GENIUS mid-term review

Report for WP3

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

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

- Nigel Hambly [1], IfA Edinburgh University (manager; 0.15 FTE funded by GENIUS rising to 0.35(GENIUS) + 0.25 = 0.6 FTE in October 2014)
- Jesus Salgado [2], ESAC Science Archives Team (deputy manager and ESAC-SAT main contact; 0.5)
- Juan Gonzalez, ESAC SAT (0.5)
- Raul Guitierrez [3], ESAC SAT (1.0)
- Juan Carlos Segovia [4], ESAC SAT (1.0)
- Wil O'Mullane, ESAC (<0.1)
- Gabriele Comoretto (<0.1)
- Harry Enke [5], AIP eScience Team (0.25)
- Adrian Partl [6], AIP eScience Team (0.75 to end June 2014)
- Jochen Klar [7], AIP eScience Team (0.75? from July 1st 2014)
- Giuliano Giuffrida [8], INAF-ASDC (0.3)
- Federica Moscato [9], INAF-ASDC (0.2)
- Paola Marrese [10], INAF-ASDC (0.2)
- Andrea Baruffolo [11], INAF-OAPd (0.2)
- Marco Molinaro [12], INAF-OATs (0.15; 0.15 FTE funded via GENIUS from Jan 2014)
- Riccardo Smareglia [13], INAF-OATs (0.15; 0.15 FTE funded via GENIUS from Jan 2014)
- I-Chun Shih (Stephen) [14], Observatoire Paris Meudon (0.1; works mainly in WP940)
- Stelios Voutsinas [15], IfA Edinburgh University (0.2 FTE from October 2013 funded by GENIUS)
- Dave Morris [16], 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)
- Jerome Berthier [17], Institut de Mecanique Celeste et de Calcul des Ephemerides (IMCCE), 1.8 staff months from TBC (funded by GENIUS)
- Enrique Solano [18], Centro de Astrobiologia (CSIC/INTA) Madrid, 6 staff months in second half of 2015 (funded by GENIUS)
- Paul McMillan [19], Universities of Oxford & Lund, from Q4 2014, XX FTE (WP936)
- Robert Butora [20], INAF-OATs (0.5 FTE funded via GENIUS from June 2014)

Work Breakdown Structure

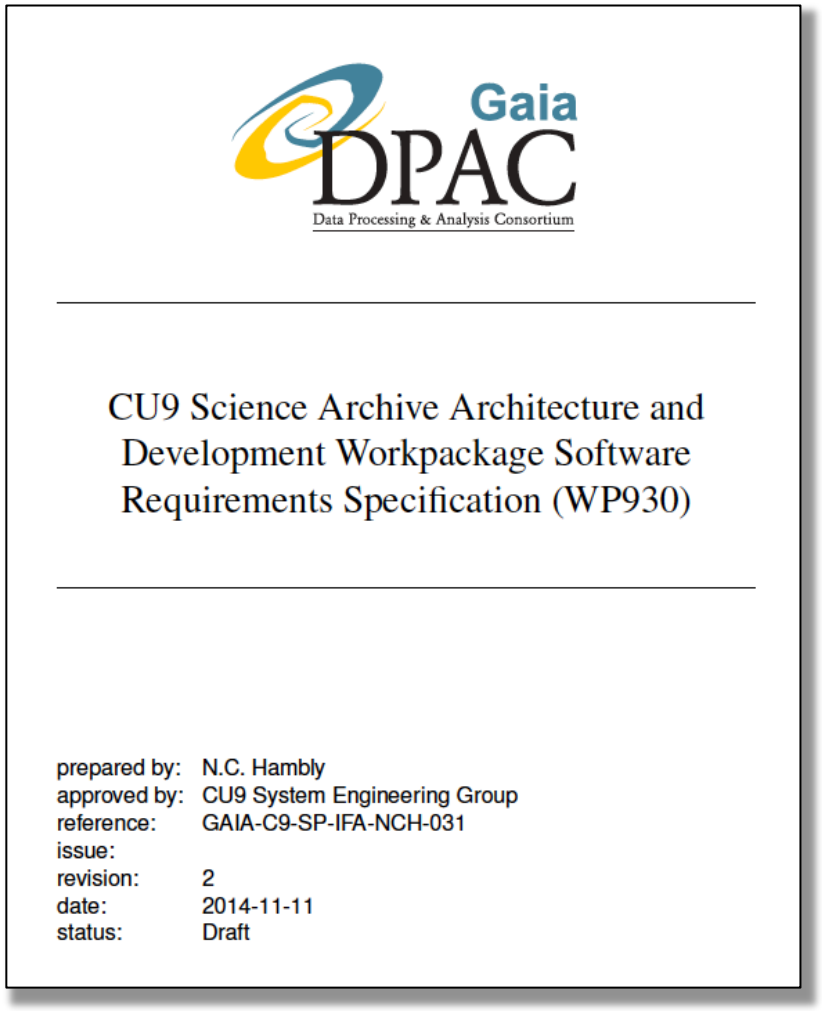
- T3.1: Technical Co-ordination
  - System Requirements Specification
  - Systems Interface Control: ICDs
- T3.2: Aspects of Archive (end-user) Interface Design
  - Subsystems interface infrastructure (affecting end-user experience)
  - Enhanced features for User Interfaces
- T3.3: VO Infrastructure
  - 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 (a total of 8.3 FTE with 3 FTE at ESAC-SAT, 3.6 FTE national funding agencies and 1.7 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

A Requirements traceability

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

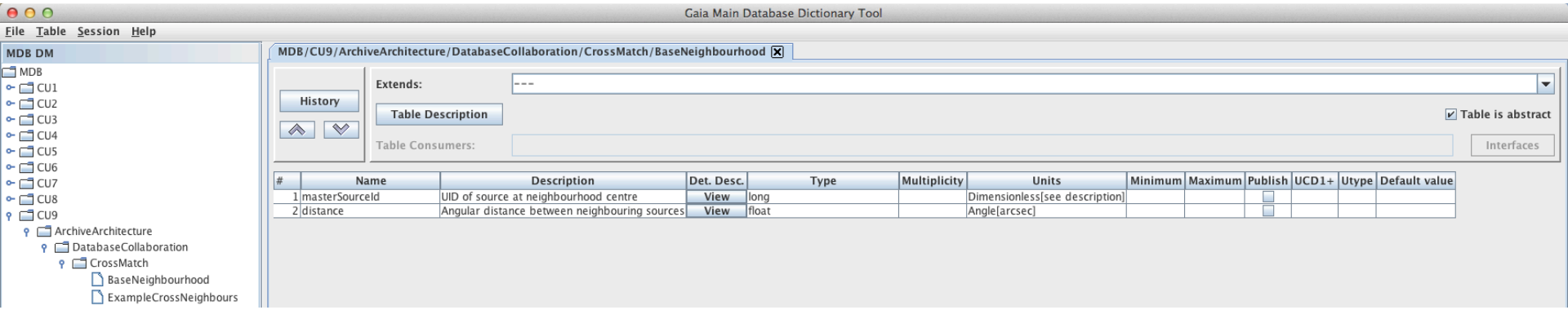
Parent Requirement	Requirements in this document
CU9-ADV-T-FUN-080	CU9-WP935-T-FUNC-100, CU9-WP935-T-FUNC-120, CU9-WP935-T-FUNC-140
CU9-ADV-T-FUN-200	CU9-WP935-T-FUNC-160
CU9-ARC-M-FUN-020	CU9-WP935-T-FUNC-180
CU9-ARC-M-PLN-020	CU9-WP931-M-PLAN-020
CU9-ARC-M-PLN-040	CU9-WP931-M-PLAN-040
CU9-ARC-M-PLN-060	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

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
  - Concentrate on these before proceeding any contributions to the User Interface itself

T3.2 (cont.): Dictionary Tool enhancements

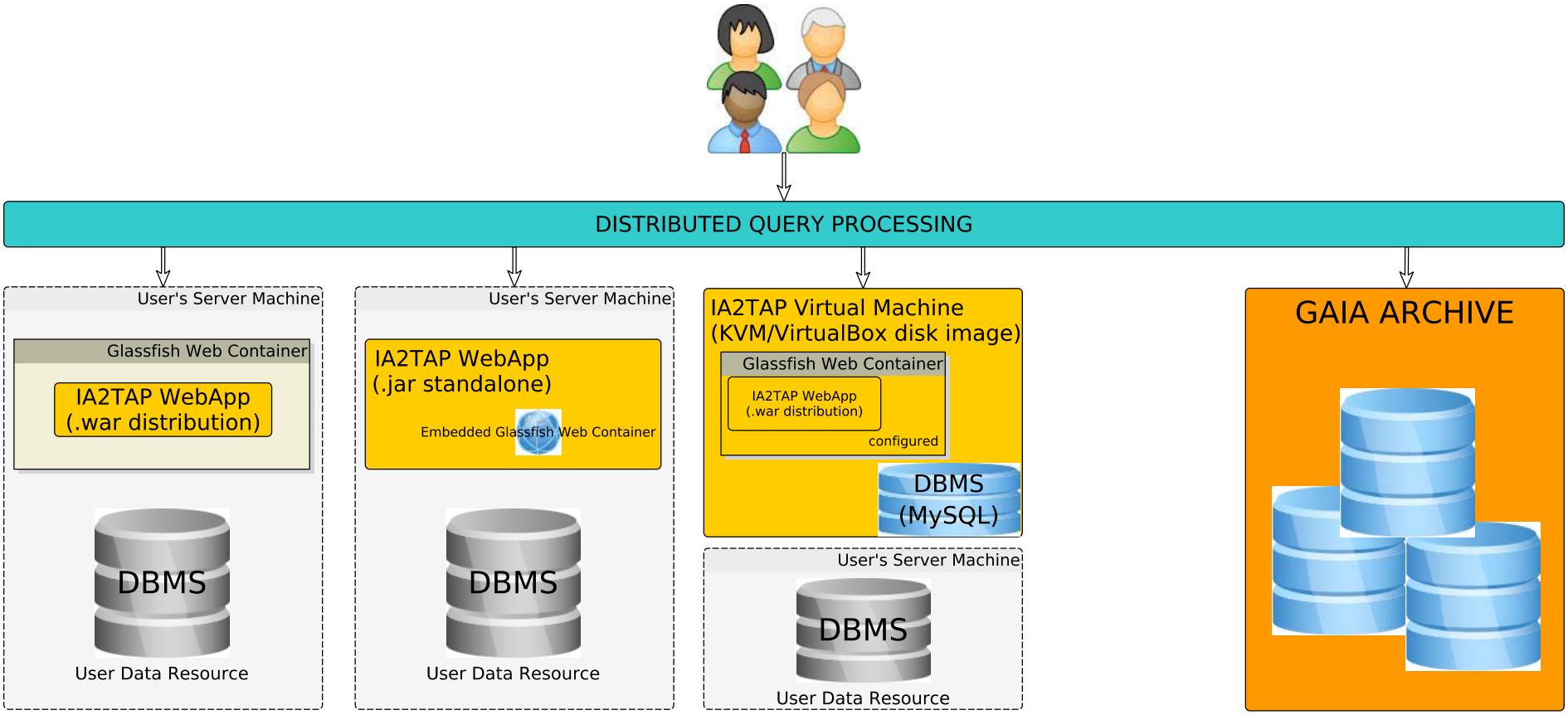
- New metadata fields
- Additional propagation features
  - Ensure UIs contain all necessary information
- New version of tool release earlier this year



T3.3: VO Infrastructure

- VO-Dance, a client-side integration tool
  - Allowing the end-user to publish to the VO
  - Work on deployment as a Virtual Appliance
- IVOA activities
  - ADQL standards
  - ADQL parser enhancements

T3.3 (cont.): VO Infrastructure

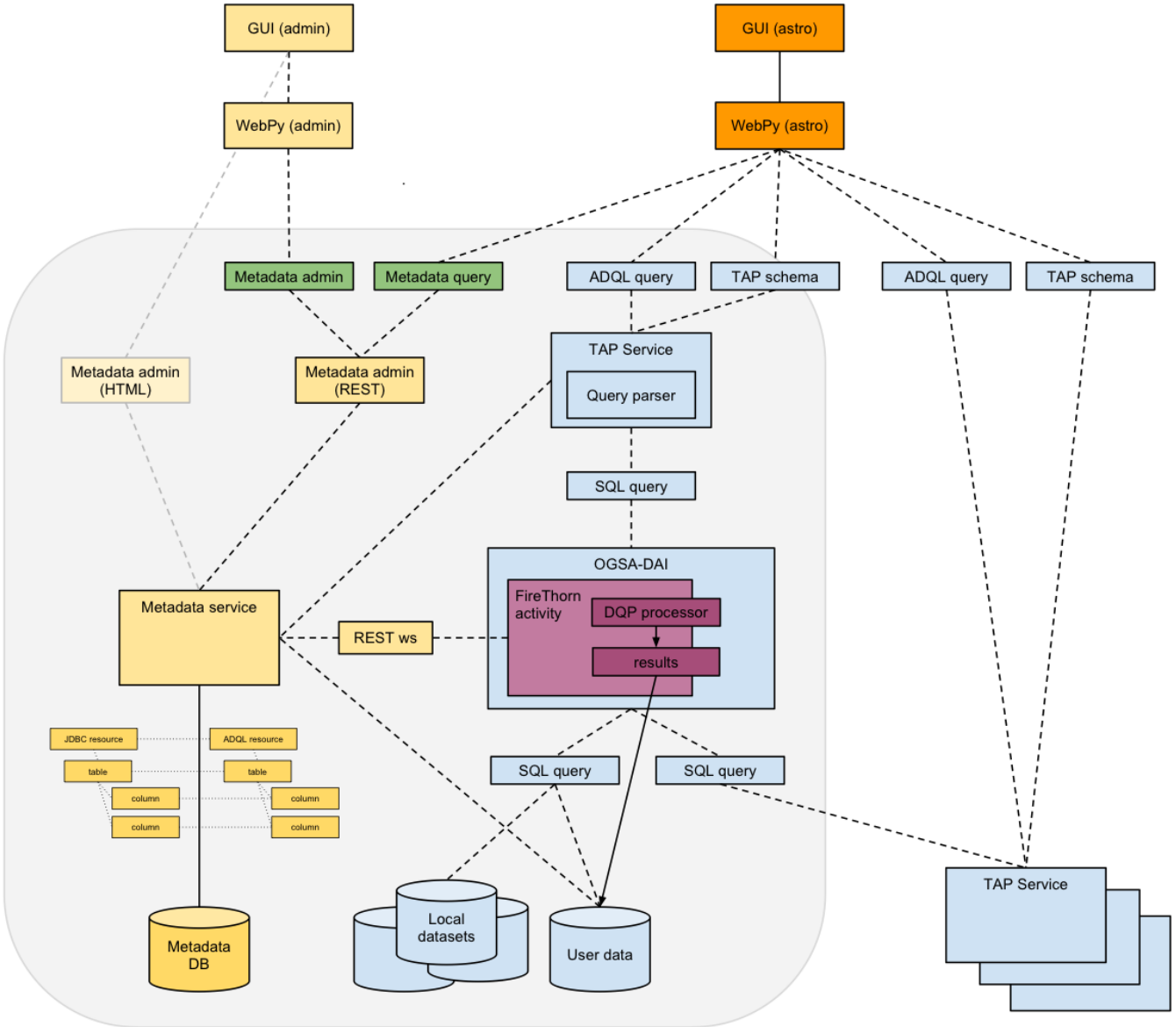


T3.4: Data Centre Collaboration

- 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

$\Rightarrow$ Distributed Query Processing is required

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



T3.4 (cont.): infrastructure demonstrator

- Deliverable D3.2: see http://genius.roe.ac.uk
  - "Web2.0" enhancements, e.g. ADQL auto-completion
  - DQP prototype

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**

In response to 1st year review…

Roadmap for second half of WP300 development programme

In response to the first year external review, an outline roadmap of supplementary milestones and deliverables for the WP300 development programme

Milestones:

Related milestones from the main plan:

GENIUS-MS9 (1/10/2015): User prototype review

GENIUS-MS12 (1/10/2015): Prototype archive tools open to the community

GENIUS-MS13 (1/9/2016): Stress testing of enhanced archive tools

GENIUS-MS16: (1/4/2017): GENIUS products availability

Deliverables:

Interim Milestones/Deliverables:

Q4 2015 (towards D3.7): First full draft of VM research environment paper for A&C

Q2 2016 (towards D3.4 & D3.6): Prototype enhanced user interface functions (Web2.0 and DQP) integrated into GACS (for beta-testing)

Q3 2016 (towards D3.4 & D3.6): First deployment of enhanced user interface functions (Web2.0) within GACS (open to end users)

Q4 2016 (towards D3.4 & D3.6): First deployment of enhanced user interface functions (DQP) within GACS (open to end users)

- See https://gaia.am.ub.es/Twiki/bin/view/GENIUS/WP300Part2Roadmap

In response to 2nd year review…

- Milestone M9 (“User prototype archive review”; 1/10/2015) is delayed 3-6 months (in order to synchronise with DPAC CU9 Gaia DR1 rehearsal and beta testing of GACS)

⇒ Working to include testing of GENIUS UI/DQP within CU9-wide beta-testing