GENIUS
Final Advisory Board meeting
WP3:
aspects of archive system design

gaia

Gaia DPAC
Data Processing & Analysis Consortium

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

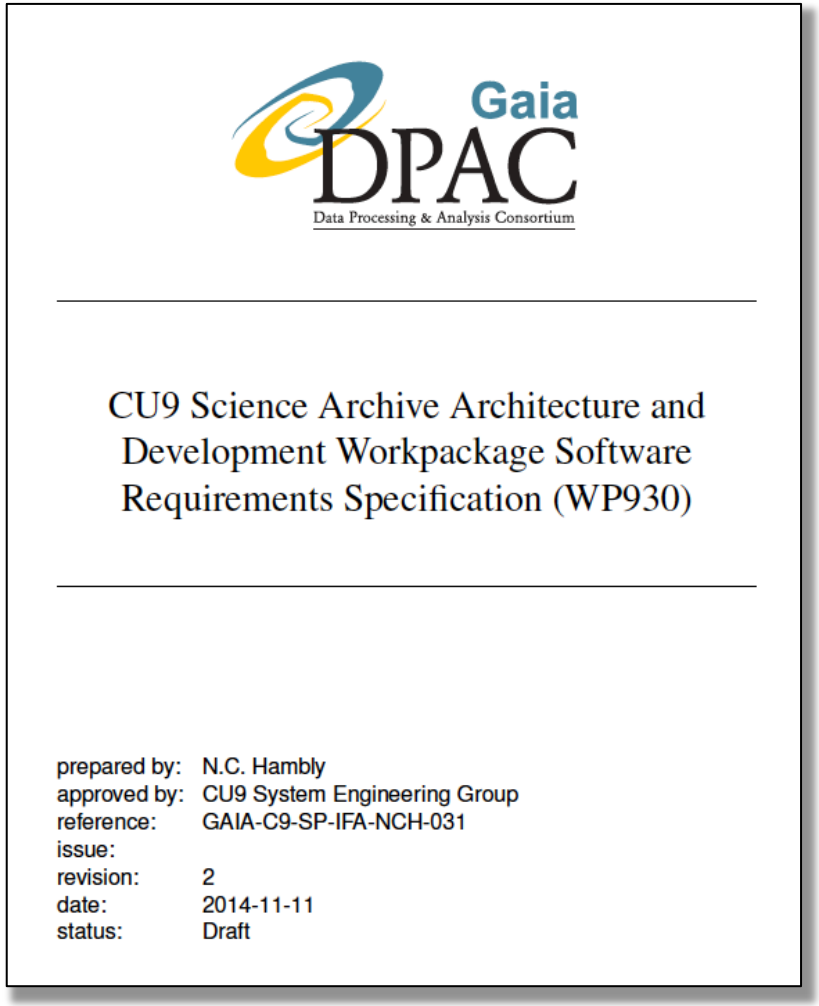
- T3.1: Technical Co-ordination
  - System Requirements Specification
  - Systems Interface Control: ICDs
- T3.2: Aspects of Archive (end-user) Interface Design
  - Subsystems interface infrastructure (affecting end-user experience)
  - Enhanced features for User Interfaces
- T3.3: VO Infrastructure
  - 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

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)

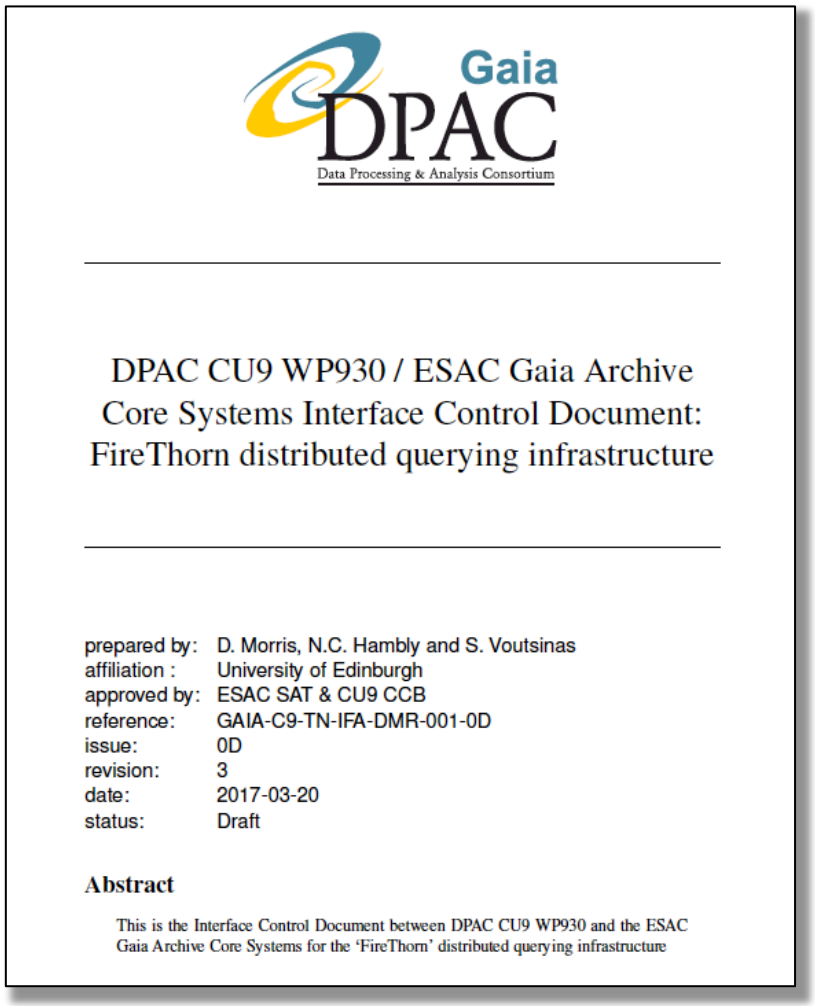
T3.1 (cont.): Integration with DPAC



Gaia DPAC
Data Processing & Analysis Consortium

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



Gaia DPAC
Data Processing & Analysis Consortium

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-DMR-001-0D
issue: 0D
revision: 3
date: 2017-03-20
status: Draft

Abstract

This is the Interface Control Document between DPAC CU9 WP930 and the ESAC Gaia Archive Core Systems for the 'FireThorn' distributed querying infrastructure

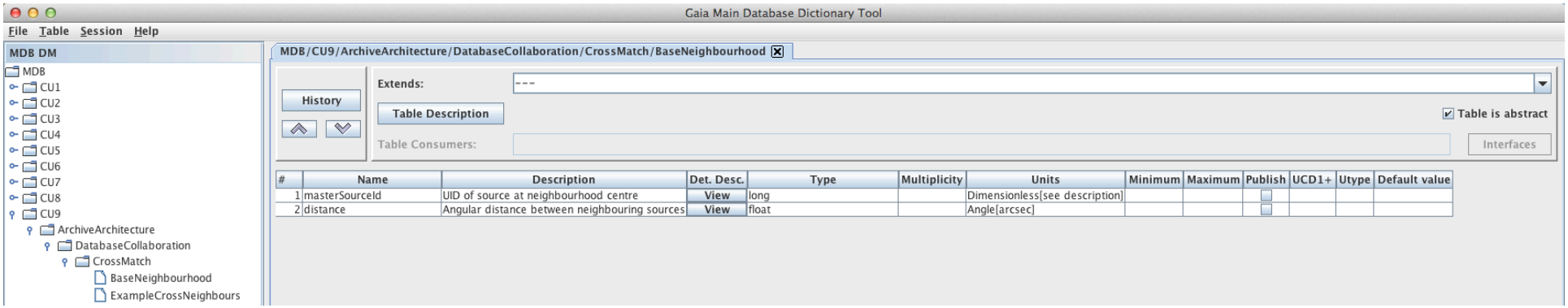
GENIUS
Deliverable D3.1

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

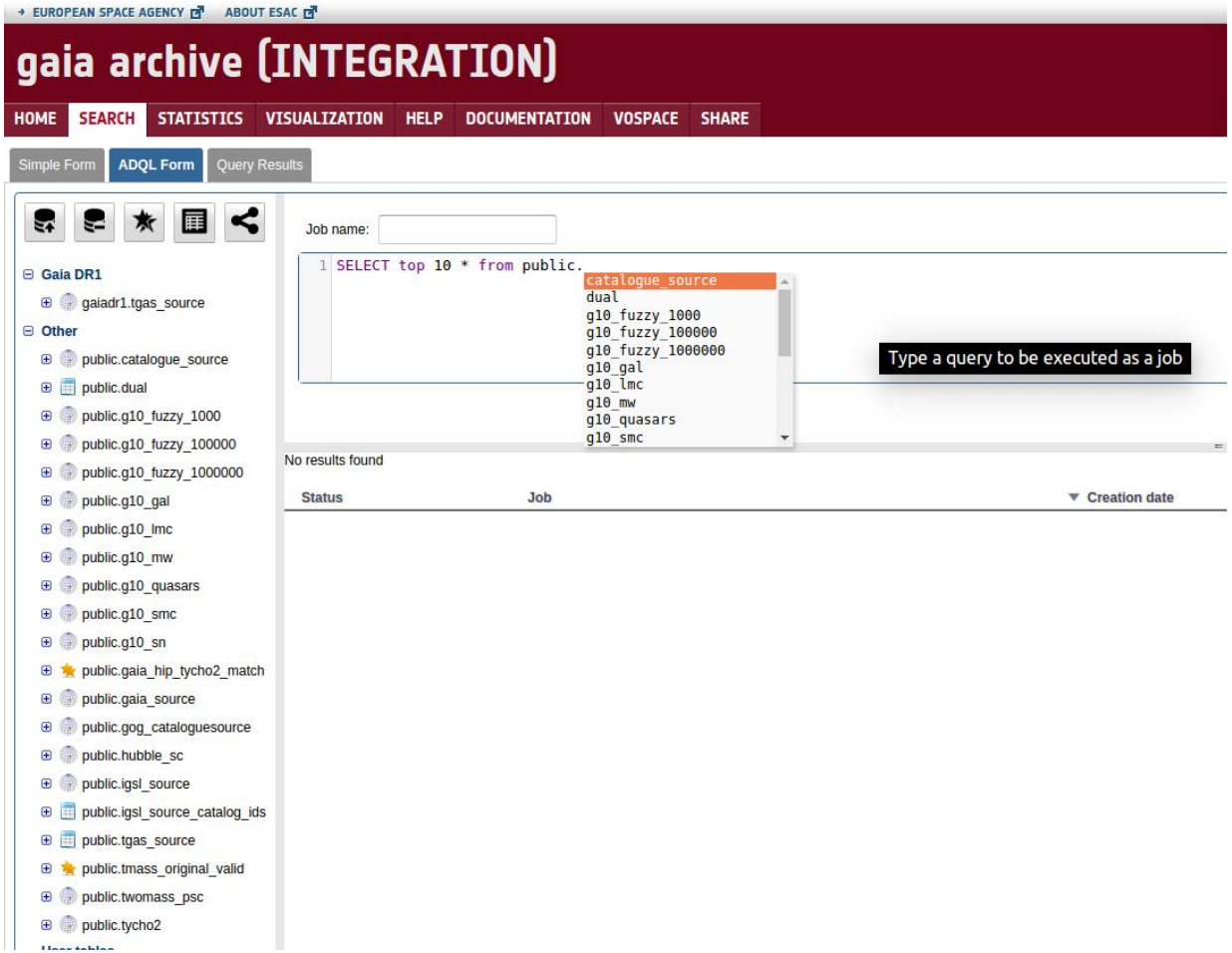
T3.2 (cont.): Dictionary Tool enhancements

- 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



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

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

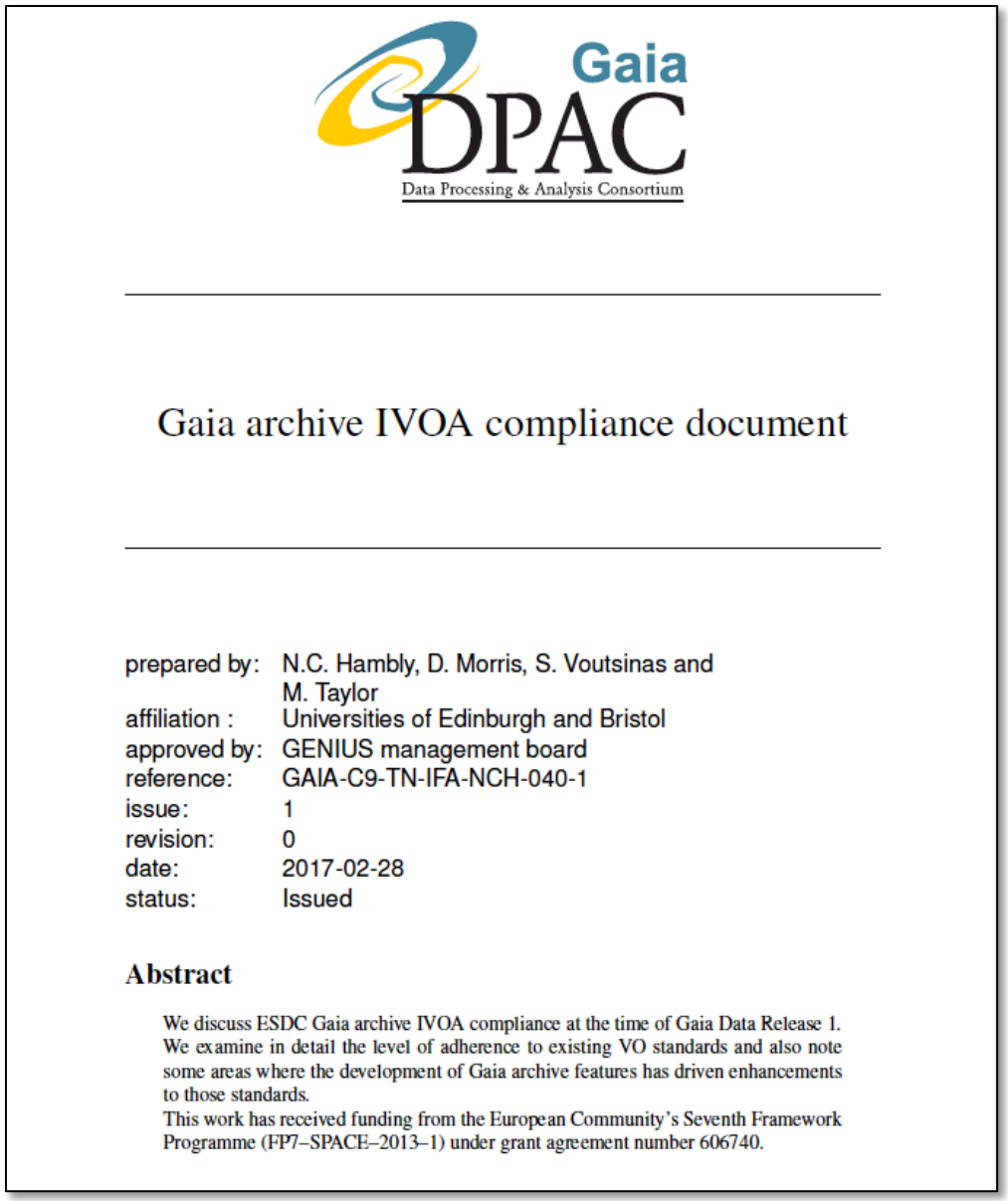


GENIUS
Deliverable D3.4

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
DPAC
Data Processing & Analysis Consortium

Gaia archive IVOA compliance document

prepared by: N.C. Hambly, D. Morris, S. Voutsinas and M. Taylor
affiliation : Universities of Edinburgh and Bristol

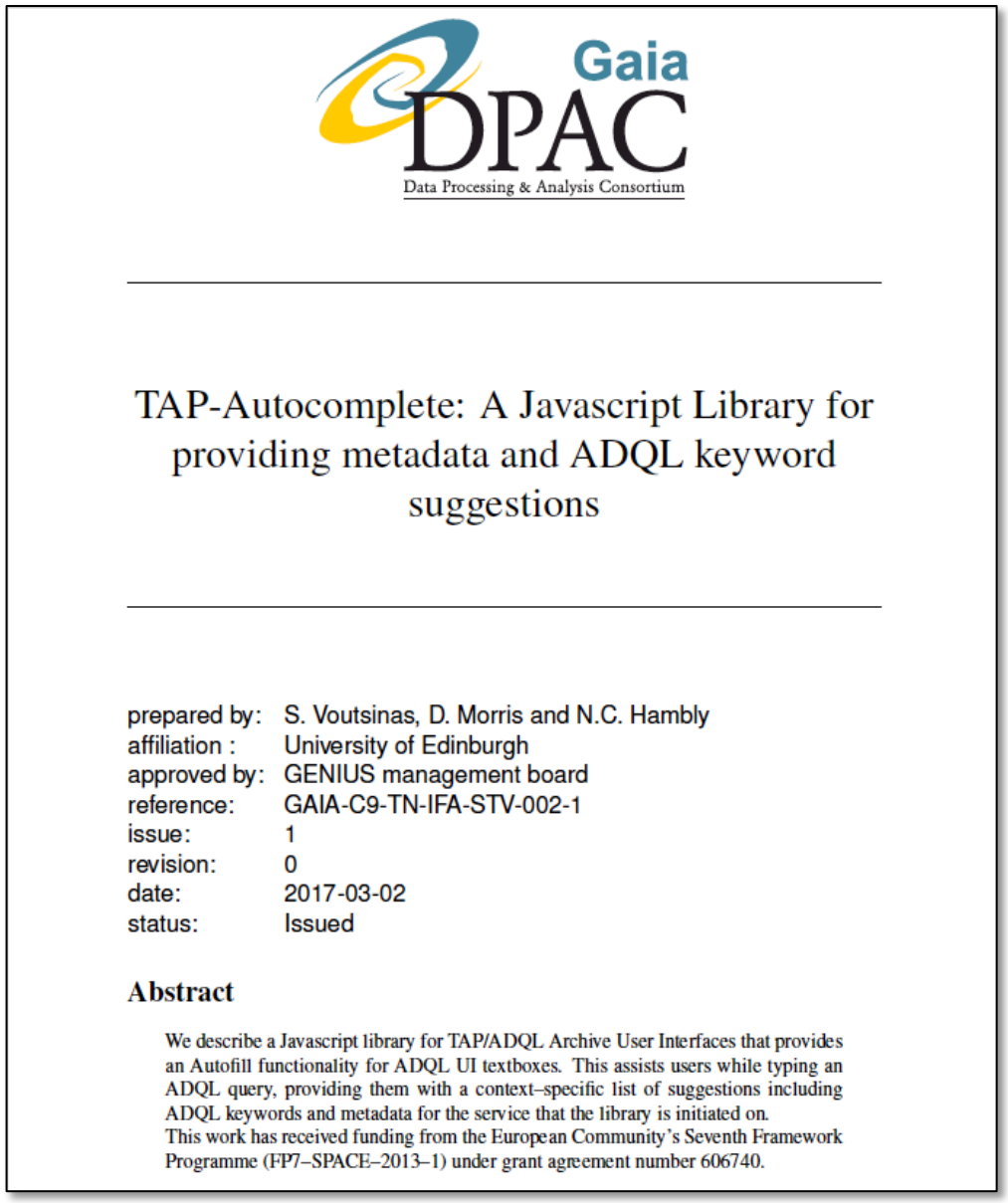
approved by: GENIUS management board
reference: GAIA-C9-TN-IFA-NCH-040-1
issue: 1
revision: 0
date: 2017-02-28
status: Issued

Abstract

We discuss ESDC Gaia archive IVOA compliance at the time of Gaia Data Release 1. We examine in detail the level of adherence to existing VO standards and also note some areas where the development of Gaia archive features has driven enhancements to those standards.

This work has received funding from the European Community's Seventh Framework Programme (FP7–SPACE–2013–1) under grant agreement number 606740.

GENIUS
Deliverable D3.3



Gaia DPAC
Data Processing & Analysis Consortium

TAP-Autocomplete: A Javascript Library for providing metadata and ADQL keyword suggestions

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-002-1
issue: 1
revision: 0
date: 2017-03-02
status: Issued

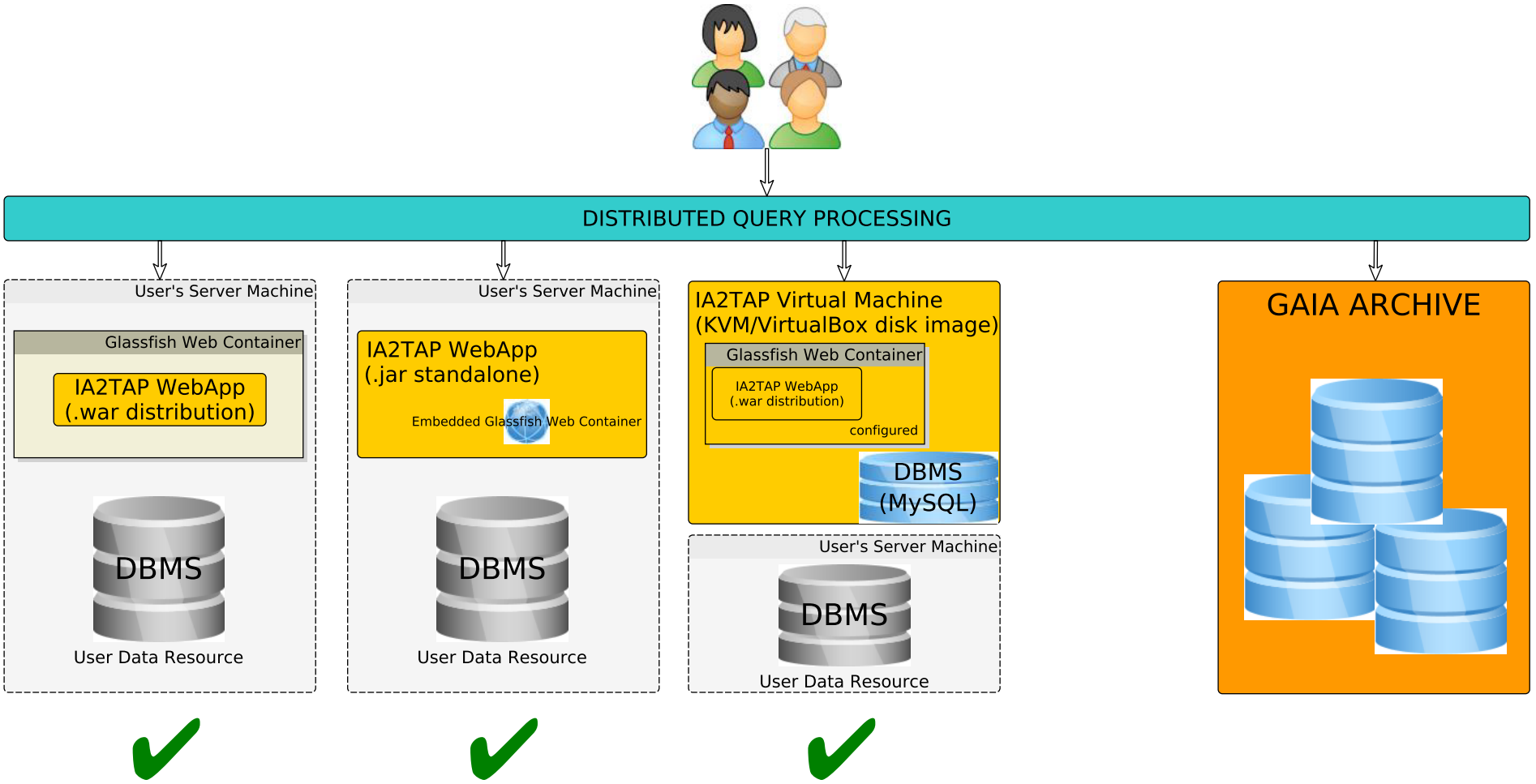
Abstract

We describe a Javascript library for TAP/ADQL Archive User Interfaces that provides an Autofill functionality for ADQL UI textboxes. This assists users while typing an ADQL query, providing them with a context–specific list of suggestions including ADQL keywords and metadata for the service that the library is initiated on.

This work has received funding from the European Community's Seventh Framework Programme (FP7–SPACE–2013–1) under grant agreement number 606740.

GENIUS
Deliverable D3.4

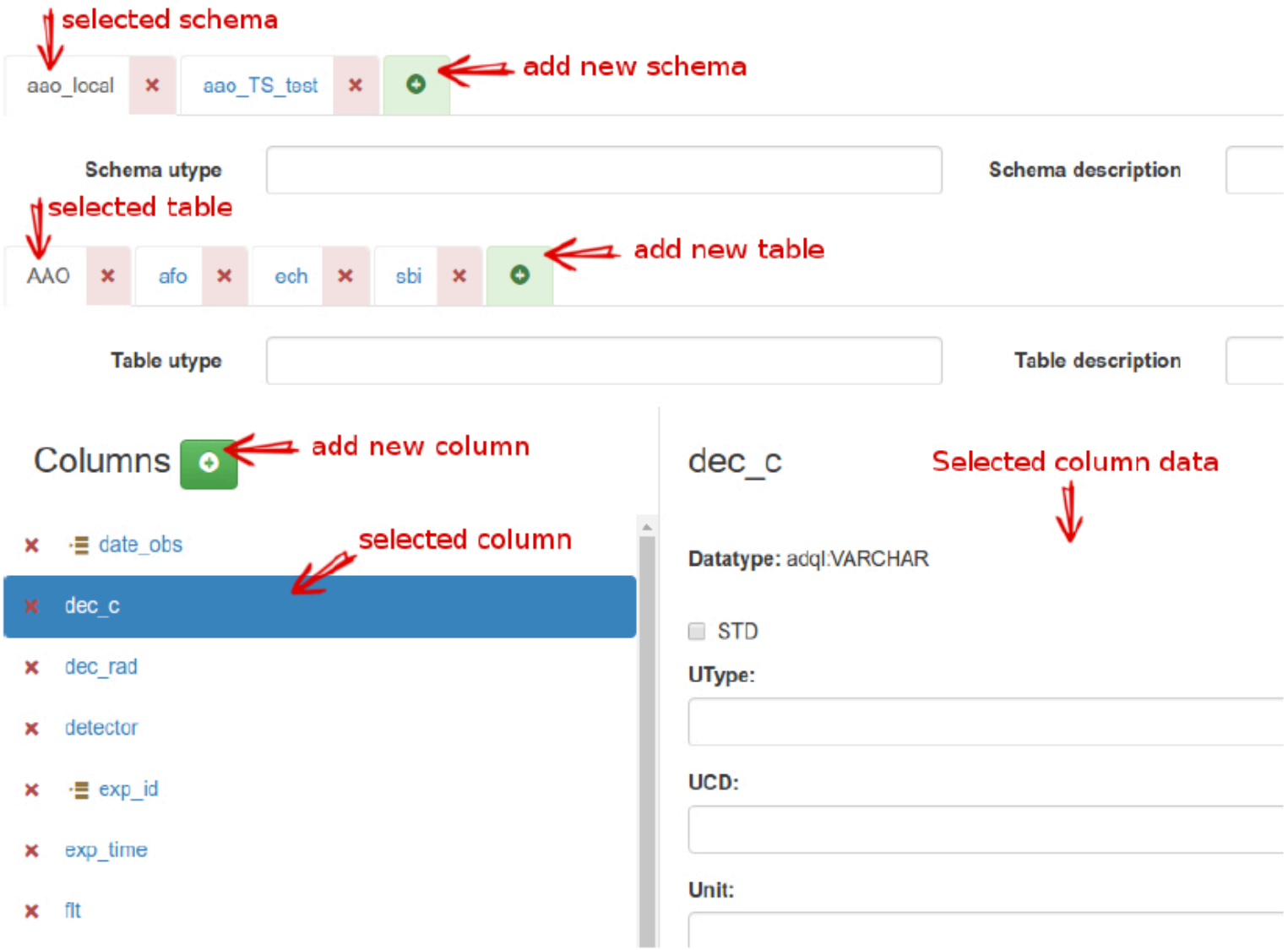
T3.3 (cont.): VO Infrastructure



✔ ✔ ✔

http://ia2.inaf.it/index.php/14-projects/39-eu-genius

GENIUS
Deliverable D3.4



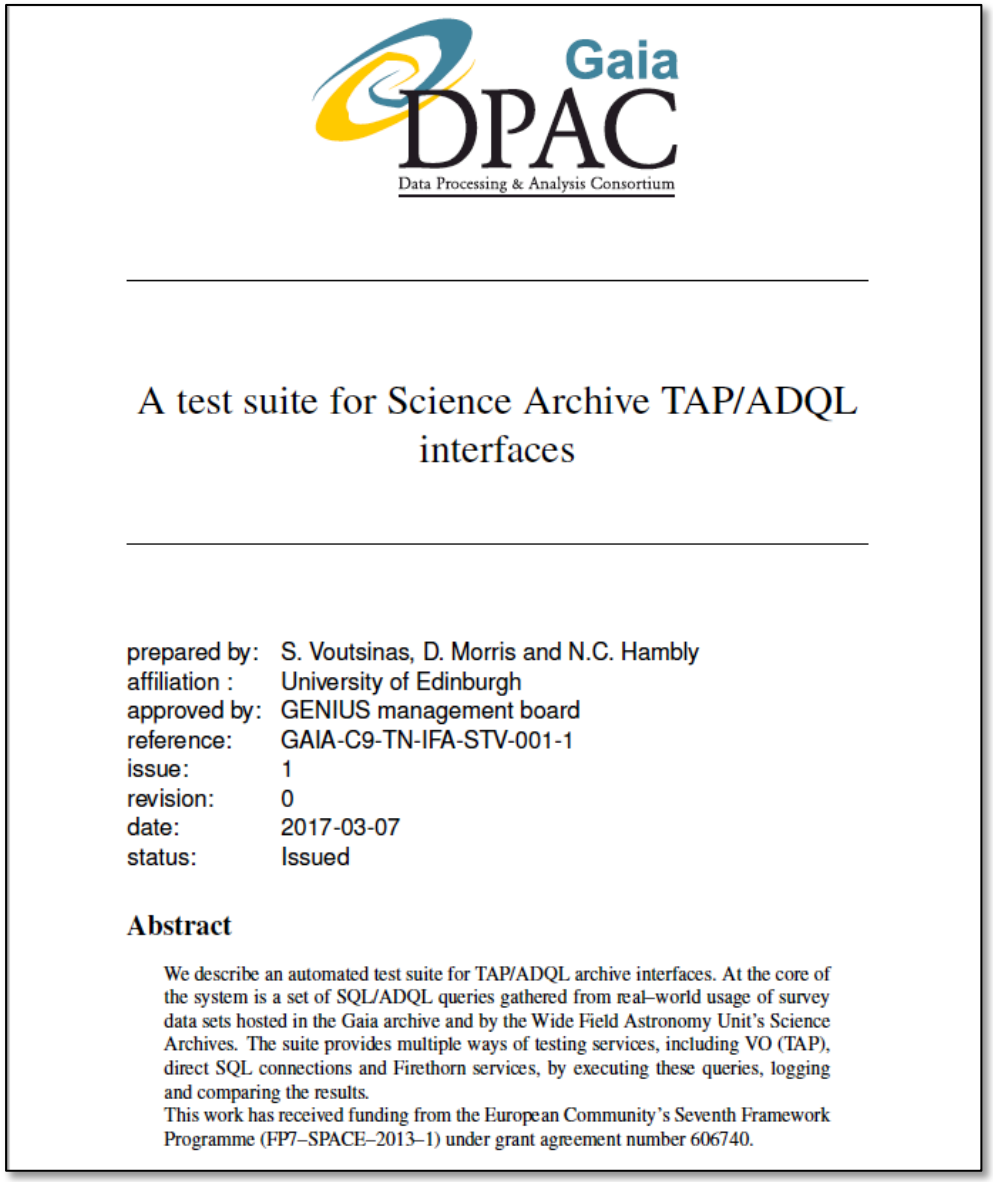
e.g. TAP Schema editing via IA2TAP GUI application

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



Gaia DPAC
Data Processing & Analysis Consortium

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-001-1
issue: 1
revision: 0
date: 2017-03-07
status: Issued

Abstract

We describe an automated test suite for TAP/ADQL archive interfaces. At the core of the system is a set of SQL/ADQL queries gathered from real–world usage of survey data sets hosted in the Gaia archive and by the Wide Field Astronomy Unit's Science Archives. The suite provides multiple ways of testing services, including VO (TAP), direct SQL connections and Firethorn services, by executing these queries, logging and comparing the results.

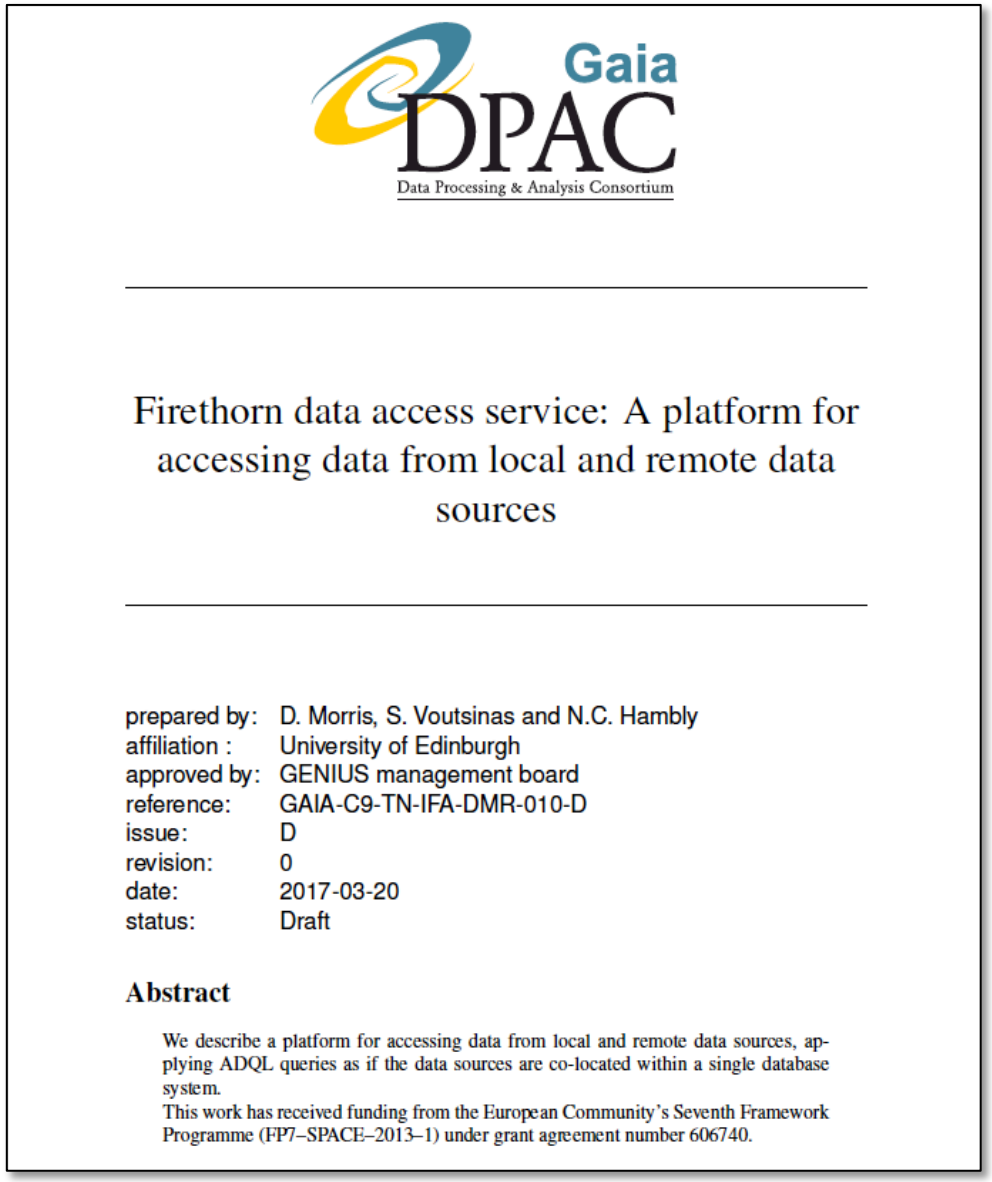
This work has received funding from the European Community's Seventh Framework Programme (FP7–SPACE–2013–1) under grant agreement number 606740.

- β-testing
- Stress-testing

GENIUS:

- Deliverable D3.5, but also covers in part
  - Milestone MS9
  - Deliverables D3.4 and 3.6

T3.4 (cont.): Distributed Query Processing



Gaia DPAC
Data Processing & Analysis Consortium

Firethorn data access service: A platform for accessing data from local and remote data sources

prepared by: D. Morris, S. Voutsinas and N.C. Hambly
affiliation : University of Edinburgh

approved by: GENIUS management board
reference: GAIA-C9-TN-IFA-DMR-010-D
issue: D
revision: 0
date: 2017-03-20
status: Draft

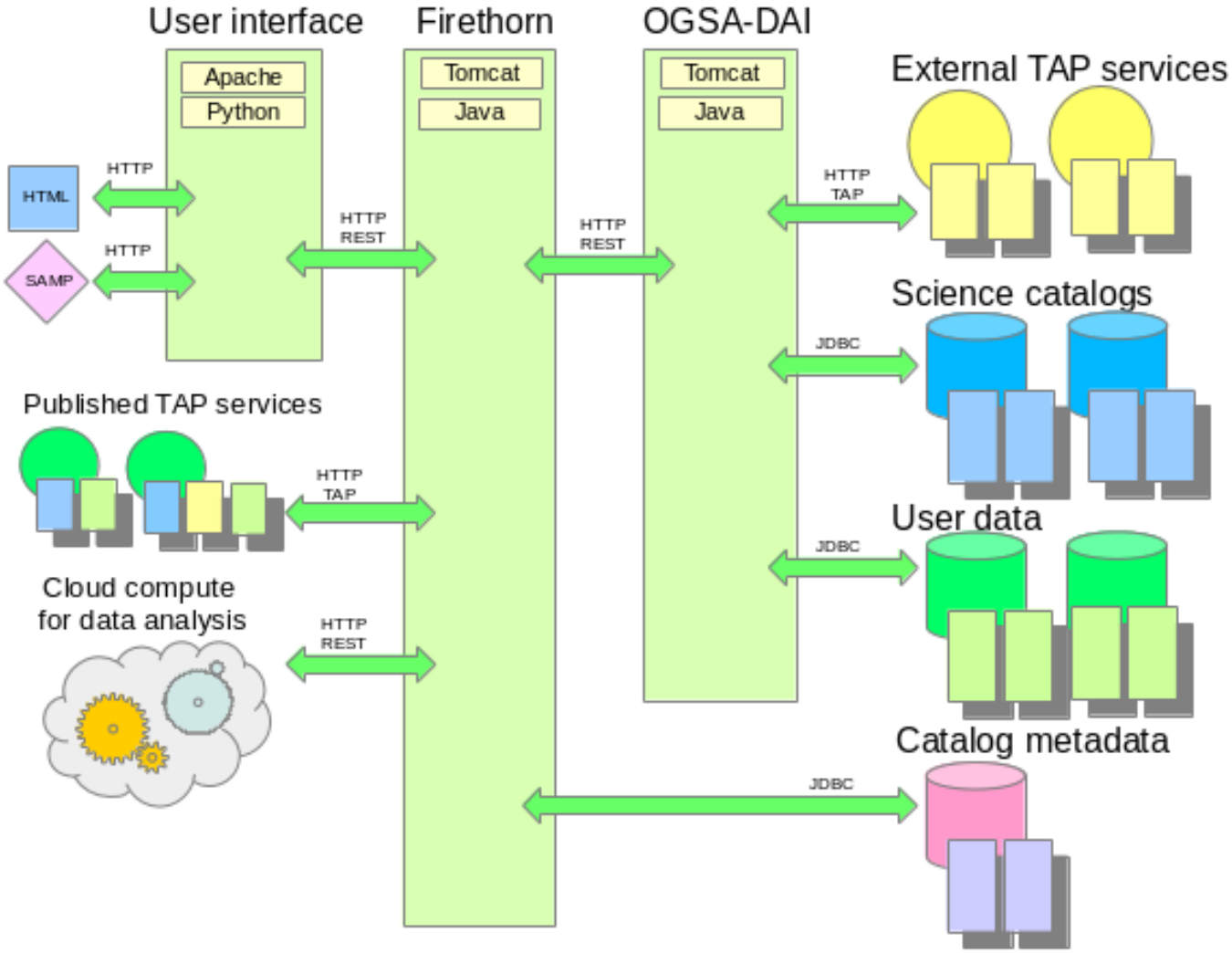
Abstract

We describe a platform for accessing data from local and remote data sources, applying ADQL queries as if the data sources are co-located within a single database system.

This work has received funding from the European Community's Seventh Framework Programme (FP7–SPACE–2013–1) under grant agreement number 606740.

GENIUS:
- Deliverable D3.6

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



T3.4 (cont.): infrastructure demonstrator

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

GENIUS
Deliverable D3.2

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)

GENIUS
Deliverable D3.7

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: Distributed Query Processing (GAIA-C9-TN-IFA-DM-001) Mirroring Applications Interface (GAIA-C9-SP-UB-XL-034)	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	01/03/2017	GENIUS deliverable D3.3, "IVOA compliance document" in the DPAC SVN as follows:GAIA-C9-TN-IFA-NCH-040.pdf Deliverable 3.3	Submitted
D3.4	Deployed web services, code and documentation	WP3	31/03/2017	22/03/2017	Deliverable 3.4	Submitted
D3.5	Data centre report and analysis document	WP3	31/03/2017		Deliverable3_5.pdf	Submitted
D3.6	TAP+ code and documentation	WP3	31/03/2017	24/03/207	Deliverable 3.6	Submitted
D3.7	Deployed CANFAR–style VM research environment and produced reports and documentation	WP3	01/04/2017		Deliverable3_7.pdf	Submitted

GENIUS legacy from WP3

- ✓ We take seriously our responsibility for dissemination of results and SW coming out of the R&D programme using well-established Digital Curation mechanisms (and avoiding ad-hoc, ephemeral web resources):
  - Software and associated documentation in GitHub
  - Mainstream electronic publishing (e.g. article in Astronomy & Computing)
  - Participation and integration in the international Virtual Observatory (developments of which adhere to the first two points)
  - Re-use of knowledge in other areas and projects (e.g. taking forward FP7 GENIUS experience into H2020 ASTERICS)

Examples from WP3

- ✓ GitHub resources (whole or in part):
  - https://github.com/stvoutsin/pyrothorn
  - https://github.com/stvoutsin/tap-autocomplete
  - https://github.com/stvoutsin/taplib
- ✓ IVOA contributions:
  - http://wiki.ivoa.net/internal/IVOA/InterOpMay2016-DAL/adql-20160509.pdf
- ✓ A&C paper (submitted end Jan 2017):
  - *"Use of Docker for deployment and testing of astronomy software"*
- ✓ Data Centres using GENIUS-developed software:
  - ESAC Science Data Centre
  - Wide Field Astronomy Unit (University of Edinburgh, UK)
  - In fact anywhere that uses IVOA infrastructure, e.g. ADQL parser