Long Term Interoperability of Distributed Research Data Infrastructures

Conference on Research Data Infrastructure
CoRDI | September 12-14, 2023 | Karlsruhe, Germany

Marius Politze, Yusra Shakeel, Sirieam Hunke, Philipp Ost, Rossella Aversa,
Benedikt Heinrichs, Ilona Lang
Motivation

Huge variety of highly complex data present in silos as various repositories, cloud storage, local storage, etc.

→ Accessing data in silos can be quite challenging
→ Lack of a harmonized representation of data and metadata for research data management

Our goal:
To achieve “Long Term Interoperability” of Distributed Research Data Infrastructures

Apply the FAIR Digital Object concept to interconnect infrastructure components
A FDO is a unit of data, represented as a sequence of bits that binds all critical information about an entity in one place and creates a new kind of actionable, meaningful and technology-independent object.

- Ability to connect different build systems to share data
- Achieved by applying the FDO concept:
  - Bridges between data repositories, disciplines, etc.
  - Applies all aspects of the FAIR principles
  - Technology agnostic
  - Implementation:
    - Handle Persistent Identifier (PID)
    - Information Record = DO’s Metadata

Source: https://fairdo.org/
NFDI-MatWerk Shared Service Architecture

Legend:
- Architecture layers
- Architecture sub-layers
- Surrounding services & external support
- Broader categories representing the architecture layers

(Additional contributors: Tilmann Hickel, Angelika Gedsun)
NFDI-MatWerk Shared Service Architecture

- Data services store the bit sequences as human interpretable content
- Metadata services complement by storing the corresponding metadata of the data stored
- Both external and internal resources are supported
**NFDI-MatWerk Shared Service Architecture**

<table>
<thead>
<tr>
<th>Development Environment</th>
<th>Provisioning Environment</th>
<th>IT Infrastructure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge Graph Representation, Tools &amp; Standards</td>
<td>Data &amp; Metadata Services</td>
<td></td>
</tr>
<tr>
<td>Access KG</td>
<td>Internal Resources</td>
<td></td>
</tr>
<tr>
<td>Interact with KG</td>
<td>External Resources</td>
<td></td>
</tr>
<tr>
<td>Validate KG</td>
<td>Tools for Services</td>
<td></td>
</tr>
<tr>
<td>Manage KG</td>
<td>Tools for Managing Data</td>
<td></td>
</tr>
</tbody>
</table>

- Stored data is represented as FDO, using services such as, Typed PID Maker and FDO-scope
- Workflow environment supports software solutions to share data, metadata and workflows, ensuring their long-term availability
- FDOs with their interconnections can be represented and explored as KG
Researchers can build on the workflows to realize applications relevant for their individual discipline.

Interfaces, such as ELNs can be developed for scientists to interact with.
Working Example 1

Scientists’ Applications & Workflows
- idCARL / OpenBIS

Knowledge Graph Representation, Tools & Standards
- Access KGs
- Interact with KGs
  - Data Collections Explorer
  - SPARQL

Workflow Environment
- Visual Programming in Chaldene (Jupyter Lab)

Data Representation: FAIR Digital Objects
- Tools for Services
  - FDO Lab
  - Typed PID Maker
  - FDO-scope

Data & Metadata Services
- Internal Resources
- MatWerk Metadata Repository
- External Resources
  - Zenodo

IT Infrastructure

Working Example II

- Scientists’ Applications & Workflows
  - idCARL

- Knowledge Graph Representation, Tools & Standards
  - Validate KGs
    - SCHACL
  - Interact with KGs
    - LDP Coscine
  - Manage KGs
    - SPARQL
  - Form Generator

- Workflow Environment
  - Pyron (communication)

- Data Representation: FAIR Digital Objects
  - Tools for Services
    - Coscine PIDs
  - PID Record

- Data & Metadata Services
  - Internal Resources
    - AIMS
    - RDS Coscine

- IT Infrastructure

Harmonization
Data, metadata, formats

Support
Participant projects, documentation
Wrap-up

- Efficiently manage their data
- Adhere to a standard for handling their research data
- To adapt the proposed architecture for other NFDI communities
- Combine data from different resources through interoperability
- Use existing services without worrying about the technical details of the backend processes

FDO Concept → Service Architecture → Researchers

Enables
Thank you!
FAIR DO can change the way we approach datasets, offering a clear and structured method of accessing and interpreting data using Persistent Identifiers (PIDs), ultimately enhancing the usability and interoperability of data by machines.