Workflows in EUDAT

reinhard.budich@mpimet.mpg.de
Erhard Hinrichs
Christian Page
Topics

- Intro
- The WFWG
- Brief History
- Presentations
- A debate?
- Activities/Next steps
- Conclusions
The WFWG

Reference: https://confluence.csc.fi/display/Eudat/Workflows

• Goals (from http://www.eudat.eu/workflows):
  – Well-described and documented scientific workflows ... becoming more and more important in all scientific disciplines ... due to ... the increasing amount of data ... and to increase the reproducibility of scientific results ... for raw data generated by sensors and software systems and ... for many areas of derived data - the long-tail data. ... institutions such as LANL and SDSC are also looking into such workflow systems ...
  – It is not yet fully clear which environment will be offered in these cases or exactly what types of services is data infrastructures should offer. EUDAT will continue to work with community experts to test service concepts that allow users to execute workflows on data stored in the EUDAT data domain.

• First Meeting in Barcelona Sep 13
• 4th meeting now
• ~ 80 member list of interested/active people
• Lately mostly active on EUDAT2020
Today

• This Workflows Session will focus on identifying EUDAT Services which could be useful for Workflows and Remote Execution.
• The first part will discuss data privacy issues with respect to the EUDAT Generic Execution Framework (GEF) and Metadata.
• The second part of the session will describe current and foreseen EUDAT Services, along with a first analysis on how these B2 Services could be useful for Workflows and Data Processing. A brainstorming will take place to deeply discuss the potential services to be selected for inclusion into the EUDAT Workflows API.

• EXPECTED SESSION OUTPUTS:
  – Wrap up findings from activities since second EUDAT conference
  – Update to work plan of the working group
  – Identify EUDAT B2 Services which should be candidate to be integrated within the GEF API in EUDAT2020: preparation of the technical report draft.
  – Discuss Data privacy issued within Workflows: identify problems and possible solutions.
Presentations

• Pawel Kamocki:
  – Protection of Data Privacy in EUDAT Workflows
• Erhard Hinrichs:
  – WP8@EUDAT2020
• Emanuel Dima:
  – Future GEF Versions
• Yann Le Franc:
  – The INCF Use Case
A Debate?

common workflows across communities

versus

sharing workflows inside communities

personal comm. Emanuel Dima, EKUT
Activities/Next steps

From the workflow working group plan v9.1

- Expand the Generic Execution Framework
  - Applicability to other communities: Report on activities and candidate communities/End Q4 2014
- Provide EUDAT Service APIs for use within Workflows
  - Identify EUDAT services candidates for use within workflows: Candidates List/Q4 2014
- Explore solutions for EUDAT Workflow Provenance Service(s)
  - Document “PNNL WFPS” /Check with PNNL, Q3 2014

1. Survey what PNNL has implemented for sharing components and describing workflows independently from concrete implementations
2. State if PNNL solution is adequate or not as a candidate for a EUDAT Workflow Provenance Service
3. Survey other institutions active in this field

More ideas

- Privacy data issues
- How do data federations dealing with privacy issues when designing workflows and interacting with metadata.
Conclusions