CU9 Science Archive Architecture and Development Workpackage Software Requirements Specification (WP930)

prepared by: N.C. Hambly
approved by: CU9 System Engineering Group
reference: GAIA-C9-SP-IFA-NCH-031
issue:
revision: 2
date: 2014-11-11
status: Draft

Abstract

This document provides the list of requirements applicable to the Gaia science archive architecture and development (WP930)

Document History

Issue	Revision	Date	Author	Comment
D	3	2014-11-11	NCH	Added GENIUS acknowledgement
D	2	2014-10-28	NCH	Comments from Harry; also included GC-001 as an applicable doc
D	1	2014-08-27	NCH	First complete draft, but could do with more detailed requirements
D	0	2014-08-26	NCH	First version of the CU9 archive system SRS

Contents

Acronym List

The following table has been generated from the on-line Gaia acronym list:

Acronym	Description
ADQL	Astronomical Data Query Language
AIP	Astrophysikalisches Institut Potsdam
ASDC	ASI Science Data Centre
AUT	AUTomated
CSV	Comma-Separated Value (database output format, e.g., for MS Excel)
CU	Coordination Unit (in DPAC)
DPAC	Data Processing and Analysis Consortium
ESACSAT	European Space Astronomy Centre Science Archives Team
FITS	Flexible Image Transport System
GACS	Gaia Archive Core Systems
GENIUS	Gaia European Network for Improved User Services
GPDB	Gaia Parameter DataBase
MAN	MANual
SRS	Software Requirements Specification
TAP	Table Access Protocol
TOPCAT	Tool for OPerations on Catalogues And Tables
VO	Virtual Observatory
WAN	Wide Area Network
WP	Work Package
XML	eXtensible Markup Language

1 Introduction

This document sets out the software requirements pertaining to the Gaia science archive architecture and development.

Note that the document is in part fulfilment of the GENIUS project milestone MS6 for Work Package 3.

1.1 Objectives

The objective of this document is to define a set of requirements that can either be explicitly tested for, or else at least verified, for the science archive development process. By setting and testing these requirements we aim at producing software subsystems that are reliable and robust both in terms of algorithmic and software performance.

1.2 Scope

This document covers all aspects of the science archive architecture, from all algorithmic aspects to implementation. The workpackage developments and implementation work are covered by CU9 WP930.

1.3 Assumptions

Some requirements have higher level CU9 requirements. These dependencies are recorded but not repeated in this document; the relevant SRS (see next Section) should be referred to for further details. The top–level requirements specification should be considered as 'level 0' requirements that are unlikely to change during the development and implementation iterations, while the derived and more detailed requirements herein may subject to significant revision consistent with a pragmatic and 'agile' development process.

1.4 Applicable Documents

When applicable documents change a change may be required in this document. The applicable documents are listed here for clarity; a full reference list is provided at the end of the document.

(WOM-086)	Software Development Plan for CU9
(WOM-033)	Gaia Catalogue and Archive SRS
(AB-026)	Gaia data access scenarios summary
(GC-001)	Gaia Archive Requirement Analysis

For convenience only the structure of this document follows that of the level 0 SRS (WOM-033).

1.5 Requirement Definition

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

CU9-*WP93x***-***X***-***SCOPE***-***xxx***

Where :

- *WP93x* is the (sub–)WP number as follows:
  - WP931: Management
  - WP932: Gaia Archive Core Systems
  - WP933: Consortium/ESAC–SAT Interface Control
  - WP934: Database Collaboration
  - WP935: Database Interface Design
  - WP936: Correlation Functions (for error propagation)
- *X* is either `S` (for **S**cientific), `T` (for **T**echnical), `Q` (for **Q**uality Assurance), or `M` (for **M**anagement)
- *SCOPE* is a four letter scope specification of the requirement following the identified list of possible values as shown in the list below
- *xxx* is a monotonically increasing counter for every unique combination

SCOPE is a 4-letter scope specification of the requirement. In this document the following scopes have been used:

GLOB: for top-level global requirements;

ALGO: for requirements on the scientific algorithm to be applied in the data processing (this should be for the detail on how the functionality is to be achieved);

CODE: for requirements on the coding activities (this should just be for how the code is written e.g. the GPDB should be used);

COOR: for requirements on the coordination activities (this should not be used for functionality requested by other WPs, but that you coordinate with them in the design etc.);

DATA: for requirements on data stream handling (this can include descriptions of input and output data);

FUNC: concerning functional requirements (this describes the functionality that is required by the system);

HARD: concerning hardware requirements;

PERF: for requirements on performances;

PLAN: for requirements on the planning activities;

QUAL: for requirements on the quality assurance (both scientific and software, robustness and quality of data could go here);

RESO: for requirements on the resource management.

DOCU: for requirements on documentation.

Each requirement is presented with its own unique label and a number of attributes in the following form:

*CU9-WP93x-X-SCOPE-**020***	$C.v$	Verification	Status
Description			
Parent: Parent			

CU9-WP93x-X-SCOPE-xxx	The unique identifier of the requirement (see above).
$C.v$	Version number of the requirement composed of major part (C) corresponding to the cycle (1, 2 and 3 corresponding to A, B and C respectively in WOM-086) in which the requirement was created and minor part (v) corresponding to the version of the requirement.
Verification	Envisaged validation method of requirement - this will be either AUT for automated or MAN for Manual.
Status	Status identifier.
Parent	Higher level requirement or requirements in a comma separated list.

2 List of requirements

2.1 General requirements

CU9-WP931-M-PLAN-020	1.1	MAN	Draft
CU9 WP930 shall operate within the agreed management structures of the CU (which in turn operates within the existing DPAC management structure).			
Parent: CU9-ARC-M-PLN-020			

CU9-WP931-M-PLAN-040	1.1	MAN	Draft
CU9 WP930 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.			
Parent: CU9-ARC-M-PLN-040			

CU9-WP931-M-PLAN-060	1.1	MAN	Draft
CU9 WP930 shall follow the engineering guidelines laid down in WOM-086.			
Parent: CU9-ARC-M-PLN-060			

CU9-WP931-M-PLAN-080	1.1	MAN	Draft
CU9 WP930 developments shall be overseen by a System Engineering coordination group.			
Parent: CU9-ARC-M-PLN-060			

CU9-WP931-M-COOR-020	1.1	MAN	Draft
CU9 WP930 shall coordinate with other CU9 WPs in defining applications interfaces of the archive system (e.g. for validation and visualisation).			
Parent:			

2.2 Community Interface

CU9-WP932-T-FUNC-020	1.1	MAN	Draft
ESAC–SAT shall design, develop, deploy and maintain a Gaia Archive Core Systems (GACS) interface for Gaia catalogue products, specification as per the parent level 0 requirements.			
Parent: CU9-CIF-T-MAN-020			

CU9-WP933-M-PLAN-020	1.1	MAN	Draft
DPAC CU9 WP930 shall work with ESAC–SAT in developing functionality over and above the GACS as required by usage scenarios identified within AB-026.			
Parent: CU9-CIF-T-MAN-020			

CU9-WP933-M-PLAN-040	1.1	MAN	Draft
CU9 WP930 shall establish interface coordination agreements with ESAC–SAT in order to develop contributions to archive interface functionality.			
Parent: CU9-WP933-M-PLAN-020			

2.3 Archive Ingestor

CU9-WP933-T-COOR-020	1.1	MAN	Draft
CU9 WP930 shall employ the Main Database Dictionary Tool (MTL-003; hereafter MDB DT) to define the detailed specification of the Gaia data to be published through the archive system.			
Parent: CU9-ING-T-FUN-020			

CU9-WP934-T-COOR-020	1.1	MAN	Draft
A data model for the core Gaia catalogues shall be defined in the MDB DT			
Parent: CU9-WP933-T-COOR-020			

CU9-WP934-T-COOR-040	1.1	MAN	Draft
A data model for crossmatch applications shall be defined in the MDB DT, including definitions for external catalogue holdings.			
Parent: CU9-WP933-T-COOR-020			

2.4 Interrogator

CU9-WP935-T-FUNC-020	1.1	MAN	Draft
GACS shall provide interactive and automated (via TAP) ADQL access to tabular datasets. *Further specific requirements for interactive access are included in Section 2.2.*			
Parent: CU9-ITG-T-FUN-020,CU9-ITG-T-FUN-140			

CU9-WP935-T-FUNC-040	1.1	MAN	Draft
The archive system shall present provenance information for derived catalogue attributes.			
Parent: CU9-ITG-T-FUN-040,CU9-ITG-T-FUN-060			

CU9-WP935-T-FUNC-060	1.1	MAN	Draft
The GACS shall provide ‘MyDB’–like local space for user’s transient intermediate results and/or the upload of private data.			
Parent: CU9-ITG-T-FUN-100,CU9-ITG-T-FUN-120			

CU9-WP935-T-FUNC-080	1.1	MAN	Draft
GACS shall allow at a minimum CSV, FITS, and VOTable (XML) output and upload data formats. *For output formats of the interactive system, this requirement may be fulfilled via a 'send to SAMP' button on the results page, which allows single–click access to richly functioned applications like TOPCAT which has options to save in multiple data formats including those mentioned above.*			
Parent: CU9-ITG-T-FUN-160,CU9-ITG-T-FUN-180			

2.5 Advanced Applications

CU9-WP935-T-FUNC-100	1.1	MAN	Draft
The archive system shall provide ADQL cross–querying of local catalogue data holdings.			
Parent: CU9-ADV-T-FUN-080			

CU9-WP935-T-FUNC-120	1.1	MAN	Draft
As well as providing ADQL cross–querying of local catalogue data holdings, the archive system shall provide a more advanced Distributed Query Processing facility for cross–querying of catalogues published to the VO and distributed over the WAN.			
Parent: CU9-ADV-T-FUN-080			

CU9-WP935-T-FUNC-140	1.1	MAN	Draft
CU9 WP930 shall develop client–side advanced applications for use with Gaia data, for example by adapting existing VO–based tools specifically for Gaia usage scenarios.			
Parent: CU9-ADV-T-FUN-080			

CU9-WP935-T-FUNC-160	1.1	MAN	Draft
CU9 WP930 shall develop a data mining Application Programming Interface for the archive system. *This is needed as a fundamental component of the execution environment for data mining applications running co-located with the archive system. This will need further requirements to be specified in a later issue of this document.*			
Parent: CU9-ADV-T-FUN-200			

CU9-WP935-T-FUNC-180	1.1	MAN	Draft
CU9 WP930 shall develop the means to mirror the core archive catalogues at other major European Data Centres (e.g. CDS, AIP, ASDC).			
Parent: CU9-ARC-M-FUN-020			

CU9-WP936-T-FUNC-020	1.1	MAN	Draft
The archive system shall provide the means for full propagation of errors, including error correlations, to end–user selections.			
Parent: CU9-ITG-T-FUN-060			

2.6 Documentation

CU9-WP935-T-DOCU-020	1.1	MAN	Draft
The archive system shall propagate all detailed documentation pertaining to catalogue attributes that is encapsulated within the MDB DT definitions.			
Parent: CU9-DOC-S-FUN-040,CU9-DOC-S-PLN-060,CU9-WP934-T-COOR-020			

2.7 Science Alerts

See elsewhere for the SRS of the relevant workpackage (WP975)

2.8 Help Desk

See elsewhere for the SRS of the relevant workpackage (WP953)

2.9 Public Outreach

See elsewhere for the SRS of the relevant workpackage (WP960)

Acknowledgements

We acknowledge European Commission funding under the auspices of the FP7 GENIUS programme.

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/cs/livelink/open/3125400

[GC-001], Costigan, G., 2014, *Gaia Archive Requirement Analysis*,
GAIA-C9-TN-LEI-GC-001,
URL http://www.rssd.esa.int/cs/livelink/open/3282045

[WOM-033], O'Mullane, W., 2009, *Gaia Catalogue and Archive Software Requirements and Specification*,
GAIA-C9-SD-ESAC-WOM-033,
URL http://www.rssd.esa.int/cs/livelink/open/2907710

[WOM-086], O'Mullane, W., Luri, X., Gracia, G., 2014, *CU9 Software Development Plan*,
GAIA-C9-PL-ESAC-WOM-086,
URL http://www.rssd.esa.int/cs/livelink/open/3237698

[MTL-003], ter Linden, M., Hernandez, J., Hutton, A., 2013, *Main Database Dictionary Tool Software User Manual*,
GAIA-C1-UG-ESAC-MTL-003,
URL http://www.rssd.esa.int/cs/livelink/open/2848107

A Requirements traceability

The following table provides traceability for derived requirements within this requirements specification, and also to level 0 requirements in WOM-033.

Parent Requirement	Requirements in this document
CU9-ADV-T-FUN-080	CU9-WP935-T-FUNC-100, CU9-WP935-T-FUNC-120, CU9-WP935-T-FUNC-140
CU9-ADV-T-FUN-200	CU9-WP935-T-FUNC-160
CU9-ARC-M-FUN-020	CU9-WP935-T-FUNC-180
CU9-ARC-M-PLN-020	CU9-WP931-M-PLAN-020
CU9-ARC-M-PLN-040	CU9-WP931-M-PLAN-040
CU9-ARC-M-PLN-060	CU9-WP931-M-PLAN-060, CU9-WP931-M-PLAN-080
CU9-CIF-T-MAN-020	CU9-WP932-T-FUNC-020, CU9-WP933-M-PLAN-020
CU9-DOC-S-FUN-040	CU9-WP935-T-DOCU-020
CU9-DOC-S-PLN-060	CU9-WP935-T-DOCU-020
CU9-ING-T-FUN-020	CU9-WP933-T-COOR-020
CU9-ITG-T-FUN-020	CU9-WP935-T-FUNC-020
CU9-ITG-T-FUN-040	CU9-WP935-T-FUNC-040
CU9-ITG-T-FUN-060	CU9-WP935-T-FUNC-040, CU9-WP936-T-FUNC-020
CU9-ITG-T-FUN-100	CU9-WP935-T-FUNC-060
CU9-ITG-T-FUN-120	CU9-WP935-T-FUNC-060
CU9-ITG-T-FUN-140	CU9-WP935-T-FUNC-020
CU9-ITG-T-FUN-160	CU9-WP935-T-FUNC-080
CU9-ITG-T-FUN-180	CU9-WP935-T-FUNC-080
CU9-WP933-M-PLAN-020	CU9-WP933-M-PLAN-040
CU9-WP934-T-COOR-020	CU9-WP935-T-DOCU-020
CU9-WP933-T-COOR-020	CU9-WP934-T-COOR-020, CU9-WP934-T-COOR-040