8th Meeting of the VT Project on

Life Science Reference Data Integration

Integrating ELIXIR reference datasets within the European Grid Infrastructure

**Date**: Wed 03 June 2015 (03/06/2015) 12:00 CET – Webex Teleconference

Contents

[Participants 1](#_Toc421104380)

[Agenda 1](#_Toc421104381)

[Minutes 2](#_Toc421104382)

[Discussion / Progress report 2](#_Toc421104383)

[Planned actions 3](#_Toc421104384)

[Useful Links 4](#_Toc421104385)

# Participants

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** |  | **Affiliation** | **Country** |
| Fotis E. Psomopoulos | FP | AUTH | GR |
| Marios Chatziangelou | MC | IASA | GR |
| William Karageorgos | WK | IASA | GR |
| Diego Scardaci | DS | INFN/EGI.eu | NL |

# Agenda

* Follow-up from the EGI Conference in Lisbon
* Steps towards the first pilot case on NGS reference data

# Minutes

## Discussion / Progress report

**Discussion**

***FP***: Quick overview of the discussions at the EGI Conference 2015 in Lisbon. Particular focus on the discussion after the ELIXIR CC meeting with Steven Newhouse. There is an agreement on the need for reference data replication in Life Sciences; however, implementation-wise, and within the context of the ELIXIR CC, a use-case on Data Replication would probably have the following format:

*A data replication workflow will use the AppDB to provide the "human" input into the system through the web forms and to trigger the transfers, whilst the EUDAT software will be used to register and record the existence of the dataset and the location of replicas, and Globus Transfer through GridFTP endpoints to move the data.*

***FP***: Although the technical implementation of the proposed workflow is currently beyond the scope of this VT, it is evident that there is significant overlap between the two use-cases.

***MC***: Agree that it is currently outside our scope. In any case, the easier solution is for EUDAT services to directly connect and update AppDB through our provided API.

MC: Also, any "request-for-replication" connection to AppDB should be eponymous (through login). On the other hand, as datasets are (or should be) publicly available, all information on the datasets (such as status, existence, number of replicas etc), should also be publicly available (i.e. without login).

***FP***: We have currently two test cases for implementation, (a) Ensembl plant data, and (b) Ensembl human index data files for NGS analysis (see also minutes of 7th teleconference).

***FP***: Regarding the first test case, we do not have much feedback yet. Regarding the second test case, we know both dataset as well as applications that can be used with that. Also regarding the second test case, after discussion with Kimmo Matila at the EGI Conference, we cannot use the Chipster case for the provided data (Chipster periodicallyproduces the index files in a manual way).

***FP***: (Action 1) Will provide a short document that will outline the second use case including information such as Dataset (URL, description, characteristics), Applications (Name, URL, syntax).

***DS***: We can use this document for a discussion over the mailing list regarding a first implementation.

***DS***: Also EGI has started a pilot case with EUDAT, aiming to connect the services of the EGI Federated Cloud with a couple of services provided by EUDAT. The pilot has already started and we can expect some first results by the end of July → investigate potential connection with this VT

***MC***: Within Task 6 (Integration of AppDB with the ELIXIR registry), there was a discussion with Jon Ison. Until June 10th we expect from ELIXIR a list of software appliances and services that will be curated from the ELIXIR point of view. This list will then be open to suggestions and comments regarding the meta-data available.

***MC***: After this process is completed (including any corrective actions that might be needed), we will be ready for a first integration of AppDB with the ELIXIR registry.

* 1st integration will focus on services and software appliances
* 2nd integration (and given that we will have a dataset listed) will allow for datasets

***DS***: BILS is also active in moving services to the Federated Cloud. As soon as this VT is mature, we can also attempt to connect with BILS.

***MC***: As IASA is currently not a member of EUDAT, we will need some contact point in EUDAT to understand which services are available (i.e. API connection to EUDAT services). DS will connect with MC over email to provide such contact.

## Planned actions

*Action 1*: Will provide a short document that will outline the use case including information such as Dataset (URL, description, characteristics), Applications (Name, URL, syntax) for Test Case 2. (***FP***)

*Action 2*: Based on the provided document from Action 1, have a discussion over the mailing list towards a first implementation. (***All***)

*Next teleconference*: Monday, June 22nd, 12:00 Amsterdam time

# Useful Links

**Pilot Wiki Page**: <https://wiki.egi.eu/wiki/VT_Life_Science_Data_Integration>

**Dedicated DocDB**: <https://documents.egi.eu/public/ShowDocument?docid=2361>

**Pilot Mailing List**: <https://mailman.egi.eu/mailman/listinfo/lsrdi-vt>

**Resources Google Sheet**: [Google Sheet](https://docs.google.com/spreadsheets/d/1gcSNCNBdwIgNN8URrj0b2w4WT18VA04BHOi5csT0HGM/edit#gid=1080851537)