Workflows at the 3rd EUDAT conference - *Bringing Data Infrastructures to Horizon 2020*

24 & 25 September 2014, Amsterdam, the Netherlands

Highlights

- Legal aspects handled in a presentation from Pavel Kamoski. Explaining about the data privacy concept. Surprisingly it seems anonymisation of data is in itself already a possible illegal action
- We were told about the workflow plans in EUDAT 2020 WP8 especially in relation to Dynamic Data by Erhard Hinrichs
- Emanual Dima had a presentation explaining about the need and possibilities to shield the GEF from the deployment environment and use of a software called "Docker"
- Yann le Franc talked about aggregating data from multiple neuro-science repositories using semantic web-technologies

Overview

The session on Workflows reported on current results and future plans (for EUDAT2020) of the Workflow Working Group. The session thereby continued the activities and meetings of this working group that was formed in September 2013 and met regularly. The group now has more than 80 members both internal and external to EUDAT. The session featured four presentations:

1. Pawel Kamocki, legal expert at the IDS Mannheim and member of the CLARIN user community, gave a presentation about current European legislation on data privacy. Respecting data privacy and awareness of legal and ethical aspects of data privacy are crucial issues for workflows, since workflows involve data collection, data sharing, and data access and data processing in general.
2. Erhard Hinrichs, representative of the CLARIN community and EUDAT partner at the University of Tübingen, gave a presentation about future work on Data Life Cycle across Communities planned for EUDAT2020. The main objective of this research activity is to generalize existing technologies that allow construction of efficient workflows on a large scale -- with special emphasis on dynamic data. Dynamic data are an important type of data common among the user communities represented in EUDAT, including earth, environmental, social sciences and humanities. Therefore, the integration of dynamic data into workflows is a timely and significant extension of the portfolio of EUDAT services. The EUDAT workflow working group will take up the agenda of dynamic data in its working group activities in EUDAT2020. It will serve as a contact point for interaction with all interested EUDAT user communities interested in the incorporation of dynamic data into workflows.
3. Emanuel Dima, representative of the CLARIN community and EUDAT partner at the University of Tübingen, gave a presentation of the Generic Execution Framework (GEF) for workflows that has been developed within EUDAT. The EUDAT2020 research activity on Data Life Cycle across Communities will build on the GEF as an existing EUDAT service and will extend it in various directions so that it can handle dynamic data. Emanuel's presentation gave an overview of which extensions will be needed and how these extensions can be implemented on the basis of current technology.
4. Yann Le Franc, CEO and founder of e-Science Data Factory, gave the presentation "Aggregating data from multiple Neuroscience data repositories using semantic web tech: putting retina data into the Big Data perspective". Yann presented an interesting use case from neuroscience that demonstrates the challenges for workflows that rely on data from multiple data repositories and showed how some of the these issues can be handled using semantic web technology, a technology that we plan to deploy in EUDAT2020 more generally.