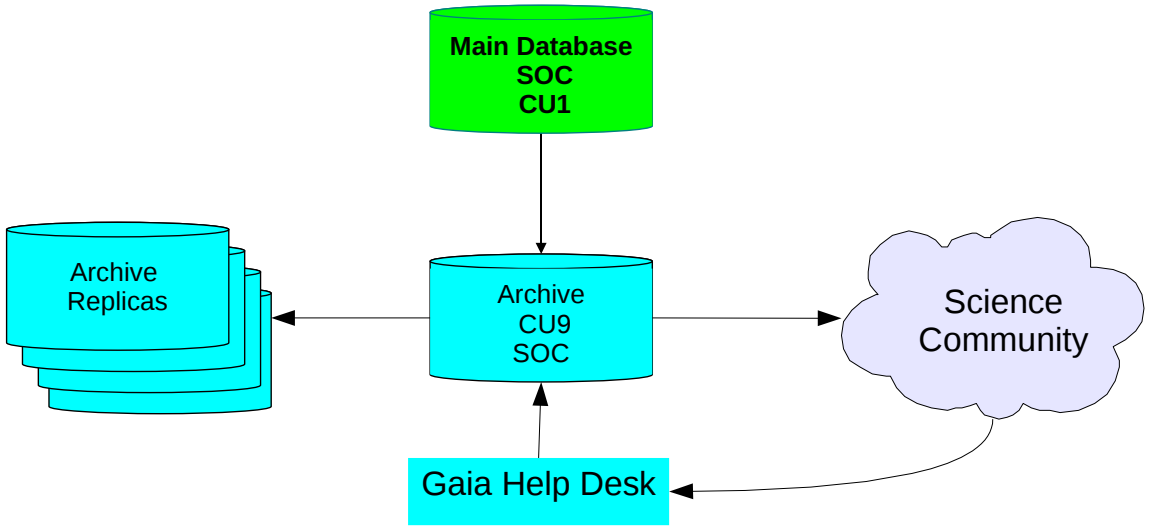

FIGURE 1: Context of the archive showing its data coming from MDB and interfaces to the community.

are not discussed in this document.

- **Interrogation System (ITG)** to effectively query the archive Sect. 5
- **Advanced Applications (ADV)** for value added access Sect. 6
- **Documentation (DOC)** to allow users understand the archive Sect. 7
- **Science Alerts (SAA)** Anomaly based Science Alerting Sect. 8
- **Public Outreach (PBO)** to engage general public Sect. 10
- **Help desk (HLP)** to answer users' questions Sect. 9
- **Community Interface (CIF)** to provide a consolidated portal Sect. 3

In addition to the specific requirement relating to these components there are some requirements concerning the archive and catalogue in general which are listed in Sect. 2.4.

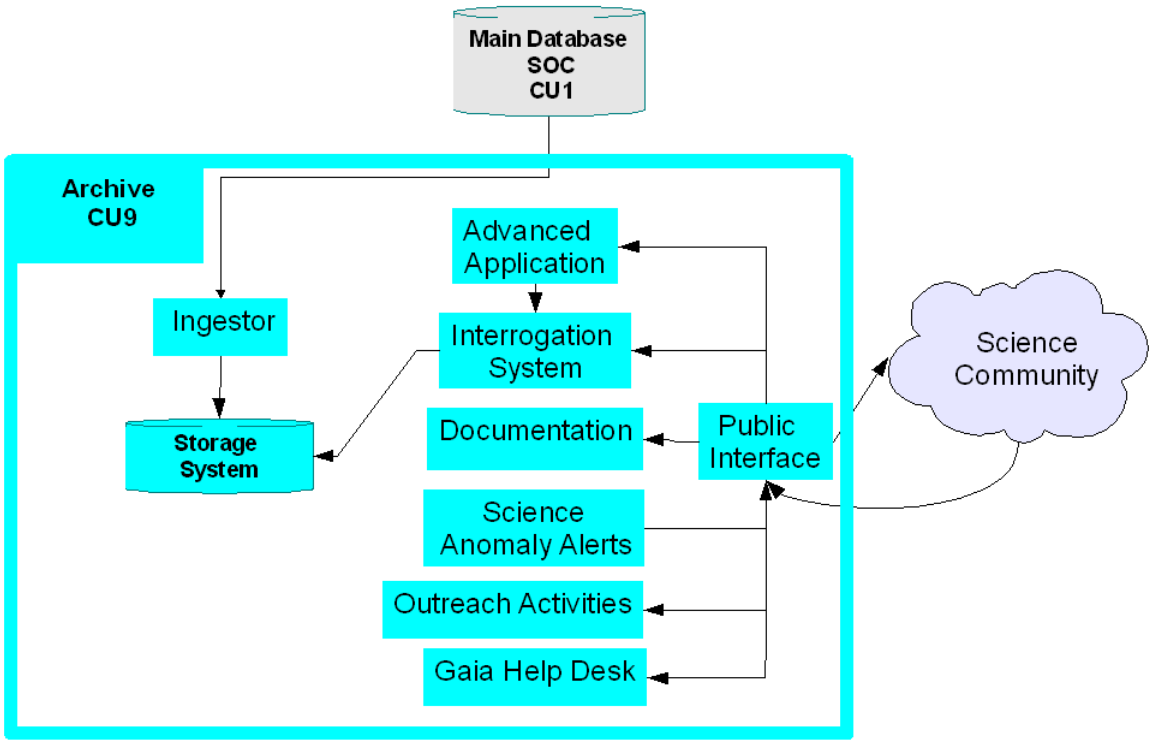

FIGURE 2: Logical decomposition of the archive.

2.3 Areas for further study

At the time of writing and in this format it is impossible to catch *all* of the requirements. The user community of the archive does not yet exist and the contents are unclear. There are some concepts which today are interesting but can not, and should not, be formulated as *requirements*. These topics should be revisited, along with others, at a later date.

2.3.1 Living Archive

The idea to be able to add to the archive is very appealing to many yet raises many questions. The term *living archive* was provided by Anthony Brown (GST member). He states :

> The catalogue will also contain intermediate data which allow a reprocessing of the observations. In particular the astrophysical parameters can be improved by incorporating complementary information obtained from other surveys or follow-up observations. For astrometric data reprocessing is relevant for multiple sources with uncertain solutions (which can be improved by incorporating follow-up radial velocity observations) or, for example, in the case of proper motion measurements for faint stars in a dwarf galaxy orbiting the Milky Way one could combine the measurements for all stars to derive a (more precise) mean motion and parallax for the dwarf. Hence the Gaia catalogue will not be a static release but will evolve over time as more information is added to it.

Of course allowing for such a system means a real quality control problem. How do we accept an update? What about conflicting updates from different parties? Once we have multiple values what about internal (self) consistency (e.g., zero points, temperature scales).

2.3.2 Archive as a Model - or model of the contents

Modeling was quite a topic an the first GREAT (Gaia Research Education and Training) meeting. The notion of how to compare ones model to the actual catalogue perhaps should be answered by the archive creators. Should a user be allowed to ask how well does my model fit the data? Or which data does and does not fit my given model.

Perhaps related is the idea of the catalogue as a model of the data described by David Hogg here `http://arxiv.org/abs/0810.3851`.

2.3.3 Cross references to other catalogues

There is certainly a need to have other catalogue information available to the Gaia users. Perhaps the virtual observatory will do this but currently VO deals with small parts of large data sets while for Gaia we would need the whole thing. CU9 will need to decide how to best facility the Gaia Archive users in this area - should a certain number of catalogues be imported so the same tools available for Gaia would be available for them?

2.3.4 Virtualisation

Currently there is a lot of virtualisation going on in data centres. One aspect of this is commercial *cloud computing* another is storage in someone else's servers. For example Amazon claim to provide free storage for data in the public domain `http://aws.amazon.com/publicdatasets/`.

This in itself may be interesting if they replicate it about for easy access. But the concept of virtualisation may be more interesting. We frequently hear that for large new astronomical data sets we must bring the programs to the data not the data to the programs. It seems virtualisation provides and excellent way to do this - once could make a range of virtual machines available *in the archive* for user to install and run their programs on. Then just let them download the result.

The local storage provided by CasJobs and VO Space partially does this but the complexity of code which can be sent is limited i.e. for CasJobs one can only send SQL programs. Virtualisation with appropriate access libraries would allow almost any code to be run. This may be the only way to bring fully general models (Sect. 2.3.2) to the data.

This obviously has very serious security implications.

2.4 General Requirements

2.4.1 Functional and Technical

CU9-ARC-M-FUN-020	5.1	MAN	Draft
CU9 shall produce the archive such that it is portable. It must be installable on a mirror site in under one day (not including data copying).			

2.4.2 Management and reporting

CU9-ARC-M-PLN-020	5.1	MAN	Draft
CU9 shall operate within the existing DPAC management structure. the CU9-L will become a member of DPACE and work and report within DPAC in the same manner as other CU leaders.			
Parent: SMP			

CU9-ARC-M-PLN-040	5.0	MAN	Draft
CU9 shall develop all code and documentation within the DPAC code repository where it shall be visible to all DPAC members but modifiable only by CU9 members.			

CU9-ARC-M-PLN-060	5.0	MAN	Draft
CU9, like other CUs, should follow the engineering guidelines laid down in O'Mullane et al. (WOM-011).			

2.4.3 Catalogue Releases

CU9-CAT-M-PLN-020	5.0	MAN	Draft
CU9 shall agree the contents of any catalogue release with DPACE and GST as per the SMP.			
Parent: SMP			

CU9-CAT-M-PLN-040	5.0	MAN	Draft
CU9 shall release the final catalogue within three years of the end of Gaia spacecraft operations.			
Parent: SMP			

CU9-CAT-M-PLN-060	5.1	MAN	Draft
CU9 shall make release 1 of the catalogue within three years (TBC by GST&DPACE after launch) of the commencement of nominal Gaia spacecraft operations.			
Parent: SMP			

CU9-CAT-M-PLN-080	6.0	MAN	Draft
CU9 shall make release 2 of the catalogue within five years (TBC by GST&DPACE after launch) of the commencement of nominal Gaia spacecraft operations.			

3 Community Interface (CIF)

A consolidated interface must be presented to the community. In today's terms this is a portal to access all Gaia data, documentation and support. Obviously the Virtual Observatory or other communities may provide some of what is needed and it is assumed CU9 will look at such technologies and report on their usefulness.

3.1 Requirements

CU9-CIF-T-FUN-020	5.0	FUN	Draft
A consolidated interface to the archive and catalogue shall be constructed for the community. Henceforth this will be termed CIF.			

CU9-CIF-T-FUN-040	5.0	FUN	Draft
The CIF shall accommodate both novice and advanced users with an easy transition. The novice and/or general public should in many cases be using the same tools as the professional astronomer.			

CU9-CIF-T-FUN-060	5.0	FUN	Draft
Access to the help desk shall be clearly visible at all times in all parts of the CIF.			

CU9-CIF-T-FUN-080	5.0	FUN	Draft
The CIF shall not require registration for access to basic functionality. It is acceptable to require registration/log in for advanced areas.			

CU9-CIF-T-FUN-100	5.0	FUN	Draft
In the case the CIF requires registration this should be automatic and not requiring human intervention on the archive side. Hence: a simple Wiki-type automatic registration where the users fill in required details and an email address. It is acceptable that requests for more than some minimum amount of resources would be done via email/manually e.g. increase in local storage.			

CU9-CIF-T-FUN-120	5.0	FUN	Draft
The CIF shall include a sky browser type interface such as provided by Skyserver or Google sky.			

CU9-CIF-T-FUN-140	5.0	FUN	Draft
The CIF shall include monitoring tools to automatically track usage an problems with the system. Frequency of use of different resources should be tracked to aid in decision making on deployment of hardware.			

CU9-CIF-T-FUN-160	5.0	FUN	Draft
The CIF shall be *localizable*. It should be easy to add new language versions of pages. This could be done using some form of content management system - a decision for CU9.			

CU9-CIF-T-FUN-180	5.0	FUN	Draft
The CIF shall have a section recording relevant publications using Gaia data. It should also contain a counter of number of publications, citations etc.			

CU9-CIF-T-FUN-200	5.0	FUN	Draft
The CIF shall have a section recording accepted corrections to the published catalogues.			

CU9-CIF-T-FUN-220	5.0	FUN	Draft
The CIF shall have documentation on how to use and navigate it. This should include tutorials for different common tasks.			

4 Archive Ingestor

Data from the MDB is nominally extracted to files (JH-001) which must then be ingested in to the archive system. This will involve filtering of data especially for an intermediate release.

4.1 Requirements

CU9-ING-T-FUN-020	5.0	AUTO	Draft
The CU9 ingestor shall operate in accordance with the MDB ICD (JH-001).			

CU9-ING-T-FUN-040	5.0	AUTO	Draft
The CU9 ingestor shall operate in a distributed manner.			

CU9-ING-T-FUN-060	5.0	AUTO	Draft
The ingestor shall not take longer than one week to ingest the Catalogue.			

5 Interrogator

An efficient query engine will need to be presented the community in a uniform manner. One may naively think this is a DBMS with SQL access but there is more involved. The DBMS needs to be tuned to answer the type of queries the astronomy community will ask. The system may also require some extra *functionality*, such as HEALPix or HTM indexing to make spatial queries faster, or coordinate transformations and the computation of galactic space velocity components.

Hogg & Lang (2008) provides an interesting view of modern archive abilities and activities.

5.1 Requirements

CU9-ITG-T-FUN-020	5.2	AUTO	Draft
The interrogator shall provide query-like (SQL or equivalent) access to the Gaia Catalogue.			

CU9-ITG-T-FUN-040	5.1	AUTO	Draft
The interrogator shall allow the user to ”drill down” in the data to the level of individual CCD transits/spectral images.			

This transit data will most probably be IDU data or better - not the raw data coming from Gaia which would be very difficult for anyone to use. In all cases GST/DPACE sign off on the final contents.

CU9-ITG-T-FUN-060	5.2	AUTO	Draft
The interrogator shall allow the user to retrieve all relevant auxiliary data (e.g. calibration, time-correlation, attitude, optical observations, reconstructed orbit of Gaia, time transformations, meteorological data, sky background etc.) for a given transit.			

CU9-ITG-T-FUN-080	5.0	AUTO	Draft
The interrogator shall provide access to all released versions of the Catalogue.			

CU9-ITG-T-FUN-100	5.0	AUTO	Draft
The interrogator shall provide identified users with limited ”local space” to store query results and to upload private data.			

CU9-ITG-T-FUN-120	5.1	AUTO	Draft
The interrogator shall allow queries involving different catalogue versions and data in the users ”local space”.			

CU9-ITG-T-FUN-140	5.0	MAN	Draft
The CU9 interrogator shall be accessible via pertinent Virtual Observatory protocols.			

At time of writing SkyNode, ADQL, ConeSearch, TAP , SSAP, SAP(SIAP) would be pertinent protocols for Req. CU9-ITG-T-FUN-140.

CU9-ITG-T-FUN-160	5.0	AUTO	Draft

The interrogator shall allow users to extract data in multiple easy-to-use formats, at least those in the following list.

- CSV : comma-separated values (ASCII)
- FITS
- HTML
- VOTable (XML)

CU9-ITG-T-FUN-180	5.0	AUTO	Draft

The interrogator shall allow users to upload data (see Req. CU9-ITG-T-FUN-100 in the same formats supported for download, see Req. CU9-ITG-T-FUN-160.

CU9-ITG-T-FUN-200	5.0	AUTO	Draft

The full functionality of the interrogator shall be exposed as a remotely accessible API to allow complex data mining programs to be written by users.

CU9-ITG-T-FUN-220	5.0	AUTO	Draft

The API Req. CU9-ITG-T-FUN-200 should also support the delivery of data on asteroids to the IAU Minor Planet Centre (MPC).

CU9-ITG-T-FUN-240	5.0	AUTO	Draft

The interrogator shall return helpful and clear error messages for erroneous queries.

5.1.1 Performance Requirements

CU9-ITG-T-PRF-020	5.2	AUTO	Draft

A series of typical queries to be answered by the interrogator shall be developed. These queries with execution times shall be published with each version of the interrogator. The queries should at least include the items of the following list.

- Per Source queries e.g. all data about source X or all astrometric data for source X.
- Spatial Queries e.g. all sources near X.
- Temporal queries e.g. all sources visible from P at time T or all sources/transits observed between T1 and T2.
- Temporal Spatial Queries e.g. all sources near X at time T.
- Variability queries, e.g. all sources with Delta m1...
- Kinematic queries, all sources with radial velocity between X and Y and parallax between p1 and p2 -this would be particularly useful for clusters, moving groups etc.
- Per Source type queries e.g. all sources of type X.
- General Queries e.g. all sources with magnitude between m1 and m2 and parallax $>$ p
- Cut queries on e.g. all objects with parallax error $<$ X(and/or photometric and/or radial velocity error) which are also variable (or multiple or ..)
- Satellite specific queries e..g Which sources were in FOV1 at time T.
- ...

CU9-ITG-T-PRF-040	5.0	AUTO	Draft

Any query executed by the interrogator shall not take longer than scanning the physical volume containing the data. Queries taking longer should be cancelled and the user should be warned.

Normally this will require the helpdesk to help the user reformulate their query.

CU9-ITG-T-PRF-060	5.0	AUTO	Draft
Query execution should be distributed such that queries finish in a time inversely proportional to the number of processors available for the query.			

CU9-ITG-T-PRF-080	5.0	AUTO	Draft
The system should have at least two modes or queues: interactive and batch. Queries in interactive mode should be stopped after 30 seconds if they do not finish. Queries in batch mode should be terminated at some time limit agreed for the system e.g. see Req. CU9-ITG-T-PRF-040.			

This is in no way meant to prevent full programmable access to the system by scientists. Queries will need to be well constructed however and some queries will simply never finish, usually these may be *rewritten* with a a little help to run in a few minutes. The system however has no way of know which will actually finish and which will not hence some *time limit* will need to be introduced.

In our work on the CasJobs system (O'Mullane et al., 2005) we found a power law correlation between query length and number of queries. The vast majority of queries took less than one minute to complete and the system could cope with hundreds of them (no need to limit). Some queries took longer - several minutes to several hours. The system could only handle about three such queries simultaneously hence they were "queued" in a batch like system with email notification etc. This approach prevented any query from bringing down the database and gave pretty much all users what they wanted.

6 Advanced Applications

Some advanced interrogation tools should be provided along with the archive. 3D visualisation tools for example as well as perhaps a sky navigator.

6.1 Requirements

CU9-ADV-T-FUN-020	5.1	AUTO	Draft
CU9 shall provide 3D visualisation tools on top of the archive utilising the best available technology. This shall allow exploration of Gaia's phase space map e.g. time and motion are also essential.			

CU9-ADV-T-FUN-040	5.0	AUTO	Draft
CU9 shall provide normal 2D visualisation tools on top of the archive when 3D devices are not available.			

CU9-ADV-T-FUN-060	5.0	AUTO	Draft
CU9 shall provide rich interconnectivity between the different types of Gaia data available in the catalogue. For example any virtual image of the sky should be mouse sensitive to allow extraction of catalogue information or spectra for object on the image.			

CU9-ADV-T-FUN-080	6.0	AUTO	Draft
CU9 shall provide access to cross match facilities to other major catalogues.			

This (Req. CU9-ADV-T-FUN-080) may be achieved by providing access to Virtual Observatory tools.

7 Documentation

A comprehensive set of documentation will be needed to describe the data, its processing , validation and access. This is a very large part of the archive activity.

7.1 Requirements

CU9-DOC-S-ACP-020	6.0	MAN	Draft
The contents of the documentation shall be agreed with the GST and no archive release shall be made without documentation.			

CU9-DOC-S-FUN-020	5.2	MAN	Draft
CU9 shall solicit (from other CUs) and present in a uniform manner documentation about all algorithms used to process and validate the data contained in the archive.			

For example see the Hipparcos catalogue volume 1.

CU9-DOC-S-FUN-040	5.0	MAN	Draft
CU9 shall put together and present a detailed explanation of all the quantities provided in the archive.			

A good deal -if not all of this information will be in the MDB dictionary tool/ICD already JH-001.

CU9-DOC-S-FUN-060	5.0	MAN	Draft
CU9 shall put together and present a project history for Gaia in conjunction with the Project Scientist.			

CU9-DOC-S-FUN-080	5.0	MAN	Draft
CU9 shall fully document the archive and access protocols. Tutorials shall be provided to show users how to perform example queries and data extraction.			

CU9-DOC-S-FUN-100	5.1	MAN	Draft

CU9 shall provide statistical plots derived from the archive contents. These shall include as a minimum the following list.

- Source count map
- Scan coverage map
- Histograms of sources observed per minute/hour/day/year
- Statistics per source type
- Histograms per source parameter
- Error maps and error histograms per source parameter
- equivalents of all the statistical plots in Vol. 1 of the printed Hipparcos Catalogue.

There are many more possible plots these are really just an indication of what is needed.

CU9-DOC-S-FUN-120	5.2	MAN	Draft

CU9 shall provide scientific plots derived from the archive contents. These shall include as a minimum the following list.

- Hertzsprung Russell Diagrams of selectable subsets
- galactic-kinematics diagrams of selectable subsets
- Hess diagrams and colour magnitude diagrams of selectable subsets
- Plots for multiple stars and for non-stellar sources

Where possible/sensible both a print quality copy and an interactive plot should be made available e.g. to take one to the subset of data represented in part of the plot clicked on.

CU9-DOC-S-PLN-020	5.0	MAN	Draft

CU9 shall produce a printed volume (or volumes) containing the algorithms documentation, processing assumptions and project history.

CU9-DOC-S-PLN-040	5.0	MAN	Draft

If suitable media exist a subset of the Gaia Catalogue should be distributed with the printed volume (see Req. CU9-DOC-S-PLN-020).

Perhaps one or two blu-ray equivalents would allow the astrometry to be provided in this manner (Req. CU9-DOC-S-PLN-040. In 2020 there will surely be interesting devices about. This may be just the sources not all transits for example.

CU9-DOC-S-PLN-060	5.0	MAN	Draft

All technical design/installation/maintenance documentation shall be produced for the archive by CU9.

8 Science Alerts

Flux-based alerts need to be delivered quickly to the community and are currently built in to the CU5 processing. These flux-based alerts will be distributed through the normal IAU channels. It is felt that introducing a new layer on top of these alerts would not benefit Gaia. However anomaly based alerts, or announcements, which are less time critical, coming from other CUs

could be distributed through CU9 (e.g. CU7 variability announcements based on the light curve analysis).

8.1 Requirements

CU9-SAA-S-FUN-020	4.1	FUN	Draft
CU9 shall provide a mechanism for other CUs to publish anomaly based science alerts.			

CU9-SAA-S-FUN-040	4.1	FUN	Draft
The publishing mechanism shall only be accessible to DPAC members and shall use the Gaia Portal LDAP system for authentication.			

Clearly we do not want just anyone to be allowed to publish alerts - we also do not want a a new system of passwords and logins for this. DPAC already have an authentication mechanism based on the portal which is used by the MDB dictionary and other tools.

CU9-SAA-S-FUN-060	4.2	FUN	Draft
CU9 shall archive all alerts raised in DPAC. The archive entry should include pertinent information such as the source id used in the alert.			

By all alerts this includes any alerts raised by CU5 independently.

CU9-SAA-S-FUN-080	4.1	FUN	Draft
CU9 shall use appropriate IAU, VO and other channels to distribute science alerts raised through CU9.			

CU9-SAA-S-FUN-100	6.0	MAN	Draft
CU9 shall monitor the results of science alerts and record their success/failure in the archive.			

This is to say that CU9 should record the success rate of the alerts whether there was something or not.

CU5 will probably raise science alerts directly as it is most efficient.

CU9-SAA-S-FUN-120	6.1	MAN	Draft
Any CU publishing flux based alerts directly shall agree with DPACE and GST the mechanism of publication.			

CU9-SAA-S-FUN-140	6.0	MAN	Draft
All flux-based alerts shall be archived by CU9 and shall be available to the community as part of all CU9 products.			

9 Help Desk

It is traditional to have a manned help desk to answer questions on the mission archive.

9.1 Requirements

CU9-HLP-M-PLN-020	5.2	MAN	Draft
CU9 shall operate a manned help desk during normal working hours after the first release of the Gaia Archive until one year after the final catalogue release.			

CU9-HLP-M-FUN-020	5.0	MAN	Draft
CU9 shall maintain an electronic ticket/tracking system for all questions/issues raised to it. Currently DPAC use Mantis but CU9 may choose another system for the public interface.			

CU9-HLP-M-FUN-040	5.0	MAN	Draft
CU9 shall maintain a list of Frequently Asked Questions (FAQ) which should be regularly updated and augmented.			

10 Public Outreach (PBO)

CU9 shall provide the ”face” of the Gaia project and its products not only to the interested science community, but also to the (interested section) of the general public.

At the DPACE-07 meeting Mark Cropper took an even more radical view: He proposed that the Gaia archive portal should be a *public* portal with an astronomers’ section, rather than vice versa. This should be considered by CU9.

10.1 Requirements

CU9-PBO-S-FUN-020	5.0	FUN	Draft
The web-based presentation of the Gaia archive shall be public-outreach oriented.			

CU9-PBO-S-FUN-040	5.0	FUN	Draft
The public-outreach component shall have a ”questions to the experts” section (interactive, with a real person answering — reasonable and serious — questions), a ”frequently asked questions” section (archival only), a kids- and youth-oriented section.			

CU9-PBO-S-FUN-060	5.0	FUN	Draft
CU9 shall consider whether scientific datamining projects ”Gaia@home”, in the manner of ”SETI@home”, the SDSS ”galaxy zoo” classification project and the aerogel stardust trail search project, can be defined.			

CU9-PBO-S-FUN-080	5.0	FUN	Draft
The public-outreach component shall include a discussion forum for direct contact between the general-public users of the archive. This forum will not provide professional interaction/advice/replies to questions. The forum will, however, be regularly monitored by a CU9 member for non-topical, offensive, commercial, ... entries or discussion threads. Such entries or discussion threads will be removed as quickly as possible.			

CU9-PBO-S-FUN-100	5.0	FUN	Draft
CU9 shall elaborate educational material as a set of exercises and guidelines for teachers to illustrate astronomical and astrophysical concepts/problems using Gaia data.			

CU9-PBO-S-FUN-120	6.0	FUN	Draft
The public material shall be *localizable*. It should be easy to add new language versions of pages. This could be done using some form of content management system - a decision for CU9. (see also Req. CU9-CIF-T-FUN-160			

CU9-PBO-S-FUN-140	6.0	FUN	Draft
An approporiate URL shall be registered for the Gaia archive website. (see also URI specification `http://www.w3.org/TR/webarch/`)			