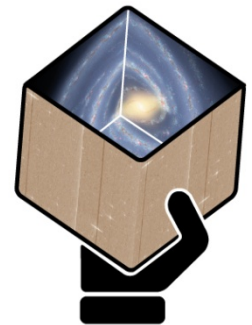


GENIUS status & wrap-up

WP3: aspects of archive system design



gaia



Nigel Hambly

Institute for Astronomy, Edinburgh University



WP3 Description

- From the GENIUS proposal:

“The objective of this work package is to design, prototype and develop aspects of the **archive infrastructure** needed for the scientific exploitation of Gaia data. The design and technology choices made will be motivated by the real user **requirements identified by WP2** – in particular, the massive, complex queries defined by the Grand Challenges – and by other initiatives, such as the GREAT project, and will be made with full recognition of the **constraints imposed by the ESAC archive** system, with which it **must interface effectively**. Prototypes will be prepared and tested in cooperation with the end user community and with the ESAC science archive team through the DPAC CU9. A **core principle** will be the **adoption of Virtual Observatory standards** and the **development of VO infrastructure** to enable ready interoperation with the other external datasets needed to release the full scientific potential of Gaia.”

GENIUS WP3 in DPAC CU9 WP930

WP930: Science Archive Architecture and Development [\[edit\]](#)

Staff [\[edit\]](#)

The following staff are currently active in WP930 at the full-time-equivalent fractional level indicated in brackets:

- Nigel Hambly [\[1\]](#), IfA Edinburgh University (manager; 0.2 FTE funded by GENIUS + 0.25 = 0.45 FTE to end Mar 2017; 0.25 thereafter)
 - Juan Gonzalez, ESAC SAT (deputy manager and ESDC main contact; 0.5)
 - Raul Guitierrez [\[2\]](#), ESDC (1.0)
 - Juan Carlos Segovia [\[3\]](#), ESDC (1.0)
 - Gabriele Comoretto (<0.1)
 - Harry Enke [\[4\]](#), AIP eScience Team (0.25)
 - Gal Matijevic [\[5\]](#), AIP eScience Team (0.75 from 1. Dec. 2015)
 - Paola Marrese [\[6\]](#), INAF-ASDC (0.2)
- Riccardo Smareglia [\[7\]](#), INAF-OATs (0.15; 0.15 FTE funded via GENIUS from Jan 2014)
- I-Chun Shih (Stephen) [\[8\]](#), Observatoire Paris Meudon (0.1; works mainly in WP940)
- Stelios Voutsinas [\[9\]](#), IfA Edinburgh University (0.2 FTE from October 2013 funded by GENIUS)
- Dave Morris [\[10\]](#), IfA Edinburgh University (0.2 FTE from October 2013 rising to 0.65 FTE April 2014 and down to 0.5 FTE from October 2014 funded by GENIUS)
 - Paul McMillan [\[11\]](#), Lund University, 0.5 FTE from Q4 2014 (WP936)
- Robert Butora [\[12\]](#), INAF-OATs (0.5 FTE funded via GENIUS from June 2014)

=> 1.8 (out of total of 6.1) FTE to end of March 2017

Work Breakdown Structure

- **T3.1: Technical Co-ordination**
 - System Requirements Specification
 - Systems Interface Control: ICDs
- **T3.2: Aspects of Archive (end-user) Interface Design**
 - Subsystems interface infrastructure (affecting end-user experience)
 - Enhanced features for User Interfaces
- **T3.3: VO Infrastructure**
 - Client-side “Table Access Protocol” (TAP) tool
 - International Virtual Observatory Alliance (IVOA) work
- **T3.4: Data Centre Collaboration**
 - Distributed Query Processing (DQP) infrastructure
- **T3.5: Cloud-based Research and Data Mining Environments**
 - Virtual Machines and containerisation



gaia



T3.1: Technical Co-ordination

- **Good communication channels**
 - On-line collaborative tools (Wiki, SVN, teleconferencing)
- **Requirements Specification**
 - SRS documented (Milestone MS6)
- **Formal interface control established**
 - e.g. DQP ICD (Deliverable D3.1)

⇒ and ensures good coordination between the 20 individuals involved in DPAC CU9 WP930 (presently a total of 6.1 FTE with 2.5 FTE at ESDC, 1.8 FTE national funding agencies and 1.8 FTE currently resourced via GENIUS in WP3)

T3.1 (cont.): Integration with DPAC



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



DPAC CU9 WP930 / ESAC Gaia Archive
Core Systems Interface Control Document:
FireThorn distributed querying infrastructure

prepared by: D. Morris, N.C. Hambly and S. Voutsinas
affiliation : University of Edinburgh
approved by: ESAC SAT & CU9 CCB
reference: GAIA-C9-TN-IFA-DM-001-0D
issue: 0D
revision: 3
date: 2014-11-11
status: Draft

GENIUS
Deliverable D3.1



GENIUS status & wrap-up January 2017

T3.2: Aspects of Archive Interface Design

- End-user experience depends on propagation of relevant information through the system
- Primary mechanism for interface specification within DPAC is a data model “Dictionary Tool”
- Some key infrastructural features missing from the GENIUS/CU9 perspective
 - Concentrated on these before proceeding any contributions to the User Interface itself
- Gaia archive UI enhancements

T3.2 (cont.): Dictionary Tool enhancements

- New metadata fields
- Additional propagation features
 - Ensure UIs contain all necessary information
- New version of tool released Q2 2016
 - In good time for DR1 developments

The screenshot shows the Gaia Main Database Dictionary Tool interface. The title bar reads "Gaia Main Database Dictionary Tool". The menu bar includes "File", "Table", "Session", and "Help". The left sidebar shows a tree view of the database structure: MDB DM, MDB, CU1, CU2, CU3, CU4, CU5, CU6, CU7, CU8, CU9, ArchiveArchitecture, DatabaseCollaboration, CrossMatch, BaseNeighbourhood, and ExampleCrossNeighbours. The main window displays the "Table Description" for "MDB/CU9/ArchiveArchitecture/DatabaseCollaboration/CrossMatch/BaseNeighbourhood". The "Extends:" field is empty. The "Table Consumers:" field is empty. The "Table is abstract" checkbox is checked. The "Interfaces" button is visible. A table with the following columns is displayed: #, Name, Description, Det. Desc., Type, Multiplicity, Units, Minimum, Maximum, Publish, UCD1+, Utype, and Default value. The table contains two rows: 1 masterSourceId (UID of source at neighbourhood centre, View, long, Dimensionless[see description]) and 2 distance (Angular distance between neighbouring sources, View, float, Angle[arcsec]).

#	Name	Description	Det. Desc.	Type	Multiplicity	Units	Minimum	Maximum	Publish	UCD1+	Utype	Default value
1	masterSourceId	UID of source at neighbourhood centre	View	long		Dimensionless[see description]			<input type="checkbox"/>			
2	distance	Angular distance between neighbouring sources	View	float		Angle[arcsec]			<input type="checkbox"/>			

T3.2 (cont.): Gaia archive UI enhancements

- Auto-complete identified under beta-testing as a “nice-to-have” feature

The screenshot shows the Gaia archive (INTEGRATION) web interface. At the top, there is a navigation bar with links for HOME, SEARCH, STATISTICS, VISUALIZATION, HELP, DOCUMENTATION, VOSPACE, and SHARE. Below this, there are tabs for Simple Form, ADQL Form, and Query Results. The main content area features a search box with the text "Job name:" and a query input field containing "1 SELECT top 10 * from public.". An auto-complete dropdown menu is visible, listing various database sources such as "catalogue_source", "dual", "g10_fuzzy_1000", "g10_fuzzy_100000", "g10_fuzzy_1000000", "g10_gal", "g10_lmc", "g10_mw", "g10_quasars", and "g10_smc". A black box with white text "Type a query to be executed as a job" is overlaid on the dropdown. Below the query input, there is a table with the header "No results found" and columns for "Status", "Job", and "Creation date".

GENIUS
Deliverable D3.4

Will appear in next
release of GACS

T3.3: VO Infrastructure

- **VO-Dance, a client-side integration tool**
 - Allowing the end-user to publish to the VO
- **IVOA activities**
 - ADQL standards
 - ADQL parser enhancements
- **VO support**
 - Coordination of content descriptors (UCD1+)
 - ADQL Cookbook in support of GDR1
 - VO Compliance Document



gaia





GENIUS Deliverable D3.3

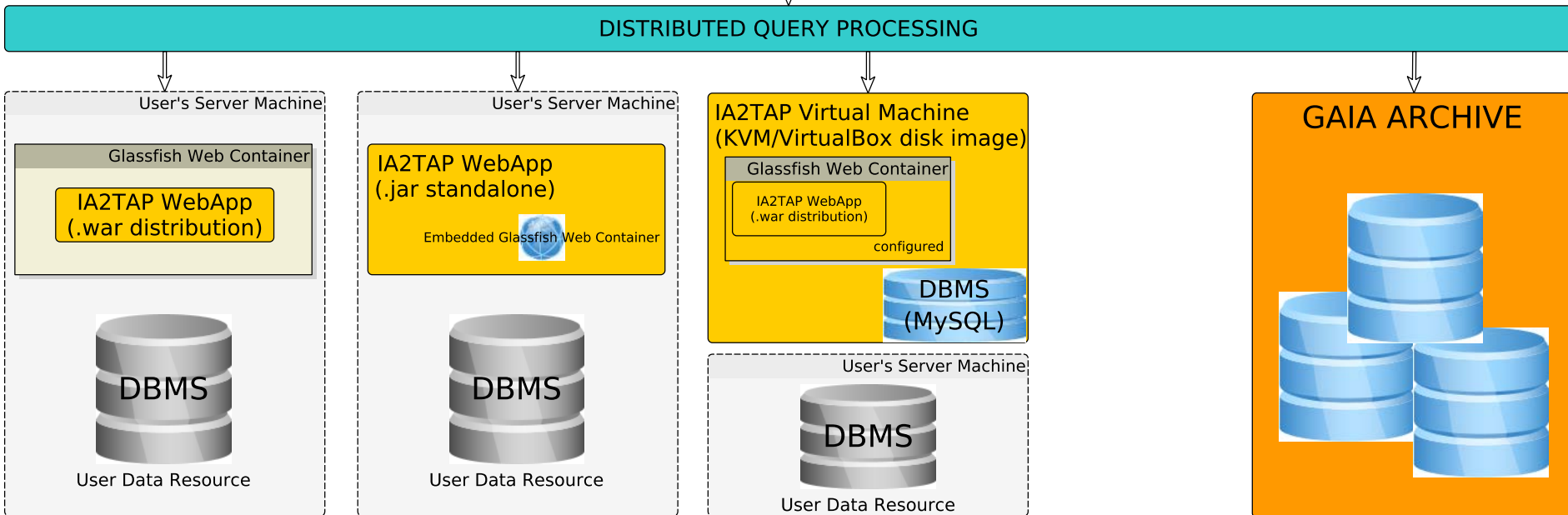
Gaia archive IVOA compliance document

prepared by: N.C. Hambly, D. Morris, S. Voutsinas and
M. Taylor
affiliation : Universities of Edinburgh and Bristol
approved by: GENIUS management board
reference: GAIA-C9-TN-IFA-NCH-040-D
issue: D
revision: 0
date: 2017-01-12
status: Draft

T3.3 (cont.): VO Infrastructure



DISTRIBUTED QUERY PROCESSING



GENIUS
Deliverable D3.4



selected schema

aao_local x aao_TS_test x + add new schema

Schema utype Schema description

selected table

AAO x afo x ech x sbi x + add new table

Table utype Table description

Columns + add new column

- x date_obs
- x dec_c selected column
- x dec_rad
- x detector
- x exp_id
- x exp_time
- x fit

dec_c Selected column data

Datatype: adql.VARCHAR

STD

UType:

UCD:

Unit:

e.g. TAP Schema editing via IA2TAP GUI application



gaia



T3.4: Data Centre Collaboration

- Archive test suite
- Recurring theme in requirements analysis is one of use of Gaia data in conjunction with other surveys:
 - Multiple wavelengths
 - Multiple epochs
 - Combination of primarily astrometric data with other surveys/missions

⇒ Distributed Query Processing is required



T3.4 (cont.): archive test suite



A test suite for Science Archive TAP/ADQL interfaces

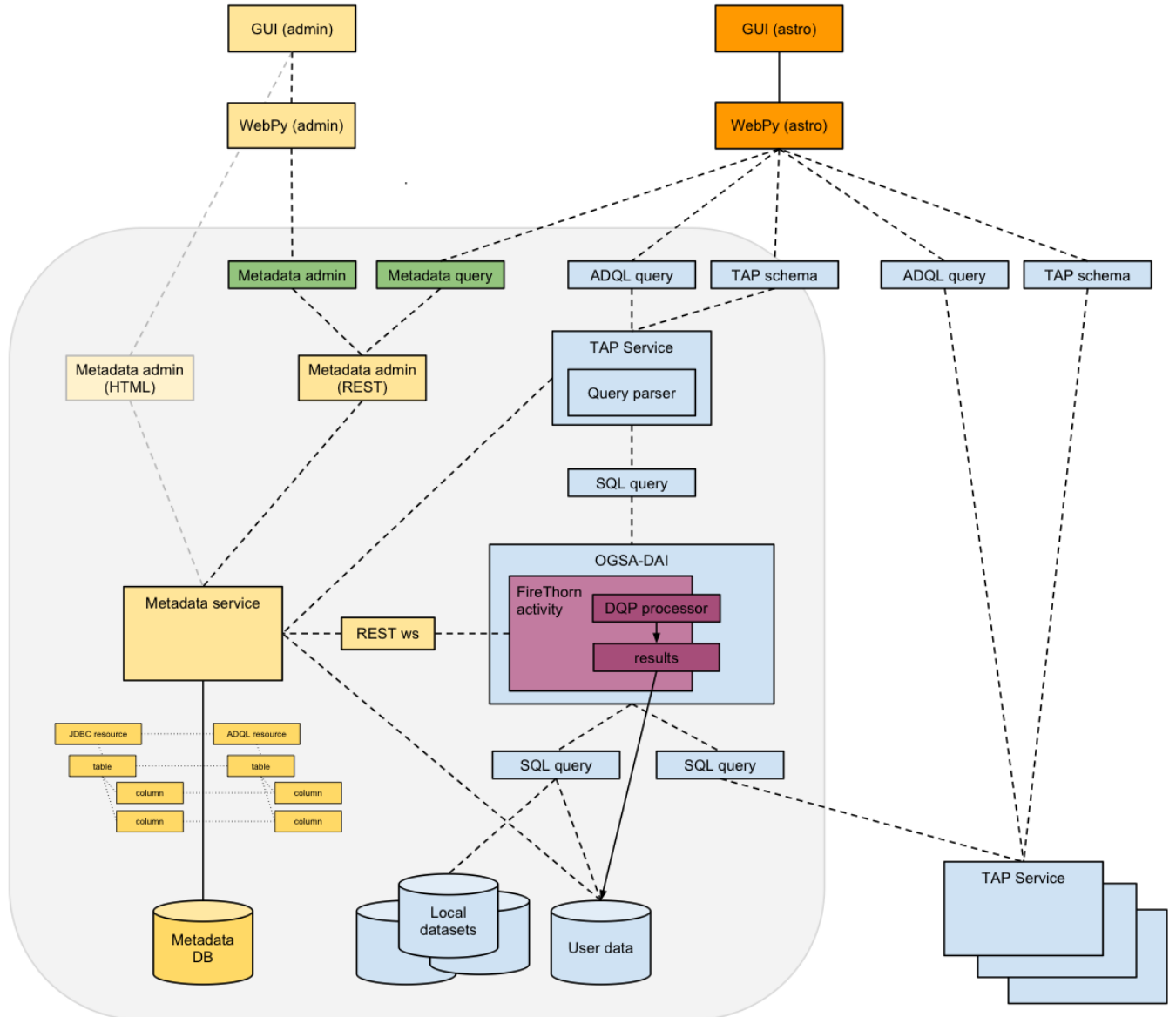
prepared by: S. Voutsinas, D. Morris and N.C. Hambly
affiliation : University of Edinburgh
approved by: GENIUS management board
reference: GAIA-C9-TN-IFA-STV-001-D
issue: D
revision: 0
date: 2016-11-21
status: Draft

- β -testing
- Stress-testing

GENIUS:

- Milestone MS9
- Deliverables D3.4 to 3.6

T3.4 (cont.): DQP under the hood



T3.4 (cont.): infrastructure demonstrator

- see <http://genius.roe.ac.uk>
 - DQP prototype

GENIUS
Deliverable D3.2

The screenshot shows the GENIUS Web2.0 demonstrator prototype website. The page has a dark grey background. On the left, there is a vertical navigation menu with a 'gaia GENIUS' logo at the top and four menu items: 'Home', 'Freeform SQL', 'Known Issues', and 'Contact us'. The 'Home' item is highlighted with a white bar. The main content area features a large red banner with the 'gaia' logo on the left and the 'esa' logo on the right. Below the banner, the title 'GENIUS Web2.0 demonstrator prototype' is displayed. Underneath the title, there is a paragraph of text: 'Gaia is an ambitious mission to chart a three-dimensional map of our Galaxy, the Milky Way, in the process revealing the composition, formation and evolution of the Galaxy. Gaia will provide unprecedented positional and radial velocity measurements with the accuracies needed to produce a stereoscopic and kinematic census of about one billion stars in our Galaxy and throughout the Local Group. This amounts to about 1 per cent of the Galactic stellar population.' At the bottom of the page, there is a footer with the European Union flag, navigation links for 'Home | Login | Freeform SQL | Contact', and the text 'FP7-SPACE-2013-1 Grant n. 606740.'

T3.5: Data mining environments

- **Docker**
 - Light-weight Virtual Machine
 - DQP has been containerised using Docker for deployment at ESAC
 - Demonstrates flexible and secure deployment of third-party code at a Data Centre
 - Has potential as a mechanism for containerisation of user code uploads

Use of Docker[☆] for deployment and testing of astronomy software

D. Morris, S. Voutsinas, N.C. Hambly and R.G. Mann

Institute for Astronomy, School of Physics and Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, EH9 3HJ, UK

Abstract

We describe preliminary investigations of using Docker for the deployment and testing of astronomy software. Docker is a relatively new containerisation technology that is developing rapidly and being adopted across a range of domains. It is based upon virtualisation at operating system level, which presents many advantages in comparison to the more traditional hardware virtualisation that underpins most cloud computing infrastructure today. A particular strength of Docker is its simple format for describing and managing software containers, which has benefits for software developers, system administrators and end-users.

We report on our experiences from two projects – a simple activity to demonstrate how Docker works, and a more elaborate set of services that demonstrates more of its capabilities and what they can achieve within an astronomical context – and include an account of how we got problems solved through interaction with Docker’s very active open source development community, which is currently the key to the most effective use of this rapidly-changing technology.

Keywords:

1. Introduction

In common with many sciences, survey astronomy has entered the era of “Big Data”, which changes the way that sky survey data centres must operate. For more than a decade, they have been following the mantra of ‘ship the results, not the

data’ (e.g. [Quinn et al., 2004](#), and other contributions within the same volume) and deploying “science archives” (e.g. [Hambly et al., 2008](#), and references therein), which provide users with functionality for filtering sky survey datasets on the server side, to reduce the volume of data to be downloaded to the users’ workstations for further analysis. Typically these science archives have been implemented in relational database management systems, and astronomers have become adept

[☆]<https://www.docker.com>

Email address: dmr_stv_nch_rgm@roe.ac.uk

(D. Morris, S. Voutsinas, N.C. Hambly and R.G. Mann)

Preprint submitted to Astronomy & Computing

January 13, 2017

GENIUS
Deliverable D3.7



GENIUS status & wrap-up January 2017

Deliverables: status summary

D3.1	GENIUS/ ESAC–SAT Coordination and Interface Control document	WP3	01/01/2014		The Gaia CU9 SDP has been adopted as the document to follow. Supplementary subsystem ICDs: <div style="border: 1px solid gray; padding: 5px; margin: 5px 0;"> Distributed Query Processing (GAIA-C9-TN-IFA-DM-001) Mirroring Applications Interface (GAIA-C9-SP-UB-XL-034) </div>	Submitted
D3.2	Web2.0 user interface demonstration prototype deployment	WP3	01/04/2015		Web2.0 Demonstrator Deliverable32_Web_User_Interface_Demonstrator_Documentation.pdf	Submitted
D3.3	Gaia Data IVOA compliance document	WP3	01/04/2017		Final draft available on DPAC SVN (NCH-040)	
D3.4	Deployed web services, code and documentation	WP3	01/04/2017		Draft docs and prototype code available (UEDIN: DPAC SVN & GitHub; INAF on request)	
D3.5	Data centre report and analysis document	WP3	01/04/2017		Final draft available on DPAC SVN (STV-001) and code (GitHub)	
D3.6	TAP+ code and documentation	WP3	01/04/2017		Prototype DQP code on GitHub	
D3.7	Deployed CANFAR– style VM research environment and produced reports and documentation	WP3	01/04/2017		“Docker” paper submitted to A&C (Morris et al, Jan 2017); containerised DQP code etc. on GitHub	



gaia

