Workflows Workshop Barcelona Summary

EUDAT Workshop Days, Barcelona, 25.09.– 26.09.2013

Morris Riedel

Juelich Supercomputing Centre

Christian Page

CERFACS

...and many other participants

Outline

- **Workshop Goals**
- **Four Recommended Services**
    - EUDAT Service APIs for Workflows
    - Workflow Provenance Support Services
    - Higher-level Analysis/Analytics Services
    - Data Workflow Recommender Services
- **Next Steps**
- **Overview & Summary**

EUDAT

Workshop Goals Nutshell

- Understanding the needs of community experts in terms of common services in the area of *'workflow support'*
    - The focus was on common services supporting researchers who need to orchestrate data processing chains
    - E.g. Execute them on computers close to where their data is located.

- Roughly twenty workshop participants have been international experts in the field of 'scientific workflows'
    - They shared their insights and experiences towards the need of common services that EUDAT might be able to provide

- Goals: consensus of workshop participants on potential *'common workflow services elements'*

EUDAT Service APIs for Workflows

- Provide EUDAT Service APIs for use within Workflows

- Projects like EUDAT should not 'create new complete WFs tools'

- Instead provide service APIs for workflows:

- Enable researchers to seamlessly take advantage of current/new EUDAT services

  - E.g. data-staging, data-transfer, data replication, or simple store, PID assignments

Workflow Provenance Support Services

- Explore EUDAT Workflow Provenance Service(s)
  - An increasing variety of WF systems exists
- Many of the communities already have chosen their solutions,
  but might re-use components of others
- We could offer a Service that enables 'workflow component sharing'
  - Represents a repository/registry where components of workflows are stored including provenance information
  - E.g. assignments of PIDs for workflow components, including concrete software elements, information about concrete execution runs of it,
  - E.g. sample data that enables other researchers to better understand the shared workflow components

Higher-level Analysis/Analytics Services

- Provide higher-level Analysis & Analytics Workflow Components & Service APIs
- Presentations across all fields has shown that statistical computing, data mining, and machine learning algorithms are used in some parts of the workflows;
  - E.g. classification, clustering, or regression techniques
- A potential set of 'higher-level data analysis/analytics services' could be hosted by EUDAT close to the data of researchers.
  - Provisioning of service APIs for a seamless integration in (existing) analysis workflows)
  - Support 'application enabling process'

Data Workflow Recommender Services

- Investigate solutions for data workflow recommender services
  - Data formats are set by user communities
  - Limited amounts of standardization is having impact
- EUDAT could investigate the possibility of recommender services
  - E.g. provide advice on suitable workflows in context depending on data formats, scalability, portability, etc.
  - E.g. benchmarks of workflows in context
  - E.g. access to (captured) best practices in the community

Next Steps

- Need to continue the discussion in order to not lose the momentum

- Focus on some concrete work actions derived from the 'broad recommended actions.

- Forming a more structured working group was considered

  - starting with identifying overlaps with existing work from the experts

EUDAT

Overview and Summary

- The field of 'harmonizing security'
  across workflows was often mentioned
  - But it was recognized that this is a problem
    needs a broader approach
- Four identified recommended actions
  - Provide EUDAT Service APIs for use within Workflows
  - Explore solutions for EUDAT Workflow Provenance Service(s)
  - Provide higher-level Analysis & Analytics Workflow Components &
    Service APIs
  - Investigate solutions for data workflow recommender services
- Next steps identified to have a more narrow focus
  - Check with overlaps of work activities of experts in the field

EUDAT