EUDAT Registry – Overview for SAF (26/04/2012)

John kennedy, Tatyana Khan

Introduction:

The Purpose of this document is to provide a more detailed overview of the EUDAT Registry status and plans and to request confirmation and feedback regarding the current course of action.

What is the EUDAT Registry:

The EUDAT Registry is a central service containing information about the sites, contacts, maintenance schedules, and services which are available within EUDAT. This information is to be maintained by site/service administrators themselves, thus ensuring flexibility and scalability within the operations system. The Registry should become a single source of information about the EUDAT Site and Service Infrastructure.

What do we gain:

The advantages of deploying such a Registry are several fold. The Registry will:

- Allow communities to see which service are available for them – service discovery
- Allow the EUDAT Operations team to see which services need to be *Managed* and *Monitored*
- Provide a Global view of the EUDAT Service Infrastructure and as such will allow us to make authoritative statements like "EUDAT consists of 87 services which are deployed at 12 core data centers, of which there are 27 iRODS instance, 4PID Instances...."
- Allow us to communicate with sites easily and reliable in case of problems – for instance service failure or security issues/alerts.
- Allow details of the EUDAT Service Infrastructure to be exposed to other e-Infrastructure projects and user communities and as such enable collaboration.

The Registry is a core component of the EUDAT operations plan.

Requirements:

To ensure that the Registry adequately covers the needs of EUDAT, use-cases, requirements and general input/feedback has been gathered from WP6,WP5,WP4(CLARIN) and the site managers at the EUDAT Operations 1st year sites (CINES,CSC,Jülich,RZG,SARA).

The use cases and requirements thus far considered include areas such as Service monitoring, Service deployment and management, Registry configuration and administration, and community support.

A more detailed description of use-cases and requirements can be found at:
https://confluence.csc.fi/display/Eudat/T6.1.2+-+Site+Registry

The use-case and requirements capture needs to be extended to better cover the communities supported by EUDAT. Currently input has been provided in the form of a CLARIN requirements document and informal discussions with some community members. A formal approach to other communities is planned to take place within May 2012.

In additional to the functional requirements it is important that a solution is found that is considered to be production ready and well supported.

Solutions in sight:

After some consideration the GOCDB was chosen as a suitable candidate for extended testing. The GOCDB is an open source product which is developed jointly by STFC (a EUDAT partner) and EGI (a European Infrastructure provider and EUDAT peer).

The GOCDB offers:

- A mature solution – which has been in productive use for several years within EGEE/EGI
- A solution which works virtually "out of the box" for EUDAT (however some customization is needed)
- Many EUDAT requirements already covered or already on development roadmap for 2012
- A Web Interface and REST API
- A central service for site/service registration

It is important to note that EUDAT sees the GOCDB as a customizable product. If the GOCDB is selected as a registry solution the EUDAT instance would not be developed in parallel to the main development stream. This ensures that no divergence in the code base occurs and that updates, both functional and security patches, to the GOCDB can be directly used by EUDAT.

In addition to the GOCDB a second registry service solution, the EMI Registry (EMIR), is considered to be of some interest. EMIR is, however, a product which is currently still under development and as such is considered to be a solution which we should monitor for possible future use. Although some functional overlap exists between the GOCDB and EMIR they are sufficiently different that they may both be of interest to EUDAT. As the current use-cases and requirements are expanded and as EMIR matures we will gain a better idea of its possible value for EUDAT.

Test System Status:

A test instance of the GOCDB was deployed at RZG in Jan 2012. Once the deployment was completed the system was customized for EUDAT. This proved to be quite time consuming for two main reasons. Firstly the new GOCDB admins at RZG needed to gain a deeper understanding of the service and secondly the GOCDB itself was strongly orientated towards EGEE/EGI – their main customer. Following discussions with the developers of the GOCDB the customization process has been simplified and future deployments are expected to be more straight forward.

Several rounds of discussions regarding the EUDAT requirements and customization needs were undertaken with the GOCDB development team.

For test purposes 1st year EUDAT Operations sites (CINES,CSC,FZJ,RZG,SARA) were asked to input data (in some cases dummy data) and provide feedback (see Fig1 and Fig2). This proved to be a good learning exercise for both the GOCDB admins and the site admins.

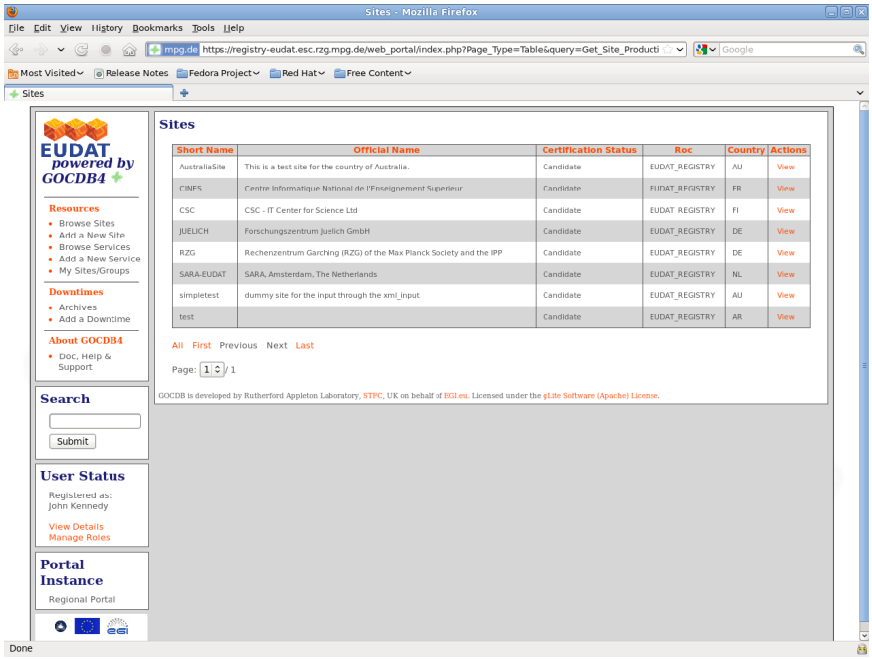

Fig1: Screen shot of Registry showing registered sites.

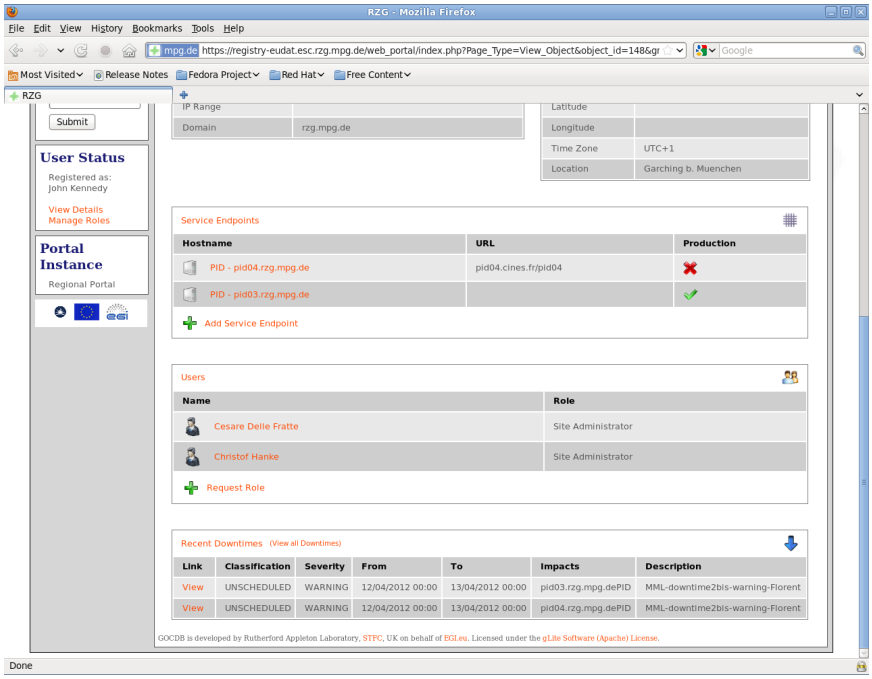

Fig2: Screen shot of registry showing detailed site information.

Once an initial understanding of the GOCDB was gained and the data for the 1st sites was added a test of the basic integration with the EUDAT monitoring services was undertaken. This was a proof of principle exercise that showed that the monitoring services could query the GOCDB via the REST interface and subsequently configure the nagios system to monitor all the available EUDAT services. Following this test we can confidently state that one of the core operations use cases is covered (see Fig 3).

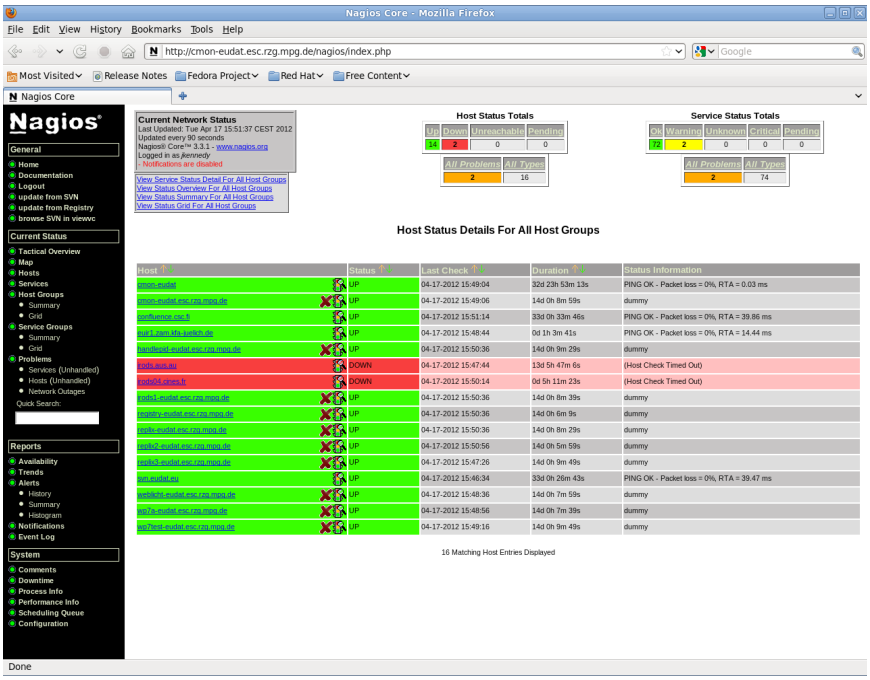

Fig3: Screen shot of Nagios service monitoring built from Registry information.

Much has been learned during this test period and the end to end test – service deployment, information input and subsequent monitoring discovery has provided us with much confidence in the GOCDB as a solution.

It should be noted that the GOCDB developers have proven to be responsive and supportive throughout the above process. During the test period we have gained the impression that the GOCDB development is well supported and is undertaken in a professional and reliable manner. Adequate documentation is provided, queries are answered promptly and reliably and development roadmaps etc are provided which allows us as a customer to understand and also influence the direction of development.

Life without a Registry?:

Following the initial use-case and requirements capture and the evaluation via the test system described above it is fair to undertake some reflection and consider the implications of life without such a registry service and to ask if a simple set of alternatives could be used.
If we consider life with no such service/functionality we find:

- No single point of information describing Infrastructure (however see simpler approach description below)
- Operations are hindered due to the lack of infrastructure description and contact details
- Communities are hindered – no service discovery
- Monitoring is hindered – no ability to dynamically create monitoring views based on registered services
- Possible integration with other Infrastructure and communities is hindered – no ability to expose our service landscape

A possible simpler approach would be information exchange via e-mail and the use of web or wiki sites to render the information persistent and make it available to to outside world. Initialy such a solution

may seem feasible, everyone can modify wiki pages, everyone can join the mailing list etc. However once more consideration is given to such a “simple” solution it is found to be lacking in several areas and as we address these, ensuring sites and services follow correct schema definitions, ensuring the system is queryable via some API (e.g. REST) etc we are essentially forced to build a registry.

Thus this brief period of reflection reinforces our initial desire for such a registry system.

Next Steps:

Here we provide a brief outline of the proposed next steps (target dates given where possible):

- **More discussion with communities:** We have already received input from CLARIN since they have produced a requirements document for a separate project. We would like to discuss with other communities regarding their needs in this area. We should note that the ability to group services into community like views is already considered as a requirement and has just become available in the latest GOCDB release (Start in May 2012 – already discussed with Willem Elbers).
- **Upgrade to latest GOCDB Release:** A new version, with functionality of interest to EUDAT, was released in early April. We aim to upgrade and test this version (May 2012).
- **Continue work on customization of our instance:** In collaboration with developers to ensure that the EUDAT instance fits our functional needs and also provides and interface which is tailored towards EUDAT.
- **Continue Monitoring Integration:** A list of requests has been gathered from the monitoring working group and we will look at fulfilling these.
- **Work through metric of use-cases and requirements:** Detail which of the current use-cases are covered and elaborate on strategies and mitigation for any areas which are not covered.
- **Move into Pre-Production:** Following the above mentioned steps, and given the acceptance of this proposal by the SAF/EB, we would like to move into a pre-production phase in August/September 2012. This would allow the service to be tested and assessed in a broader manner.

The above steps aim to put in place a pre-production registry service in M11-M12 (EUDAT time-frame). The milestone for the deployment of this service is M18, as such a 6 month period for extended testing and integration of solutions for community requirements exists.

Open Issues:

Several open issues remain, however in some cases it is currently unclear to what extent they will impact the project. For the sake of openness we detail them here (adding mitigation info where appropriate):

- **AAI – Technical Issue:** The current AAI solution used by the GOCDB is X.509 based. This is unlikely to be acceptable as a final solution for EUDAT, however it is not seen as a problem for the initial testing phase. The GOCDB development team has already agreed to adapt the GOCDB to allow for alternative AAI solutions to be used (e.g. Shibboleth, SAML). EUDAT would need to provide a suitable plugin to provide support for our chosen AAI solution.
- **Oracle Database – Technical Issue:** The current database solution used by the GOCDB is Oracle, a commercial solution. A license may be required in the future depending on the requirements placed on the system. Porting to an alternate, free, database solution would probably entail significant effort because the business logic for the GOCDB is embedded into the database itself. A port would also need to be supported to ensure that no forking of the GOCDB source code is required. Currently the Oracle XE free database is in use for testing and

no scalability issues have as yet be observed.

- **Exploding Requirements – Technical Issue:** The current EUDAT requirements appear to be either covered by the GOCDB or part of the already defined development roadmap. The development team have also expressed a willingness to support EUDAT, extending their functionality where needed, however if the EUDAT requirements become too extensive a more scalable solution would be needed, for instance the assignment of some EUDAT STFC PMs to aid with the implementation.
- **Site, Service-Type registration procedure – Policy Issue:** A procedure needs to be put in place that defines how a new site or service-type is added to the registry. A simple and flexible solution could be to allow the WP4 and WP6 leaders to decide with the SAF's seal of approval.
- **Exposing contact details – Policy Issue:** We need to define a policy regarding which details are to be opened internally to EUDAT and/or to the outside world regarding site admin contacts (email,phone numbers etc). It may also be the case that different centers have different policies regarding this.

Questions to SAF:

The SAF is requested to confirm that the current course of action is considered to be correct for EUDAT and that the workplan, including the next steps and timeline is acceptable. Furthermore the SAF is asked to raise any concerns that they have and provide suggestions where appropriate.