HELmholtz Scientific Project WORkflow PlaTform
Status and Roadmap
December 2020
Oliver Knodel // contact: o.knodel@hzdr.de
The Challenge: An End-to-End Digital Data Lifecycle

— We support all stages with tools:
  • electronic lab books,
  • interactive analysis,
  • (automated) publication,
  • scientific workflow management.

— **Requirement**: Transfer data as early as possible to our data center.
HZDR Data Management Strategy

DMS Guidance System (Webfrontend)

- Proposal Management (GATE)
- Data Management Plan (RDMO)
- Device-Control
  - Facility-Control (Environment)
  - Detector-Control (Experiment)
- Data Source (Analyzer, Experiment, Simulation, ...)
- Lab Documentation
  - Knowledgebase (MediaWiki)
  - ELN (OpenBis)
  - Lims
- Workflow Engine
  - Version Control (GitLab)
- Compute (HPC/OpenStack)
- Publication
  - ROBIS
  - RODARE (Invenio)
- Data Transport Engine (Globus/Rucio)
- Archiving
- Elastic Search

Abstraction for all Raw/Metadata

Infrastructure

- Time Series Database
- Object, Parameter, Metadata Database
- Docker/VM Container/Images
- Compute Resources

Data Storage
HZDR Data Management Strategy
A Guidance System Should Enable a **Workflow on Project Level**

![DMS Guidance System (Webfrontend)](image)

Create DMS Project
Proposal Management
Create DMS Project
Data Management Plan
Configure/Monitor
Data Sources & Sinks
Configure/Monitor
Compute Workflows
Publication
Data Transport
Archive/Close
DMS Project

**Multiple iterations possible to enable high flexibility**
HELmholtz Scientific Project WORkflow PlaTform — Version 0.2.0

Project Workflow of ELBE ETL Workflow (HZDR.FWCC.2020.275336)

Computational/Scientific Workflow
Project Plan of ELBE ETL Workflow (HZDR.FWCC.2020.275336)
Scientific Workflow

Computational workflows describe the complex multi-step methods that are used for data collection, data preparation, analytics, predictive modelling, and simulation that lead to new data products.

Scientific Workflow

HELIPORT has an built-in Integration of Scientific Workflows
Scienﬁc Software Development and Workﬂows

— Analysis and Pre-/Postprocessing steps needs to be:
  • Documented and
  • Reproducible
— Capsuling every step in a
workﬂow adapts the FAIR principles.
Scientific Workflow Example

A scientific workflow must be:

• Comprehensibly
• Archivable,
• Reusable/Reproducible,
• Publishable,
• …
based on an open standard:

OWL or WDL

The description of the workflow itself...
Different Types of Publications
An Example for a Complete FAIR Research Experiment
Our Objective

— In all stages as much information as possible is stored in our HELIPORT PostgresDB.
— Metadata is used to transfer the information between different (unknown) systems.
— HELIPORT is designed to export metadata for known schemes.
— Every published object should have describing metadata in an open and widely used format to be
HELIPORT Roadmap

First Draft: Project Plan (August 2020)
- Project and user management
- Configurable stages
- REST API for proposal information
- CWL visualization prototype

Modular Structure (February 2021)
- Subdivision of the stages into modular and configurable Django apps to allow individual extensions
- Refactoring of the project
- Documentation using GitLab pages

Documentation of a TELBE user experiment*
- Integration of all related data sources
- Automated workflow initiation
- Publication of all data products

Initial Version (June 2020)
- Webinterface with user authentication (LDAP)
- DMS Projects and additional proposal information from the HZDR GATE database

Improved Project Plan (December 2020)
- Fully configurable stages and modules
- Infrastructure and database updates
- Daily proposal database update
- CI pipeline for test and deployment
- Advanced logging and monitoring

Integration of various Apps and Features*
- Integration of (different) Metadata Schemas
- Computational/Scientific workflow execution
  - Workflow management and monitoring
  - CWL support
- Data Management Plan support
- (Global) Handle management

*Depending on the pending approval of our HMC proposal!
Modular HELIPORT Design — Draft for Version 0.4.0

**Infrastructure**
- HZDR Gate
- HELIPORT HZDR Theme
- Data Management Plan (RDMO)

**Related Systems**
- Version Control
- OpenBis
- MediaWiki

**Project Files and Data**
- Data Sources and Tools
- CWL Workflows

**Scientific Workflow**
- CWL Management
- Slurm Integration

**Publication**
- Robis
- Rodare
- Internal Archive

**Settings.py**

**Django Apps**
HELIPORT Architecture Concept

- **Scientific Project Workflow Frontend**
- **Digital Object Manager**
- **Message Broker (RabbitMQ)**
- **Message Queue (Celery)**
- **Hemera**
- **Gate Interface**
- **Postgres DB**
- **DMS App**

Asynchron Database Synchronization with Celery Beat
HELIPORT

HELmholtz ScIentific
Project WORkflow PlaTform

https://vlsdms.fz-rossendorf.de