1st Conference on Research Data Infrastructure

Connecting RDM

https://doi.org/10.52825/CoRDI.v1i.348

Published: 07 Sept. 2023

Long Term Interoperability of Distributed Research Data Infrastructures

Marius Politze[1] [https://orcid.org/0000-0003-3175-0659] , Yusra Shakeel[2] [https://orcid.org/0000-0001-5135-4325] , Sirieam Hunke[1] [https://orcid.org/0000-0001-9316-4220] , Philipp Ost[2] [https://orcid.org/0000-0002-7198-0566] , Rossella Aversa[2] [https://orcid.org/0000-0003-2534-0063] , Benedikt Heinrichs[1] [https://orcid.org/0000-0003-3309-5985] , and Ilona Lang[1] [https://orcid.org/0000-0002-7202-5982]

[1]RWTH Aachen University, Germany

[2]Karlsruher Institut für Technologie, Germany

Abstract:

Research institutions have established a variety of research data infrastructures that orient towards discipline or methodology specific needs of their respective research community. Technically, these infrastructures ultimately are based on off-the-shelf hardware and software building blocks – both commercial and open-source. While such enterprise ready infrastructures can scale well, they apt to data silos and typically do not adhere to scientific standards like the FAIR (Findability, Accessibility, Interoperability, Reusability) principles. Using common architecture concepts such as the FAIR Digital Object (FAIR DO) allows interconnection of these silos by adding a long term interoperability layer on top of the existing infrastructure components.

A FAIR DO is a digital representation of data as a sequence of bits, identified by a globally-unique, persistent, and resolvable identifier, described by an information record, and classified by a type. This interoperability layer on top of the existing infrastructure components allows distributed data resources to be connected and related, regardless of where they are stored.

For this purpose, we propose a shared service architecture that allows the implementation of discipline specific services and applications. The data services enable storing and processing of the data, complemented by the metadata services that describe this stored data. While separate aspects of a digital object may be stored in specialized systems like databases or object storages, a data representation layer ties these aspects together as an entity, i.e. the FAIR Digital Object. The entirety of these digital objects with their interconnections can then be represented and explored as a knowledge graph that supports specific workflows for instance machine learning or quality assurance. Finally, researchers can build on top of these workflows to realize applications relevant for their individual discipline and develop interfaces, such as connected Electronic Lab Notebooks (ELNs). These technical infrastructure need to be framed with discipline specific support, teaching and consulting activities, data governance services and harmonized agreements on interfaces, data, and metadata formats. Figure 1 gives a schematic overview of such an architecture incorporating an infrastructure and software developer's perspective.

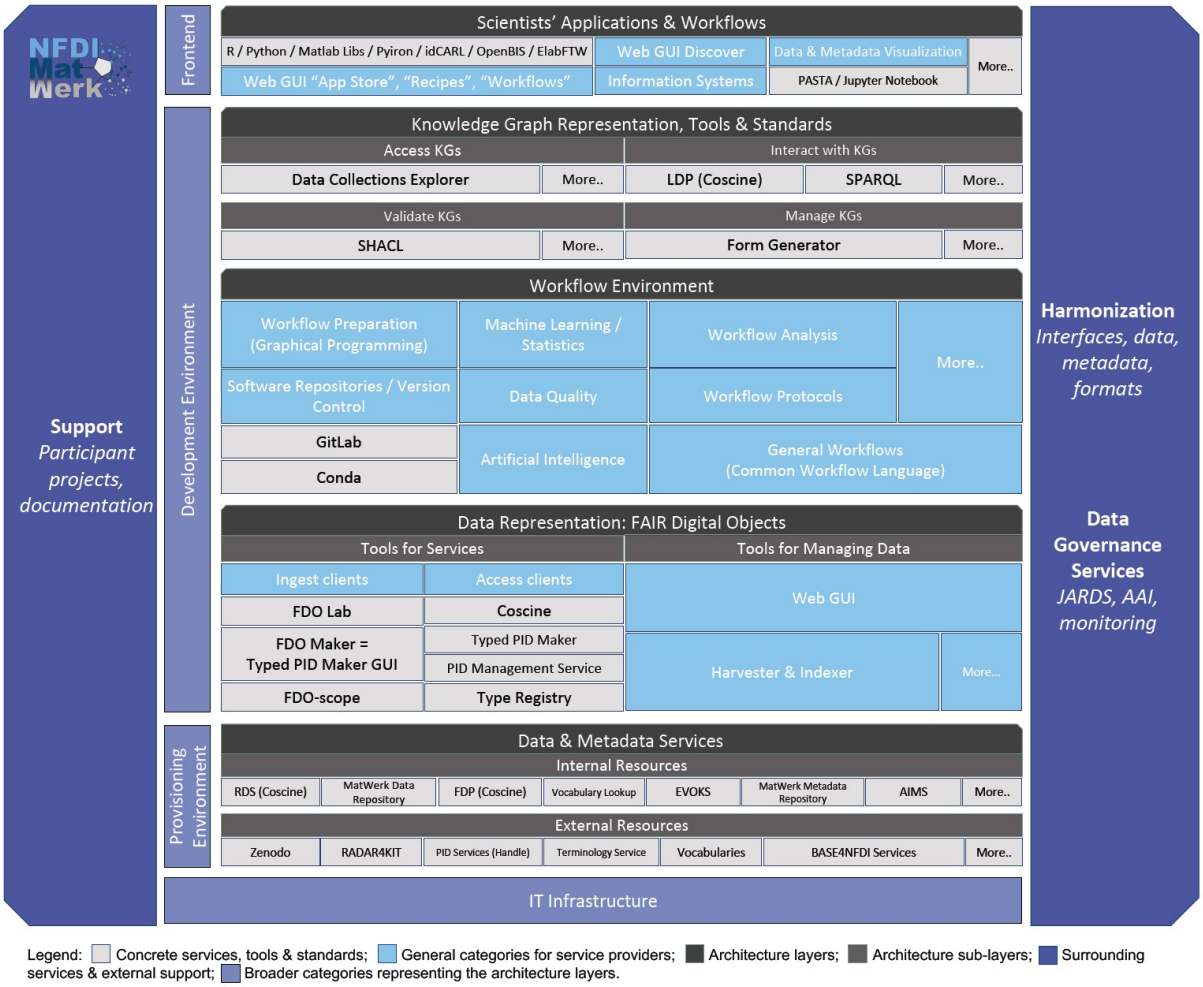

Figure 1. Schematic Overview of the envisioned architecture and its software components.

Based on a requirements analysis and researchers' needs within the NFDI-MatWerk consortium, we present a set of collectively working examples (integrated platforms: Coscine and MatWerk Data Repository, creation: FDO Maker, exploration: FAIR-DOscope) of this approach based on the concept of FAIR DO as recommended by the RDA and the European Commission. A combination of service offers and installable applications allows implementing the FAIR principles within the existing research data infrastructures. The presented results are widely based on W3C standards and are currently being evaluated in the context of the NFDI-MatWerk consortium. After that, the results will be transferred to other NFDI consortia, including NFDI4Ing, NFDI4Chem, and NFDI4Microbiota.

The approach provides a practical solution for interconnecting distributed research data infrastructures to national (like NFDI) and international (like EOSC and Gaia-X) infrastructures and preventing the creation of data silos. By allowing existing data infrastructures to make data FAIR, we enable researchers to access and reuse data from different domains, facilitating cross-disciplinary research and advancing new methods for scientific discoveries.

Contributions

- Marius Politze — Project administration, Supervision, Writing – original draft
- Yusra Shakeel — Investigation, Visualization, Software, Writing – review & editing
- Sirieam Hunke — Investigation, Visualization, Software, Writing – review & editing
- Philipp Ost — Conceptualization, Writing – review & editing
- Rossella Aversa — Project administration, Supervision, Writing – review & editing
- Benedikt Heinrichs — Software, Writing – review & editing
- Ilona Lang — Conceptualization, Investigation

The work was supported with resources granted by NFDI-MatWerk, funded by Deutsche Forschungsgemeinschaft (DFG) with project number 460247524.

Keywords: FAIR, FAIR Digital Object, Service Architecture, Standardization