10th Meeting of the VT Project on

Life Science Reference Data Integration

Integrating ELIXIR reference datasets within the European Grid Infrastructure

**Date**: Mon 20 July 2015 (20/07/2015) 12:00 CEST – Webex Teleconference

Contents

[Participants 1](#_Toc422757415)

[Agenda 1](#_Toc422757416)

[Minutes 2](#_Toc422757417)

[Discussion / Progress report 2](#_Toc422757418)

[Planned actions 4](#_Toc422757419)

[Useful Links 5](#_Toc422757420)

# Participants

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** |  | **Affiliation** | **Country** |
| Fotis E. Psomopoulos | FP | AUTH | GR |
| Afonso Duarte | AD | ITQB | PT |
| Kostas Koumantaros | KK | GRNET | GR |
| Marios Chatziangelou | MC | IASA | GR |
| William Karageorgos | WK | IASA | GR |

# Agenda

* Discussion on the 1st use-case for implementation (see also relevant [document](https://documents.egi.eu/secure/RetrieveFile?docid=2361&version=2&filename=RefGenomeIndexReplication_v1.0.docx))
* Next steps towards implementation
* AOB

# Minutes

## Discussion / Progress report

**Discussion**

***FP***: Follow-up on the Pilot Proposal discussed with Steven Newhouse after the EGI Conference in Lisbon. There is a planned meeting between Steven and EUDAT on Monday 20/07/2015, so hopefully there will be further developments in the near future. The work described in this Pilot Proposal is conceptually similar (i.e. replication of reference data), but the underlying technologies have significant differences.

***FP***: KK and his team have setup TopHat and Cufflinks on all HellasGrid sites using CVMFS. The releavant paths are:

Path TopHat : $VO\_SEE\_SW\_DIR/tophat/bin

Path Cufflinks: $VO\_SEE\_SW\_DIR/cufflinks/bin

(for testing purposes, use the following jdl requirement for quickly identifying errors/issues:

requirements = RegExp("cream-ce01.marie.hellasgrid.gr",other.GlueCEUniqueID); )

***FP***: Next steps are the following (Actions on FP):

1. Step 1: Run a job on Grid (SEE VO) using the pre-compiled index files (wget from locations described in previous meetings) with small input files.
2. Step 2: Use LFC to replicate the index files to a second location (***KK***: LFC is simply away to register/reference the copies, you can use whatever technology you like for the transfers)
3. Step 3: Register the dateset and consequent replicas on AppDB (developer version).

***KK***: Do we need a cloud pilot as well?

***FP***: Ideally we would have a best-practices / guidelines document for both Grid and Cloud. However, our current aim is to provide a proof-of-concept that can be applicable on a Grid (Pilot 1.1) and on the Cloud (Pilot 1.2).

***MC***: Bear in mind that whatever data is used in the developer version of AppDB, when we move to the production level, they need to be inserted again.

***FP***: Understood. We will need some accounts in order to manually insert the dataset records.

***WK***: We can provide those (to ***FP*** and ***AD***)

## Planned actions

*Action 1*: ***WK***&***MC*** provide ***FP*** and ***AD*** for accounts on the development site of AppDB.

*Action 2*: ***FP***: Run the pilot case on the NGI-GR sites and insert the necessary dataset records on AppDB.

*Next teleconference*: Thursday, September 10th, 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)