5th Meeting of the EGI-ELIXIR Pilot

Integrating ELIXIR reference datasets within the European Grid Infrastructure

**Date**: Mon 23 Mar 2015 (23/03/2015) 12:00 CET – Webex Teleconference

Contents

[Participants 2](#_Toc409094196)

[Agenda 2](#_Toc409094197)

[Minutes 3](#_Toc409094198)

[Part A: Overview of the general actions for the pilot 3](#_Toc409094199)

[Part B: Progress report in Pilot Tasks 4](#_Toc409094200)

[Task 1: Identify existing life science datasets in EGI 4](#_Toc409094201)

[Task 2: Identify reference datasets for replication 4](#_Toc409094202)

[Task 3: EGI AppDB extension to a dataset registry 5](#_Toc409094203)

[Task 4: Tools for data replication 5](#_Toc409094204)

[Task 5: Analysis tools to work with data replicas 6](#_Toc409094205)

[Task 6: Integration with ELIXIR Registry 6](#_Toc409094206)

[Part C: Discussion and further Actions 7](#_Toc409094207)

[Useful Links 7](#_Toc409094208)

# Participants

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** |  | **Affiliation** | **Country** |
| Fotis E. Psomopoulos | FP | AUTH | GR |
| Afonso Duarte | AD | ITQB | PT |
| Diego Scardaci | DS | INFN/EGI.eu | NL |
| Marios Chatziangelou | MC | IASA | GR |
| William Karageorgos | WK | IASA | GR |
| Gergely Sipos | GS | EGI.eu | NL |
| Brane Leskosek | BL | NGI\_SI | SI |
| Lukasz Dutka | LD | CYFRONET | PL |
| Alexandre Bonvin | AB | eNMR | NL |
| Jan Jona Javorsek | JJ | JSI | SI |
| Kostas Koumantaros | KK | GRNET | GR |

# Agenda

* Discussion/actions on the feedback regarding naming and context
* Presentation/discussion of the Survey results
* Review of the agreed actions (incl. ELIXIR Node endorsements)
* Progress report on the 6 tasks
* Discussion for next steps / actions (incl. a face-to-face meeting at the EGI Conference in Lisbon)

# Minutes

## Part A: Overview of the general actions for the pilot

**1. Change of Project Name**

There have been several discussions with respective Heads of ELIXIR Nodes, as well as with the Director of ELIXIR, concerning the branding (ELIXIR) and the miscommunications (Pilot) that may rise. In order to address these concerns, we will rename this task as following:

EGI Virtual Team Project on "Integrating life science reference datasets within EGI"

*Discussion:*

GS: There are two options; (a) we keep the name and according to the approval or not by ELIXIR, the effort continues or it is completely stopped. (b) we change the name and we continue the effort. It could fit within the ELIXIR CC umbrella or, at worse, as an independent effort of interest to EGI.

MC: It must be noted that a change in name (i.e. removing the "ELIXIR" from the title) might lead to a different, more general focus, and consequently changes in technical level which will require new specifications.

FP: This is not the case. The change in name does not essentially change the nature of this work.

GS: Indeed only in title. However, this effort should be handled in order to produce general guidelines. In other words, setup a discussion and a set of best-practices that will be applicable to other disciplines.

KK: How much effort is required for this project still? Virtual teams / virtual project and/or pilots should have a specific target and timeframe otherwise they end up being never-ending.

FP: The project was initiated Dec 2014 with an expected duration of 9 months. We aim to finish by August 2015 (same time that the CC's start the work)

KK: AppDB should use these use cases to provide common functionally as a generic service for all disciplines

* *FP*: to send an email to all participants in the mailing list for comments (Deadline 26/03/2015)
* *FP*: when all in agreement, send to IT Support [<it-support@egi.eu>](mailto:it-support@egi.eu) to change the names of the wiki page, DocDB and email list.

**2. ELIXIR Letters of Endorsement**

The 2015 ELIXIR all-hands meeting will take place on Monday 30th and Tuesday 31st March 2015 at the Wellcome Trust Genome Campus, Hinxton, UK

We have two Letters of Endorsement, from the SI and GR Nodes respectively. No response from the Italian, Swedish and Portuguese Nodes so far.

Actions

* The project will be presented as a poster in the upcoming all-hands ELIXIR meeting (30-31/03/2015). We aim to increase awareness and get some feedback from the ELIXIR community then
* *GS* to create the poster using the presentation by *FP*
* *BS* and *GS*: regarding the higher feedback from ELIXIR nodes (and better adoption within ELIXIR) have an open talk with Niklas in Hinxton (hopefully clearing the position of this activity).

**3. Presentation of the Project and the Survey responses**

*FP* presented a brief overview of the project, as well as the results of the survey responses so far (presentation file in DocDB: EGI-ELIXIR\_v2.5.pptx)

*AD*: (Slide 11) Don't all of the datasets listed here have subsets?

*FP*: Not necessarily. Yes, you can define filters to retrieve only partial data, but you can always retrieve a dataset listed as "single" as one compact structure.

*AD*: ok so it refers to the way the DB is organized

*AB*: The EBI must be replicating most of those databases... The info is not complete. E.g. for PDB, the EBI is one center part of the wwPDB (worldwide) PDB. See www.pdbe.org

*FP*: Indeed it is not complete yet. The primary location is listed, but this doesn't mean that there aren't any mirrors in other central repositories.

*AD*: How could we achieve a higher feedback from ELIXIR nodes?

*FP*: We've already asked for circulation of the survey through the GR Node, as well as the UK node. However, we still need more disparity in locations.

*KK*: I'm only worried if we have the important datasets in our list

*FP*: Indeed we should have; the current state is the one we expected in the first place.

*GS*: We should add the names of the people that supported one of the top 10 datasets, in order for us to do a follow-up.

## Part B: Progress report in Pilot Tasks

### Task 1: Identify existing life science datasets in EGI (GS)

*GS*: No update. No reason to expect any information from NGIs at this point. We should really focus only on the output of T2.

### Task 2: Identify reference datasets for replication (FP)

Progress

* Questionnaire on Reference Datasets is still live (target audience: end users)  
  URL: https://survey.egi.eu/index.php/652565/lang-en
* First analysis of the relevant datasets (presentation file and document)

Actions:

* Construct a second survey for dataset providers, after receiving some feedback from the Dataset Users survey. The second survey can be in the form of a document to directly engage Dataset providers
  1. who is using it,
  2. how is it provided,
  3. how often is it accessed, etc

Will create a Google Doc to be concurrently filled with the respective questions. All will contribute, both from the technical side as well as the user side.

### Task 3: EGI AppDB extension to a dataset registry (WK)

Progress

* Refined front-end implementation for the Datasets in AppDB  
  URL: <https://appdb-dev.marie.hellasgrid.gr/browse/datasets>

General Comments

*WK*: Current implementation is based on versions of datasets; each entry contains all relevant versions in different tabs

*WK*: We aimed for compatibility with the ELIXIR Dataset Registry (<https://elixir-registry.cbs.dtu.dk/>). However, the Registry view is mostly ontological, so additional investigation is required.

*WK*: Several discussions with the ELIXIR Task Force (Jon Ison, Kristoffer Rapacki) regarding connection to the ELIXIR ontology.

Actions

* *WK:* Modify/Rephrase the link to the ELIXIR Dataset Registry so as to avoid confusion (homepage == dataset elixir page???)
* *FP:* Setup a small group of people to review the front end and provide feedback.

### Task 4: Tools for data replication (LD in place of GD)

General Discussion

LD: The OneData framework software will be deployed next week.

LD: There is a list of potential sites, but currently it will be deployed in INFN.

GD: Finishing the table that could summarize the key capabilities of each tool.

### Task 5: Analysis tools to work with data replicas (AD)

No extra development.

Actions

* Continue updating the table, with additional focus on application already listed on AppDB

### Task 6: Integration with ELIXIR Registry (MC)

Progress:

MC: Meeting with Jon Ison, Kristoffer Rapacki regarding integration with ELIXIR registry (a follow-up to be held on 24/03/2015).

MC: Focus on the vocabulary ELIXIR registry is using, and comparison with the AppDB back end

MC: This information can be seen in the following document:

https://docs.google.com/spreadsheets/d/1UU8h88ajH3C9gkP76fAKHYMILMNuEMU7BXgmi5mExaw/edit?pli=1#gid=0

MC: Identified the list of the metadata that would be most useful from the ELIXIR registry

Actions

* Further investigate the metadata fields in order to identify relevant fields in AppDB and datasets in particular.

## Part C: Discussion and further Actions

*Actions for next period*

* Follow-up on the ELIXIR Nodes’ Endorsement Letter (**General**)

*Next teleconference*: Tuesday, April 21st, 12:00 CET. After this, *teleconferences will be* ***every 2 weeks***.

# Useful Links

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

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

**Pilot Mailing List**: <https://mailman.egi.eu/mailman/listinfo/elixir-pilot>

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