



1ST BASE4NFDI USER CONFERENCE



NOV 20-21,
2024



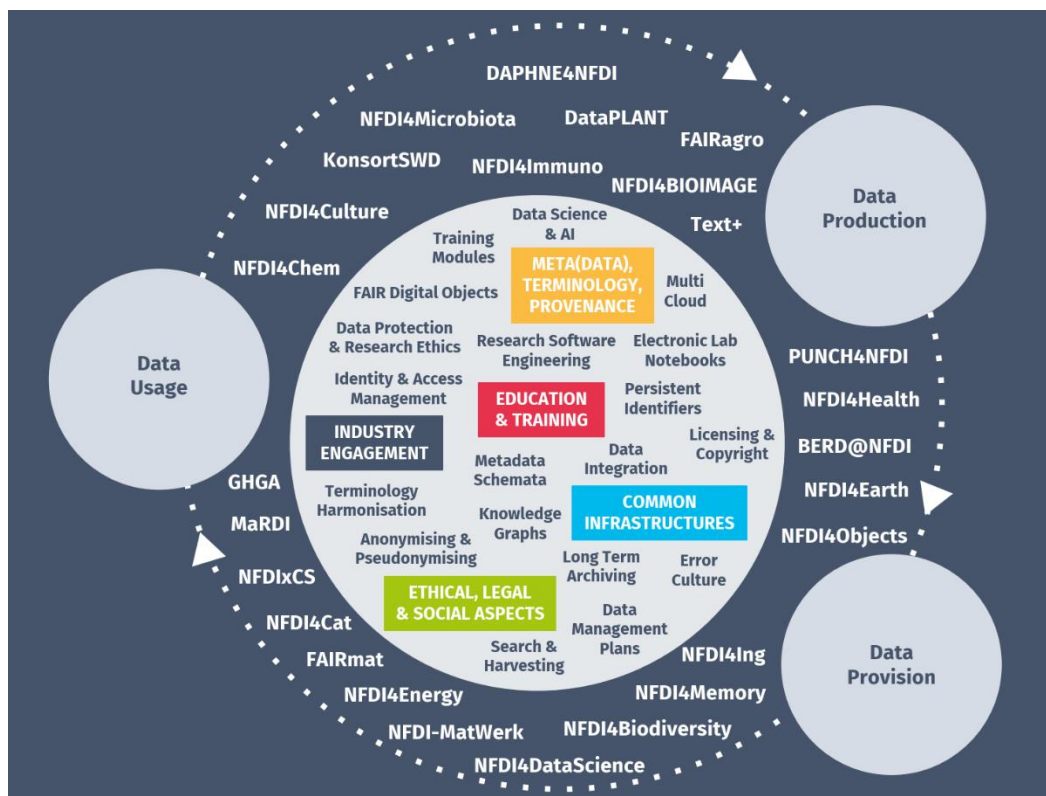
BERLIN

For those interested in actively using & shaping future basic services of the NFDI.

Proceedings

What is Base4NFDI?

Today, research data is often stored in many different places, difficult to find and only available for a limited time. The NFDI aims towards better findability, accessibility and reusability of research data. Base4NFDI integrates and establishes basic services as common, interoperable solutions. Already existing services are adapted or extended to be usable for researchers from other disciplines. This way, parallel developments are avoided - since many scientific fields have similar requirements for a number of research data management services.



How does it work?

All institutions in the scientific community (and beyond) can actively participate in the development and implementation of these services and submit proposals via the . A panel of experts selects the proposals of the development teams for Base4NFDI, which then receive funding from the German Research Foundation (DFG). These teams also receive support and guidance in areas such as development, implementation, and training. After development, the NFDI basic services will be offered permanently to the scientific community. Base4NFDI thus actively contributes to the systematic opening and networking of the German science system.

Keynote

EPOS, a community-driven research infrastructure for open access to interdisciplinary data and services from solid Earth science

Lilli Freda and Daniele Bailo (both National Institute of Geophysics and Volcanology, EPOS ERIC)

Mi, 20.11., 13:00 – 14:00 CET



*Lilli Freda
(INGV, EPOS-ERIC Executive Director)*

The European Plate Observing System (EPOS, <https://www.epos-eu.org/>) pan-European Research Infrastructure (RI) enables interdisciplinary collaboration by providing integrated and open access to high-quality data, services, and computational resources in solid Earth science. The EPOS RI is operated by a European Research Infrastructure Consortium (ERIC) and adopts a federated governance model to ensure the coordination of its heterogeneous distributed resources. The challenge of long-term sustainability is tackled by addressing technical, legal, governance and financial dimensions in a synergic way.

The EPOS Platform, in compliance with FAIR principles, acts as a multidisciplinary gateway, enabling seamless access to harmonized data and services from almost 300 providers from 26 European countries. The Platform relies on a federated architecture, rich metadata, and service-driven approach. EPOS overcomes challenges related to data interoperability and standardization and provides a unique point of access to the wealth of resources in the solid Earth domain. This fosters excellent science and contributes to address societal challenges such as natural hazard risk assessment and environmental sustainability. Available as an open-source software package, the Platform envisages to include data analysis, processing, and advanced visualization tools.



*Daniele Bailo
(INGV, EPOS-ERIC IT Officer)*

Conference Program

Program Wednesday, 20.11.2024

Location for all events: Veranstaltungszentrum, Fraunhofer FOKUS

12:00 – 13:00 CET	Lunch
13:00 – 14:00 CET	<p>Welcome <i>Manfred Hauswirth (Fraunhofer FOKUS)</i> <i>Sonja Schimmeler (Fraunhofer FOKUS)</i></p> <p>Keynote: EPOS, a community-driven research infrastructure for open access to interdisciplinary data and services from solid Earth science <i>Lilli Freda and Daniele Bairo (EPOS ERIC)</i></p> <p><i>Session details: see p.4</i></p>
14:00 – 14:30 CET	<p>Poster Minute Madness <i>Chair: Lukas Weimer (SUB Göttingen)</i></p> <p><i>Session details: see p. 7</i></p>
14:30 – 15:15 CET	Coffee break (posters displayed)
15:15 – 16:45 CET	<p>Basic services in initialisation phase: Jupyter4NFDI, DMP4NFDI, KGI4NFDI, nfdi.software <i>Chair: Raphael Ritz (MPCDF)</i></p> <p><i>Session details: see p.9</i></p>
16:45 – 17:30 CET	Coffee break (posters displayed)
17:30 – 19:00 CET	<p>Further fields <i>Chair: Reinhard Altenhöner (Staatsbibliothek zu Berlin – PK)</i></p> <p><i>Session details: see p.9</i></p>
19:00 – 21:00 CET	Reception (posters still displayed)

Program Thursday, 21.11.2024

Location for all events: Veranstaltungszentrum, Fraunhofer FOKUS

09:00 – 10:30 CET	Basic services in integration phase: IAM4NFDI, TS4NFDI, PID4NFDI <i>Chair: Raphael Ritz (MPCDF)</i> <i>Session details: see p. 10</i>
10:30 – 11:15 CET	Coffee break (posters displayed)
11:15 – 12:45 CET	Incoming services <i>Chair: Reinhard Altenhöner (Staatsbibliothek zu Berlin – PK)</i> <i>Session details: see p. 10</i>
12:45 – 13:00 CET	Closing <i>Sonja Schimmmler (Fraunhofer FOKUS)</i>
13:00 – 14:00 CET	Lunch

Poster Minute Madness

#30	The NFDI Data Stewardship Network <i>Daniela Hausen (RWTH Aachen), Jens Dierkes (Cologne University and City Library), Cord Wiljes (NFDI)</i>
#23	NFDI4Cat Research Data Management System: A Proposal for Enhanced Data Sharing in Catalysis Research <i>Nadiia Huskova, Taras Petrenko, Vuliia Dikova, Volodymyr Kushnarenko, Preston Rodrigues, Thomas Bönisch (all HLRS Stuttgart)</i>
#35	Development of Sustainable Aggregating Services for an Improved FAIR Research Software Ecosystem <i>Maxence Azzouz-Thuderoz, Madhurima Deb, Moritz Schubotz (all FIZ Karlsruhe)</i>
#39	Overlaps between NFDI Knowledge Graphs <i>Daniel Mietchen (FIZ Karlsruhe)</i>
#40	A terminology service for BERD@NFDI and its knowledge graph infrastructure via TS4NFDI <i>Renat Shigapov (Mannheim University), Oliver Koepler (TIB), Roman Baum (ZB MED)</i>
#41	Facilitating Comprehensive Metadata Capture and Validation in Data Repositories (nmrXiv) through Terminologies and Terminology Service Suite Widgets <i>Venkata Nainala, Noura Rayya, Christoph Steinbeck (all Jena University), Oliver Koepler (TIB)</i>
#43	Basic services to the community – facilitating the application of basic services through interactive case study textbooks <i>Melanie Seltsmann (HU Berlin), Evgenia Samoilova (Potsdam University), Hannes Schnaitter, Philipp Schneider (both HU Berlin), Maria Chlastak (GI), Ulrike Lucke (Potsdam University), Sonja Schimmler (TU Berlin and Fraunhofer FOKUS)</i>
#44	Software repository for the NFDI <i>Dirk von Suchodoletz (University of Freiburg), Timo Mühlhaus, Christoph Garth (both RPTU Kaiserslautern), Björn Usadel (HHU Düsseldorf)</i>
#45	Cloud based flexible service infrastructure stack for the NFDI <i>Jonathan Bauer, Marcel Tschöpe, Paul Chr. Schnürle, Julian Weidhase (all University of Freiburg), Christian Garth, Timo Mühlhaus (both RPTU Kaiserslautern)</i>

#27	IAM4NFDI – JARDS Incubator <i>Marius Politze, Marcel Nellesen, Camilla Lummerzheim (all RWTH Aachen)</i>
#42	BITS - a Use Case for Terminologies in Earth System Sciences <i>Claudia Martens (DKRZ), Anette Ganske (TIB), Alexander Wolodkin (SGN)</i>
#49	Semantic linked data and metadata storage made easy with LARAsuite <i>Mark Dörr (University Greifswald)</i>
#51	Base4NFDI in the German National Research Data Infrastructure (NFDI) e.V. Association <i>Cord Wiljes, Stefanie Fuchsloch, Lukas Jansen, Anne Lehmann, Hendrik Seitz-Moskaliuk (all NFDI)</i>
#08	A Cloud-Native Virtual Bioimage Analysis Research Desktop (BARD) for Deployment of Containerised Bioimage Tools on Kubernetes <i>Yi Sun, Christian Tischer, Jean-Karim Hériché (EMBL)</i>
#11	Enhancing FAIR Research Data Management: advancing fine-grained data citation by PIDs for dataset elements <i>Janete Saldanha Bach, Peter Mutschke (both GESIS)</i>
#12	NFDI4Objects and the Base4NFDI Services: An Overview <i>Fabian Fricke (DAI), Florian Thiery (LEIZA)</i>
#14	Persistent Identifier (PID) a journey of making data machine-actionable. <i>Preston Rodrigues (HLRS Stuttgart), David Linke, Thomas Bönisch (both LIKAT)</i>
#17	Open source Virtual Desktop Infrastructure solution for remote visualization and training in the NFDI <i>Armin Saur, Michael Scherle, Dirk von Suchodoletz, Rafael Gieschke (all University of Freiburg), Yi Sun, Jean-Karim Hériché (both EMBL)</i>
#22	Infrastructure for Research Based Learning and the NFDI Basic Service Jupyter: Searching for Opportunities for Cooperation and Areas of Differentiation <i>Jan F. Krämer, Lilian Löwenau (both HU Berlin)</i>
#47	NFDI4Cat Web Application for User-Friendly Generation and Browsing of Catalysis Metadata within a Semantically Rich RDF Framework <i>Taras Petrenko, Yuliia Dikova, Nadiia Huskova, Volodymyr Kushnarenko, Preston Rodrigues, Thomas Bönisch (all HLRS Stuttgart)</i>

Sessions

Basic Services in Initialization Phase

Chair: Raphael Ritz (MPCDF)

10	DMP4NFDI - a Basic Service for Data Management Planning <i>Jürgen Windeck (ULB Darmstadt)</i>
18	Jupyter4NFDI: a central JupyterHub providing access to various software stacks and computing resources across the NFDI consortia <i>Tim Kreuzer (Forschungszentrum Jülich GmbH)</i>
38	Reproducibility assessment as a service: a knowledge graph about the reproducibility of Jupyter notebooks <i>Daniel Mietchen (FIZ Karlsruhe & NFDI/MaRDI)</i>
9	Facilitating creation, (re)use and interoperability for Knowledge Graphs in NFDI: the vision behind KGI4NFDI <i>Dr Lozana Rossenova (TIB – Leibniz Information Centre for Science and Technology, Hannover), Prof. Konrad U. Förstner (ZB MED - Information Centre for Life Sciences)</i>
20	A Knowledge Graph-Based Data Integration Workflow for NFDI4Culture and Beyond <i>Jonatan Jalle Steller (Academy of Sciences and Literature Mainz)</i>
25	nfdi.software - unlock software, activate research data <i>Martin Hammitzsch (Helmholtz-Zentrum Potsdam Deutsches GeoForschungsZentrum GFZ)</i>

Further fields

Chair: Reinhard Altenhöner (Staatsbibliothek – PK)

15	The Research Commons approach – an opportunity for Base4NFDI? <i>Tilo Mathes (Research Space)</i>
19	Research Software Engineering within the NFDI (INFRA-WG-RSE) <i>Bernd Flemisch (Universität Stuttgart), Martin Hammitzsch, Florian Thiery (LEIZA)</i>
26	Challenges, Opportunities and Experiences of setting up sustainable Research Data Infrastructure <i>Andreas Ganzenmüller (HLRS)</i>
7	Beyond lifetimes: what does archiving mean for research data? <i>Dr Kai Naumann (Landesarchiv Baden-Württemberg)</i>

21	The FID Middle East-, North Africa- and Islamic Studies – as a user and partner of base4nfdi <i>Josef Peter Jeschke (GNOI)</i>
34	Software metadata for BASE4NFDI services <i>Dhwani Solanki (ZB MED Information Centre for Life Sciences)</i>
36	Cooperation in Research Data Management across Base4NFDI, regional RDM initiatives, and beyond – considerations from SaxFDM <i>Moritz Kurzweil (Leibniz-Institut für Geschichte und Kultur des östlichen Europa (GWZO))</i>

Basic services in integration phase: IAM4NFDI, TS4NFDI, PID4NFDI

Chair: Raphael Ritz (MPCDF)

16	IAM4NFDI – The Community Experience <i>Dr Marius Politze (RWTH Aachen University)</i>
24	Polished Policies for the NFDI-AAI <i>Marcus Hardt (KIT), Wolfgang Pempe (DFN-Verein)</i>
37	Experiences and status quo of IAM / CAAI in DataPLANT <i>Dirk von Suchodoletz (GNOI)</i>
6	There will be FAIR. Insights into the PID landscape analysis <i>Torsten Kahlert (German National Library of Science and Technology (TIB) - Leibniz Information Centre for Science and Technology)</i>
46	First steps of the basic service Terminology Services 4 NFDI - Step by step through the initialization to the integration phase <i>Roman Baum (ZB MED - Information Centre for Life Sciences)</i>
32	Terminology service for research data management and knowledge discovery in low-temperature plasma physics <i>Markus M. Becker (Leibniz Institute for Plasma Science and Technology (INP))</i>

Incoming services

Chair: Reinhard Altenhöner (Staatsbibliothek – PK)

31	The NFDICore Ontology And Related Modular Domain Ontologies For NFDI4Culture - NFDI-MatWerk - NFDI4DataScience - NFDI4Memory And Beyond <i>Tabea Tietz (FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur)</i>
----	--

48	A Platform for the Provision of Teaching and Training Contents in Data Literacy: DALIA4NFDI Jochen Ortmeyer (RWTH Aachen University, Institute of Inorganic Chemistry)
5	Meet the digital data steward Jarves Jonas Werheid (WZL of the RWTH Aachen University), Tobias Hamann (WZL RWTH Aachen University Intelligence in Quality Sensing (IQS))
13	ASSURED. A training and accreditation service for safe research Deborah Wiltshire (GESIS Köln), Simon Parker (DKFZ Heidelberg), Vanessa González Ribao (DKFZ Heidelberg)
33	Connecting the RDM Helpdesks by a shared Basic Framework Judith Engel (Universität Bremen - MARUM (NFDI4Biodiversity)), Martha Stellmacher (SLUB Dresden/NFDI4Culture)
4	FAIR principles and benchmarking of research software as essential data Dr Jörg F. Unger (Bundesanstalt für Materialforschung und -prüfung)
29	RDMTraining4NFDI - Consolidating Training of Research Data Management Skills in the NFDI Konrad Förstner (ZB MED - Information Centre for Life Science)

Abstracts

Abstracts start on next page

Basic Services in Initialization Phase

Chair: Raphael Ritz (MPCDF)

DMP4NFDI - a Basic Service for Data Management Planning

Authors:

Jürgen Windeck^{1*}, Katja Diederichs², Sabine Schönau³, and David Wallace⁴

*Lead presenter

¹ Juergen.Windeck@tu-darmstadt.de, TU Darmstadt

² ZB MED

³ RWTH Aachen University Library

⁴ TU Darmstadt

Abstract: [max. 500 words]

DMP4NFDI is a Basic Service to support consortia providing data management plan (DMP) services or analogues such as software management plan (SMP) services to their respective communities. DMP4NFDI offers RDMO tool hosting for the consortia, support and training for the creation of DMP templates, and coordinates the standardization by developing a DMP template framework. Its main goal is to implement machine-actionable DMPs across the NFDI.

The NFDI aims to establish processes for standardized data handling, quality assurance, and sustainable and interoperable research data management. DMPs can be an aid to achieve these goals and foster the necessary cultural change, if based on machine-actionable, community-curated DMP templates and if standardized by a common framework within the NFDI. DMP4NFDI aims to establish such a standardized framework utilizing the DMP tool *Research Data Management Organiser (RDMO)* in a centralized multi-tenant hosting environment for consortia-wide use. We address identified shortcomings of current DMP usage and implementations, e.g. the lack of discipline-specificity, standardized answer sets, controlled vocabularies, and automation processes. The envisioned role of the service includes supporting NFDI consortia in making the DMP information more user-friendly and seamlessly integrated into the research process.

During its initialization phase, DMP4NFDI cooperates with several consortia as early adopters to refine requirements and prototype service integrations. We set up RDMO clients and prototype service integrations to other Basis Services or specific communities. These clients can be tailored according to the community needs, e.g. theming, plugins and templates, while still being connected to the overall NFDI framework. We will establish an editorial board for this framework to ensure coherence and interoperability of DMP templates across disciplines. In addition, DMP4NFDI offers support and training for consortia to create DMP templates and services.

The Basic Service is largely shaped by previously gained knowledge from workshops, use cases and stakeholder interactions, which will be adapted through further exchange with consortia over the course of the first year. In subsequent phases of the Basic Service, we will add more interested consortia, accompanied by further standardization and integration of existing services.

Keywords: [max. 5 keywords]

Data Management Plan, Basic Service, DMP Templates, Standardization

Jupyter4NFDI: a central JupyterHub providing access to various software stacks and computing resources across the NFDI consortia

Authors:

Tim Kreuzer^{1*}, Björn Hagemeier¹, Bernd Flemisch², Arnim Bleier³, Matthias Lieber⁴, George Dogaru⁵ and Klaus Reuter⁶

*Lead presenter

¹t.kreuzer@fz-juelich.de, Forschungszentrum Jülich GmbH

²University of Stuttgart

³GESIS – Leibniz-Institute for the Social Sciences

⁴Technische Universität Dresden

⁵Gesellschaft für wissenschaftliche Datenverarbeitung mbH (GWDG)

⁶Max Planck Computing & Data Facility (MPCDF)

Abstract:

The widespread usage of Jupyter notebooks across scientific disciplines highlights their fundamental importance in current research practices. However, their deployment across various NFDI consortia currently occurs through individual JupyterHubs, resulting in access barriers to computational and data resources. Only some of the services are widely available, and others are barricaded within VPNs, creating a fragmented landscape that complicates access. Our base service Jupyter4NFDI aims to unify these efforts by offering a centralized service. We simplify access, significantly improve the user experience, and extend the reach of Jupyter to a broader audience within the NFDI and beyond. We integrate with IAM4NFDI to govern access to the central service and external resources, with future integrations planned with upcoming basic services such as Multi-Cloud or nfdi.software. Moreover, HPC resources from the GCS and the NHR Alliance will be accessible via the centralized Jupyter service, which lowers HPC entry barriers for a wide range of users. Importantly, the centralized system not only simplifies access but also supports the import of projects along with their necessary dependencies, fostering an ecosystem conducive to creating binder-like 1-click reproducible FAIR Digital Objects (FDOs), possibly along with notebook identifiers supported by PID4NFDI. By ensuring a consistent user experience across all consortia, we promote continued innovation. Our commitment is to a collaborative and centralized approach that brings Jupyter notebooks within everyone's reach and democratizes access to the infrastructures for reproducible computational research within the NFDI and beyond.

The project's initialization phase has started in June 2024. During this phase, we build up a central base service, initially hosted as part of the already existing JupyterHub environment at JSC, which has been providing Jupyter based access to high-performance computing resources for over seven years by now and is continuously improved and extended.

We will present an update on the current status of the project and the intended next steps, including an outlook for the next project phases.

Keywords: Jupyter, Interactive Computing, HPC, NHR

Reproducibility assessment as a service: a knowledge graph about the reproducibility of Jupyter notebooks

Authors:

Daniel Mietchen^{1*} and Sheeba Samuel²

*Lead presenter

¹daniel.mietchen@mardi4.de, FIZ Karlsruhe — Leibniz Institute for Information Infrastructure

²Technical University of Chemnitz

Abstract:

Reproducibility has many facets, including computational ones. Assessing reproducibility at scale requires consistent workflows. In this context, the Basic Service Jupyter4NFDI offers a unique opportunity for standardization in terms of infrastructure, assessment criteria, documentation and training. In this contribution, we will describe how we used scalable workflows for automatically assessing key aspects of the computational reproducibility of Python-based Jupyter notebooks associated with research publications [1]. While we have focused on publications from the biomedical domain, the approach does not require an association with manuscripts and is transferable to other domains.

Our dataset has detailed metadata about publications, about associated code repositories and about Jupyter notebooks and their reproducibility. We have made it available in its original format as a SQLite database [2] and in addition converted it into a knowledge graph [3]. This knowledge graph can be explored in detail by way of SPARQL queries [4], of which we provide a set of examples catering to multiple use cases. These use cases include the documentation of common problems regarding the reproducibility as well as training contexts in which students or instructors might want to find real-world examples of notebooks from a particular research field and being either fully reproducible or failing in some specific way, e.g. with a given error message relevant for a training session.

Such a conversion of a rich dataset from its original format into a knowledge graph could also be an interesting use case for the Basic Service KGI4NFDI. We would be particularly interested in exploring how these two Base Services could interact to facilitate (1) routine reproducibility assessments of Jupyter notebooks relevant in NFDI contexts, (2) feeding such a knowledge graph with information about these reproducibility assessments, (3) documentation of common issues and best practices and (4) training for Jupyter, reproducibility, knowledge graphs and other facets of Base Services, like terminology or persistent identifiers.

References:

- [1] Sheeba Samuel, Daniel Mietchen, Computational reproducibility of Jupyter notebooks from biomedical publications, GigaScience, Volume 13, 2024, giad113, <https://doi.org/10.1093/gigascience/giad113>
- [2] Sheeba Samuel, & Daniel Mietchen. (2023). Dataset of a Study of Computational reproducibility of Jupyter notebooks from biomedical publications [Data set]. In GigaScience (Vol. 13). Zenodo. <https://doi.org/10.5281/zenodo.8226725>
- [3] Sheeba Samuel, Daniel Mietchen, FAIR Jupyter: a knowledge graph approach to semantic sharing and granular exploration of a computational notebook reproducibility dataset (2024), arXiv preprint arXiv:2404.12935, <https://doi.org/10.48550/arXiv.2404.12935>
- [4] <https://w3id.org/fairjupyter>

Keywords: Jupyter notebooks, Python, computational reproducibility, knowledge graphs

Facilitating creation, (re)use and interoperability for Knowledge Graphs in NFDI: the vision behind KGI4NFDI

Authors:

Lozana Rossenova^{1*}, Renat Shigapov², Moritz Schubotz³, Fidan Limani⁴, Benjamin Zapilko⁵, Till Sauerwein⁶, and Konrad U. Förstner⁷

*Lead presenter

¹ lozana.rossenova@tib.eu, TIB – Leibniz Information Centre for Science and Technology, Hannover

² renat.shigapov@uni-mannheim.de, Mannheim University Library, Mannheim

³ Moritz.Schubotz@fiz-Karlsruhe.de, FIZ Karlsruhe – Leibniz Institute for Information Infrastructure, Berlin

⁴ flimani@zbw-workspace.eu, ZBW – Leibniz Information Centre for Economics, Kiel

⁵ Benjamin.Zapilko@gesis.org, GESIS - Leibniz Institute for the Social Sciences, Cologne

⁶ sauerwein@zbmed.de, ZB MED – Information Centre for Life Sciences, Cologne

⁷ foerstner@zbmed.de, ZB MED – Information Centre for Life Sciences, Cologne

Abstract:

A graph-structured knowledge base known as a Knowledge Graph (KG) consists of a terminology (vocabulary or ontology) along with interconnected data entities, all based on semantic web technologies like RDF and SPARQL. Knowledge Graphs represent an important and powerful tool for achieving interoperability in and across research domains and fulfilling the mission of the NFDI (National Research Data Infrastructure). Several consortia are building their own KG solutions, embedded in their overall data management strategy. The Working Group "Knowledge Graphs" (WG KGs) was established in the NFDI Section "(Meta)data, Terminologies, Provenance" to coordinate the development and use of KGs in all NFDI consortia. It has carried out an evaluation of the state of the art of KG adoption as well as the need for additional support in the NFDI. This led to the development of the KGI4NFDI (Knowledge Graph Infrastructure for the German National Research Data Infrastructure) service proposal which will support NFDI consortia by providing guidance and documentation around development practices as well as software dedicated to the creation and (re)use of KGs, including tools for data import, validation and export, collaborative frontends, search APIs, SPARQL endpoints, and tools for visualizing query results.

Besides decentralised tooling, the service will establish a registry of KGs utilized by NFDI consortia, which will be presented in the form of its own KG. Alongside this, the service will devise an interoperability strategy, conduct surveys, and demonstrate the application of KGs across various research fields and scenarios using diverse methods. The implementation will leverage a widely-used FLOSS technology stack to ensure maximum reusability and sustainability, with any generated solutions being made available under open source and content licenses for others to benefit from.

This presentation will outline the mission and core objective of KGI4NFDI that will be developed starting in the summer of 2024. We would like to use this opportunity to gather feedback as well as engage with the NFDI and research community.

Keywords: Knowledge Graph, Semantification, Research Data Management

A Knowledge Graph-Based Data Integration Workflow for NFDI4Culture and Beyond

Authors:

Jonatan Jalle Steller^{1*}, Linnaea Söhn¹, Torsten Schrade¹, Oleksandra Bruns^{2,3}, Tabea Tietz^{2,3}, Etienne Posthumus², Sarah Rebecca Ondraszek^{2,3}, and Harald Sack^{2,3}

*Lead presenter

¹jonatan.steller@adwmainz.de, Academy of Sciences and Literature Mainz, Geschwister-Scholl-Straße 2, 55131 Mainz, Germany

²FIZ Karlsruhe – Leibniz Institute for Information Infrastructure, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany

³Karlsruhe Institute of Technology (AIFB), Kaiserstr. 89, 76133 Karlsruhe, Germany

Abstract:

Despite covering specific scientific disciplines, all NFDI consortia share similar concepts such as organizations, people, institutions, areas of expertise, data repositories, projects, data sets, and research (meta)data, just to name a few [1, 2]. When interconnected, such information has the potential to open up new research horizons. To achieve this, however, the data needs to be available as linked open data (LOD). Hence a comprehensive workflow, which includes data discovery, harvesting, preprocessing, mapping, and integration into a knowledge graph (KG) is required, which is the topic of the present work [2, 4].

Taking the NFDI4Culture KG as an example, we present a workflow for data ingestion as LOD, which is, in principle, also applicable to other NFDI consortia. Here, the NFDI4Culture KG acts as a single point of access to various decentralized research data resources, and aggregates diverse and isolated data from the research domain, enabling discoverability, interoperability and reusability of cultural-heritage data [3].

The NFDI4Culture KG consists of the Research Information Graph (RIG), which describes metadata such as publishers, contact points, standards, licenses, and data portals, and the Research Data Graph (RDG), which interconnects the content metadata provided by data portals to make granular items accessible for search. Taking into account the challenges and objectives of NFDI4Culture to aggregate a diverse landscape of cultural-heritage research data for improved interoperability, we designed a Python package of reusable LOD components, harvesters using these components, a SPARQL endpoint explorer (*shmarql*), and an ETL (Extract, Transform, Load) pipeline. The latter consists of six modular workflow components, adaptable for independent use or within a comprehensive, automated ingest routine.

Step 1: run harvest routines. This works through a set of RDF-based action files with schema.org-based step definitions to scrape remote data, connect the feed to its metadata in the RIG, and generate persistent identifiers for imported resources. To ensure harmonization and interoperability across harvested data, transformations available in our Python package are applied to generate triples according to the *nfdicore/cto* ontology from various common data formats in the cultural-heritage domain, if necessary.

Step 2: clean harvested data. To ensure harmonization between the harvested data feed and its associated action file, triples representing the harvesting state are added or deleted.

Step 3: commit harvest state. Changes made by a harvesting run are pushed to the pipeline's own repository to ensure up-to-date action files.

Step 4: prepare and index data. If there are changes in a data feed, data directories are automatically updated or created and search indexes are produced.

Step 5: build a new endpoint. To prevent downtimes, a new SPARQL endpoint container is built while the previous version remains available. Once the new endpoint becomes operational, the old container is stopped and removed.

Step 6: publish statistics. In a last step, statistics about the integrated data feeds are pushed to a public dashboard. It supports data analysis and visualizations based on the execution of provided SPARQL queries.

Deletions and alterations in the KG are handled by the same routine as data feeds are periodically re-harvested and included in the next endpoint.

References:

- [1] Sack, H., Schrade, T., Bruns, O., Posthumus, E., Tietz, T., Norouzi, E., Waitelonis, J., Fliegl, H., Söhn, L., Tolksdorf, J., Steller, J.J., Azocar Guzman, A., Fathalla, S., Zainul Ihsan, A., Hofmann, V., Sandfeld, S., Fritzen, F., Laadhar, A., Schimmeler, S., Mutschke, P.: Knowledge Graph Based RDM Solutions: NFDI4Culture - NFDIMatWerk - NFDI4DataScience. In: 1st Conf. on Research Data Infrastructure (2023).
- [2] Bruns, O., Tietz, T., Söhn, L., Steller, J.J., Ondraszek, S.R., Posthumus, E., Schrade, T., Sack, H.: What's Cooking in the NFDI4Culture Kitchen? A KG-based Research Data Integration Workflow. In: 4th Workshop on Metadata and Research (objects) Management for Linked Open Science (DaMaLOS), co-located with ESWC (2024).
- [3] Tietz, T., Bruns, O., Fliegl, H., Posthumus, E., Schrade, T., Sack, H.: Knowledge Graph-basierte Forschungsdatenintegration in NFDI4Culture. In: Busch, A., Trilcke, P. (eds.) DHd2023: Open Humanities, Open Culture (2023).
- [4] Bruns, O., Söhn, L., Tietz, T., Steller, J., J., Posthumus, E., Schrade, T., Sack, H., Gotta Catch'em All: From Data Silos to a Knowledge Graph. In: ESWC Satellite Events: Poster and Demo Track (2024).

Keywords: knowledge graph, ontology, semantic web, research data management, FAIR

nfdi.software - unlock software, activate research data.

Authors:

Martin Hammitzsch^{1*}, Vasiliy Seibert², Aleksander Marčić³, Stephan Ferenz⁴, Oliver Werth⁴, Astrid Nieße⁴, Matthias Löbe⁵, Björn Grüning⁶, Leyla Jael Castro⁷, Michael Zacharias⁸, Beate Hetenyi¹ and Sonja Schimmler^{9 10}

*Lead presenter, martin.hammitzsch@gfz-potsdam.de

¹Helmholtz-Zentrum Potsdam Deutsches GeoForschungsZentrum GFZ

²Technische Universität Clausthal

³Nordrhein-Westfälische Akademie der Wissenschaften und der Künste

⁴OFFIS – Institute for Information Technology

⁵Universität Leipzig

⁶Albert-Ludwigs-Universität Freiburg

⁷ZB MED – Information Centre for Life Sciences

⁸Landessternwarte, Universität Heidelberg

⁹Fraunhofer FOKUS

¹⁰Technische Universität Berlin

Abstract:

nfdi.software aims to improve the access to and the use of research data with appropriate software. The proper linking and contextualisation of research software opens up the possibility of using research data sustainably and expanding the spectrum of analysis and processing. The need for a central access point for research software of the NFDI consortia arises from the growing need of the scientific disciplines of the cultural sciences, humanities and social sciences, engineering and the natural and life sciences to ensure the sustainable use and further development.

Although much work regarding research software is planned and conducted by at least 20 NFDI consortia, there is currently a lack of a cross-consortia initiative focused on common objectives, bringing together reliable contributions from these consortia with the necessary impetus and producing results that can be used by the majority of the consortia. nfdi.software is intended to link and coordinate independent individual developments from these areas in the federated data infrastructure.

Initialisation Phase 1, scheduled to start in late summer 2024, will establish the necessary communication for a concerted effort, bring together stakeholders and providers to jointly design the interplay of related consortia services, and provide an integrative prototype for first-hand experience to support the practical identification of issues and gaps and subsequently the design of Integration Phase 2.

nfdi.software provides the foundation for establishing standards and for networking across the NFDI. During the initialisation phase, an approach is developed based on the experience of previous initiatives, leading to a 'Prototype for Integration'. In its initial phase, it will build on top of four proven solutions, thus minimising duplication of work while maximising compatibility to other similar and/or complementary efforts. Bio.tools (1) and the Research Software Ecosystem is a portal supported by the European Infrastructure for Biological Information (ELIXIR), where researchers can find and compare bioinformatics tools thanks to curated metadata. The Research Software Directory (2) aims at finding and reusing cross-domain research software, stimulating reuse, encouraging proper citation, and making the impact of research software more visible. Betty's Research Engine (3) is a search engine that finds software repositories and links them to corresponding scientific publications and enriches the results with harmonised metadata. Physics.tools (4) searches in publications for referenced software and automatically retrieves accessible metadata information from the code repository. The criteria catalogue of the TMF ToolPool Health Research ensures exemplarily the high quality of registry entries.

Moreover, nfdi.software will utilise further solutions, e.g., Papers with Code, CodeMeta, CFF, and Zenodo, learn from them and integrate. The EOSC is the target destination when it comes to aligning with European efforts. It is aimed to push this concept by focusing on the needs in the national context towards “One NFDI”. As a central service, nfdi.software will harvest, aggregate, and harmonise software metadata from NFDI consortia marketplaces and registries, and will provide back enriched and curated metadata. The marketplace is therefore a complementary platform that also reflects the relevance, acceptance, and possibilities of the respective software within the research domains.

Keywords: Research Software, RSE, marketplace, metadata

Further fields

Chair: Reinhard Altenhöner (Staatsbibliothek – PK)

The Research Commons approach – an opportunity for Base4NFDI?

Authors:

Rory Macneil¹ and Tilo Mathes²

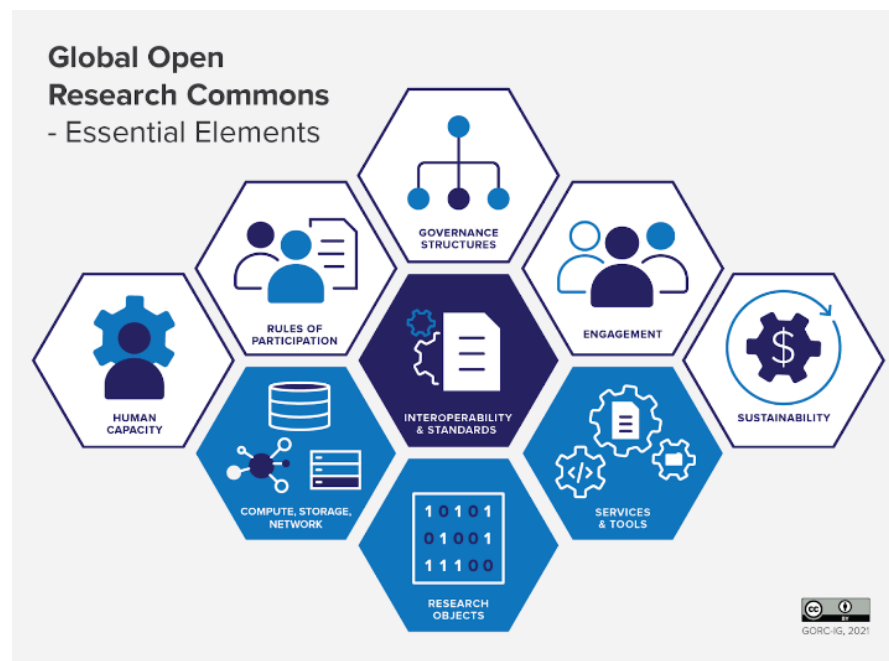
¹rmacneil@researchspace.com, Research Space

²tilo.mathes@researchspace.com, Research Space

Abstract:

This presentation reflects on how the growing number of Research Commons initiatives could be used as references in thinking about how Base4NFDI services could be offered as an integrated package of services or a Research Commons.

The publication by the RDA Global Open Research Commons Working Group in October, 2023 of the Global Open Research Commons International Model (GORC Model) made available a fully featured template for a Research Commons.ⁱ

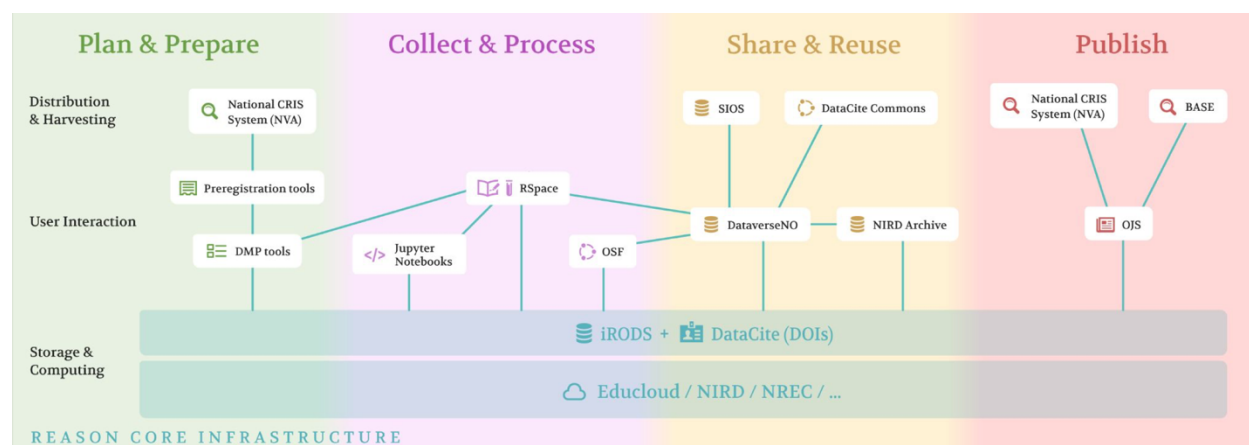


Since the publication of the GORC model, national organizations in Sweden, The Netherlands, Germany, and elsewhere, and ELIXIR, are using the GORC Model to explore establishing Research Commons. The Common Infrastructures group of the NFDI is also exploring the creation of a Research Commons.

The co-presenters were involved in the development of the GORC model and the proposed Norwegian Research Commons for Norway (REASON). REASON is the first detailed conceptualization of a Research Commons based on the GORC Model, and is being used as a reference by many groups exploring the establishment of Research Commons.

Two key features of Research Commons generally and including REASON are: first, they offer researchers access to an integrated series of complementary services accessible from a single platform; second, the platform is explicitly designed with reference to the research data lifecycle. The services offered are associated with different phases of the lifecycle, are interoperable, and

facilitate passage of data and metadata between tools throughout the research lifecycle. These principles are evident in the below graphic depicting the technical core of REASON.



Base4NFDI offers a growing number of Basic Services, e.g., data management planning and support for PIDs, that are identified in the GORC model and included in REASON. The poster explores how the GORC model and REASON can serve as a useful reference for the evolution of Base4NFDI into a Research Commons.

The poster is divided into the following sections:

1. Introduction to the GORC Model
2. Use/adaption of the GORC model to design REASON
3. REASON's applicability as a reference for other Research Commons
4. Comparison of REASON with the Basic Services offered by Base4NFDI
5. Exploration of a 'Base4NFDI Research Commons' with reference to REASON

Keywords: Research Commons, Research Infrastructure, GORC

ⁱ <https://zenodo.org/records/10694490>

Research Software Engineering within the NFDI (INFRA-WG-RSE)

Authors:

Bernd Flemisch^{1*}, Martin Hammitzsch^{2*}, and Florian Thiery^{3*}

*Lead presenter

¹ bernd.flemisch@iws.uni-stuttgart.de, University of Stuttgart

² martin.hammitzsch@gfz-potsdam.de, GFZ German Research Centre for Geosciences

³ florian.thiery@leiza.de, Leibniz-Zentrum für Archäologie

Abstract:

Research Software Engineering (RSE) is fundamental to the German National Research Data Infrastructure (NFDI). Within the NFDI e.V., several sections deal with overarching topics, including RSE technologies and knowledge, especially the "Section Common Infrastructures" with its working groups on "Data Integration (DI)", "Data Management Planning (DMP)", "Data Science and Artificial Intelligence (DSAI)", "Electronic Lab Notebooks (ELN)", "Persistent Identifiers (PID)" and "Research Software Engineering (RSE)".

The RSE working group connects the NFDI consortia in software-related aspects. It focuses on Research software, software communities, and software infrastructure at NFDI. The working group operates a central forum in an advisory and supportive capacity. It establishes the necessary software ecosystem within NFDI for the professional development of software infrastructure components, which represent an integral part of the NFDI. In addition, the working group serves as an interface for the NFDI to compare European and international initiatives to promote the connectivity of the NFDI with other infrastructures.

The WG RSE comprises several tasks that are also related to Base4NFDI:

- Task *Jupyter Services*: bring together the JupyterHub providers in Germany; collect and list available hubs at a common website; evaluate the possibility of an NFDI-wide Jupyter service. This is addressed in the Jupyter4NFDI Basic Service (in its initialisation phase. Jupyter4NFDI aims to address the fragmented deployment of Jupyter Notebooks across NFDI consortia by offering a centralised service. This service will simplify access, improve user experience, and extend Jupyter's reach to a wider audience within and beyond the NFDI.
- Task *Software Marketplace*: Allow access to the portfolio of research software in the NFDI; develop a concept for the Marketplace; consider requirements from the consortia and available solutions. This has been substantiated in a proposal to Base4NFDI.
- Task *Training Materials*: Create recommendations to spread good practices using available content.
- Task *Mission Statement*: Develop a common mission statement, "Software Engineering in the NFDI".
- Task *Status Quo Survey and Report*: Apply an RSE survey for a "Landscape Analysis" within the NFDI.
- Task *Overarching Initiatives*: Ensure connectivity and compatibility with European and international efforts.
- Task *Software Ecosystem*: Develop a draft strategy for establishing a software ecosystem within the NFDI.

- Task *Use Cases*: Identify use cases for a cross-consortium, prototypical development of the software ecosystem.
- Task *Quality Criteria for Research Software*: Identify and develop suitable, easy-to-record indicators.
- Task *Criteria for NFDI Components*: Ensure reliability, compatibility, and security; provide checklists/test suite.

This contribution presents the WG RSE activities and the possibilities for participation (small subgroups work on individual tasks, monthly WG-wide coordination meetings) in the working group discussion rounds that bring the NFDI consortia together in all software-related aspects.

Keywords: Research Software Engineering

Challenges, Opportunities and Experiences of setting up sustainable Research Data Infrastructure

Authors:

Andreas Ganzenmüller^{1*}, Preston Rodrigues², Volodymyr Kushnarenko², Miroslav Puskaric², and Thomas Bönisch²

*Lead presenter

¹andreas.ganzenmueller@hlrs.de, High-Performance Computing Center Stuttgart (HLRS), University of Stuttgart, Germany

²High-Performance Computing Center Stuttgart (HLRS), University of Stuttgart, Germany

Abstract:

Setting up a sustainable research data infrastructure is not just about hardware and software. It is also about expertise, routines and workflows on how to handle research data on a daily basis, as well as the necessity to have the required human resources to maintain the infrastructure in the long term.

As an infrastructure provider HLRS is collaborating in several projects [1, 2, 3, 4] concerning research data management. In all these projects there is a requirement that some services (primarily research data repositories) are set up and perspective kept running for many years. To accomplish this, we at HLRS decided to develop a platform dedicated primarily to the needs of research data management. This platform is supposed to accommodate all the required services. Currently our platform comprises a small-scale virtualisation cluster running Proxmox Virtual Environment, an open-source server virtualisation management solution [5], and a dedicated storage system.

Since we began developing the platform, we have gained important insights which we would like to share with the community. Starting with hardware selection, integration into existing infrastructure, network configuration, hosting web-services and ensuring sustainable and secure maintenance. We would like to provide insight into our architecture. We will elaborate motivation behind our design decisions and demonstrate how research data management software can be deployed on the platform. Additionally, we would like to emphasize the critical role of detailed technical documentation and the dedicated team responsible for its operation, particularly in ensuring the platform's long-term operation.

Overall, we identified three issues to be major design challenges: flexibility, scalability and sustainability.

Flexibility is a key issue, because the platform must be able to run a wide variety of software and integrate different services to work together. To accomplish this, some small adaptations of software components are inevitable. But any large software customisation needs to be met with caution, because maintenance effort can increase drastically. The goal should be to generate synergy effects by having multiple projects being administrated by the same service provider and benefit from the accumulated experience and shared resources.

Scalability is critical, because the platform must be expandable to accommodate further services in the future. At the same time, it is necessary to aim for an efficient hardware utilisation to lower over-all costs for all participants. Extending the hardware base of the

platform therefore is a continuous task that requires careful planning in advance.

Concerning sustainability, the overall challenge lies in setting up organisational structures that allow to maintain the platform independent of the financial contributions of single service partners. Therefore, the platform requires a certain minimum size to serve enough projects and accommodate their services, to ensure operations in case one service partner unexpectedly drops out.

Our platform can serve as an example for sustainable research data infrastructure and our designs and workflows can be used as guidelines to develop, configure and maintain similar platforms and related services for big communities such as NFDI.

References:

- [1] SDC4Lit Homepage. URL: <https://www.sdc4lit.de/>, last access on 15.05.2024.
- [2] NFDI4Cat Homepage. URL: <https://nfdi4cat.org/>, last access on 15.05.2024.
- [3] KI Allianz Baden Wuerttemberg Homepage. URL: <https://ki-allianz.de/>, last access on 15.05.2024.
- [4] Meridional Homepage. URL: <https://meridional.eu/>, last access on 15.05.2024.
- [5] Proxmox Homepage. URL: <https://www.proxmox.com/en/>, last accessed on 15.05.2024.

Keywords: Infrastructure, Research Data, Data Management, Sustainability

Beyond lifetimes: what does archiving mean for research data?

Authors:

Kai Naumann

kai.naumann@la-bw.de, Landesarchiv Baden-Württemberg, NFDI4Memory, NFDI Association, Section Infrastructures

Abstract:

Archiving is an everyday task, but in terms of the German legal system, it is also a defined process assigned to specialized institutions. State archives, academic, municipal, business, political, and private archives hold the historical record of the institution they take care of. Historical records are defined very broadly and not confined to the subjects people expect in history books. There are registers, journals, data timelines, photographs, audio and video recordings, both in analog and digital form. Similar processes are carried out by some research libraries.

As a person involved in the NFDI Sections Common Infrastructures and Ethical and Legal Aspects, most prominently of the Long Term Archival and Access Group, and as proxy of my colleagues in the memory institutions mentioned above, I would like to engage with the NFDI community on how the German academic society would like to configure archiving of research data, i.e. keeping data of various kinds available beyond the lifespan of hardware and software. This includes steps beyond the end of life of data centers, data formats, encoding schemes, or even business models. Our current legislation supports the view that some fine tuning in the legal setting is changing, maybe even more. Are these moves well-reflected or ad hoc solutions? Are these movements similar in all areas of science?

See (partly in German)

<https://archive20.hypotheses.org/13005>

<https://www.zotero.org/groups/5240872/nestortoolbox/collections/CGRK6EFP>

Keywords: legislation, continuity, archival law, sustainability

The FID Middle East-, North Africa- and Islamic Studies – as a user and partner of base4nfdi

Authors:

Dr. Josef Jeschke^{1*}, Dr. Volker Adam²

*Lead presenter

¹josef-peter.jeschke@bibliothek.uni-halle.de, FID Middle East-, North Africa- and Islamic Studies, ULB Saxony-Anhalt

²FID Middle East-, North Africa- and Islamic Studies, ULB Saxony-Anhalt

Abstract:

FID and NFDI share some goals: better findability, accessibility and reusability of research data. Both networks strive to avoid parallel developments. Base4NFDI might and should function as a hub for basic RDM developments and tools, that are made accessible to those who generate and manage the research data: the whole research community.

The approach of a central, coordinated infrastructure meets one obstacle that many services have to address: Germany's decentralized and diverse research environment. Over the last decade the FID system has gained a rich experience how to integrate the heterogeneous demands of their respective research community into a coordinated supply of research resources like printed and electronic research literature, databases, publishing services, counseling, and more. As result to its strong connections into the research communities the FID system fulfills a twofold function: provider of resources and broker of services.

The FID activities concerning RDM are primarily amongst the second function, but since research data are resources too, the FID aims to make reusable research data findable. What the FID do not do is to provide RDM tools, but researches often seek our advice about them. Accordingly, the FID might be seen as a possible user and broker of BASE4NFDI services.

From the FID Middle East-, North Africa- and Islamic Studies point of view four of the current six projects run by BASE4NFDI are of special interest: IAM4NFDI, PID4NFDI, TS4NFDI, DMP4NFDI.

IAM4NFDI aims to provide authentication systems with in NFDI. The obstacle of identification and authentication has to be addressed by our FID as well, since our research community is often confronted with data that may not be made available openly, because of copyright, data security or the CARE principles. The main challenge is the access management to data of finished projects.

Even though PID are an essential part of an effective RDM, their usage is not fully applied in our research community. Consequently, training on the assignment and usage of PID is one of the FIDs tasks. So, the FID strongly welcomes that PID4NFDI will design a work programme to build a PID service.

Our FID has a great experience in the work with controlled terminologies, especially the GND. That is why our FID sees a great potential in an involvement of the FID in the development Terminology Services 4 NFDI (TS4NFDI).

Since the FID offers counselling on research data management plans we regard the development of DMP4NFDI as highly relevant and would like to connect our research community to its services.

The proposed contribution aims to point out the FID's demand as a user of BASE4NFDI services and to emphasize the potential of a collaboration in the development. Only by communication about demands and by collaboration in development we may efficiently set up a joint research infrastructure and avoid parallel structures. The contribution will strongly advocate the importance to design services that reach beyond the system of NFDI consortia. The contribution will be prepared in close contact to the FID workgroup RDM and aims reflect a wider FID perspective.

Keywords: [FID] [NFDI] [collaboration] [efficiency] [RDM]

Software metadata for BASE and NFDI

Authors:

Leyla Jael Castro¹, Dhvani Solanki^{1*}, Nelson Quiñones¹, Lukas Geist¹, and Dietrich Rebholz-Schuhmann¹

*Lead presenter

¹{ljgarcia, solanki, quinones, geist, rebholz-schuhmann}@zbmed.de, ZB MED Information Centre for Life Sciences

Abstract:

Structured, semantic, and machine-actionable metadata for all sorts of research artifacts is a must when it comes to the Findable, Accessible, Interoperable and Reusable (FAIR) principles [1]. Metadata makes it easier for search engines, recommenders, aggregators, archives, and registries to provide a harmonized and connected view across research disciplines and artifacts, including BASE4NFDI services. Metadata for scholarly publications is already consolidated thanks to efforts such as [CrossRef](#) and [DataCite](#) (although richer metadata is still possible) while metadata for data is gaining ground as data is already recognized as a key research artifact. The case of metadata for software is still trying to find its way and getting stronger thanks to initiatives such as FAIR for Research Software (FAIR4RS) [2], [CodeMeta](#) [3], [Bioschemas](#) [4], Software Management Plans (SMPs) (e.g., by ELIXIR [5] or Netherlands eScience Centre [6]) and [machine-actionable SMPs \(maSMPs\)](#) [7–9] (i.e., a semantic metadata layer to describe SMPs and the developed software, based on schema.org [10]).

Various efforts around metadata for software are currently discussed in NFDI sections and consortia, including the Metadata for Research Software Engineering Working Group (MetaRSE WG, Section Metadata), the nfdi.software BASE service, and the NFDI4DataScience consortium –in particular wrt maSMPs and their connection to Machine Learning (ML) Models. Related discussions in other communities include EOSC with the analysis of quality characteristics for research software and its alignments to FAIR4RS [11], and the Research Data Alliance FAIR4ML Interest Group as research software is used to create ML models. Additional efforts look to automatically extract metadata from software repositories (e.g., SoMEF [12]) or to find mentions of software in literature (e.g. SoMeSci [13]). Despite these efforts, there is still a need to better understand the common metadata layer for software across NFDI consortia, a task that is ongoing in the MetaRSE WG, plus a need to connect data and software metadata at management plans level, a task that will be soon started at NFDI4DataScience as this connection is key to enable Data Science Management Plans.

Supporting metadata for software across NFDI consortia and BASE services will facilitate the creation of knowledge graphs of software together with their scholarly adoption and use in data-driven approaches. We argue that schema.org can be used to get a common lightweight layer across multiple disciplines and consortia [14,15] while more specialized vocabularies including ontologies can be used for more in-depth analysis (that will become possible thanks to the existence of the lightweight layer).

Keywords: Metadata, Research Software, machine-actionability

References

1. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*. 2016;3: 160018. doi:10.1038/sdata.2016.18
2. Barker M, Chue Hong NP, Katz DS, Lamprecht A-L, Martinez-Ortiz C, Psomopoulos F, et al. Introducing the FAIR Principles for research software. *Sci Data*. 2022;9: 622. doi:10.1038/s41597-022-01710-x
3. Jones MB, Boettiger C, Mayes AC, Arfon Smith, Slaughter P, Niemeyer K, et al. CodeMeta: an exchange schema for software metadata. *KNB Data Repository*. KNB Data Repository; 2016. doi:10.5063/SCHEMA/CODEMETA-1.0
4. Gray AJG, Goble C, Jimenez RC. From Potato Salad to Protein Annotation. ISWC Posters and Demo session. Vienna, Austria; 2017. p. 4. Available: <http://ceur-ws.org/Vol-1963/paper579.pdf>
5. Alves R, Bampalakis D, Castro LJ, González JMF, Harrow J, Kuzak M, et al. ELIXIR Software Management Plan for Life Sciences. *BioHackrXiv*; 2021. doi:10.37044/osf.io/k8znb
6. Martinez-Ortiz C, Martinez Lavanchy P, Sesink L, Olivier BG, Meakin J, de Jong M, et al. Practical guide to Software Management Plans. *Zenodo*; 2022 Oct. doi:10.5281/zenodo.7248877
7. Castro LJ, Giraldo O, Geist L, Quiñones N, Solanki D, Rebholz-Schuhmann D. machine-actionable Software Management Plan Ontology (maSMP Ontology). *Zenodo*; 2024. doi:10.5281/zenodo.10582073
8. Giraldo O, Geist L, Quiñones N, Solanki D, Alves R, Bampalakis D, et al. A metadata schema for machine-actionable Software Management Plans. *PUBLISSO-FRL*; 2023. doi:10.4126/FRL01-006444988
9. Giraldo O, Dessi D, Dietze S, Rebholz-Schuhmann D, Castro LJ. Machine-Actionable Metadata for Software and Software Management Plans for NFDI. *Proceedings of the Conference on Research Data Infrastructure*. 2023. doi:10.52825/cordi.v1i.279
10. Guha RV, Brickley D, Macbeth S. Schema.org: evolution of structured data on the web. *Commun ACM*. 2016;59: 44–51. doi:10.1145/2844544
11. David M, Colom M, Garijo D, Castro LJ, Louvet V, Ronchieri E, et al. Ensure Software Quality. *Zenodo*; 2024 Feb. Available: <https://zenodo.org/records/10723608>
12. Mao A, Garijo D, Fakhraei S. SoMEF: A Framework for Capturing Scientific Software Metadata from its Documentation. 2019 IEEE International Conference on Big Data (Big Data). 2019. pp. 3032–3037. doi:10.1109/BigData47090.2019.9006447
13. Schindler D, Bensmann F, Dietze S, Krüger F. The role of software in science: a knowledge graph-based analysis of software mentions in PubMed Central. *PeerJ Comput Sci*. 2022;8: e835. doi:10.7717/peerj-cs.835
14. Gray A, Castro LJ, Juty N, Goble C. Schema.org for Scientific Data. *Artificial Intelligence for Science*. WORLD SCIENTIFIC; 2022. pp. 495–514. doi:10.1142/9789811265679_0027
15. Castro LJ, Fluck J, Arend D, Lange M, Martini D, Neumann S, et al. Schema.org as a Lightweight Harmonization Approach for NFDI. *Proceedings of the Conference on Research Data Infrastructure*. 2023. doi:10.52825/cordi.v1i.280

Title: Cooperation in Research Data Management across Base4NFDI, regional RDM initiatives, and beyond – considerations from SaxFDM

Authors:

Moritz Kurzweil^{1*}, Alina Bastian², Carolin Hundt³, Dr. Guido Juckeland⁴, Dr. Ralph Müller-Pfefferkorn², Maria Rein⁵, Johannes Sperling¹, Dr. Barbara Weiner³

* Lead presenter

¹ moritz.kurzweil@leibniz-gwzo.de, Leibniz-Institut für Geschichte und Kultur des östlichen Europa (GWZO)

² Technische Universität Dresden

³ Universität Leipzig

⁴ Helmholtz-Zentrum Dresden Rossendorf

⁵ Hochschule Mittweida

Abstract:

Effectively disseminating and developing research data management (RDM) tools and services requires a multifaceted approach that leverages the expertise of various actors within the RDM landscape. This contribution explores how the activities of Base4NFDI, regional initiatives and other RDM actors can complement each other to contribute to this shared goal.

Building on the "Joint Plea for Close Integration of the State Initiatives for Research Data Management into the National Research Data Infrastructure" (<https://doi.org/10.5281/zenodo.4524655>), we examine the RDM landscape made up of its various actors: national consortia (NFDI), regional initiatives (e.g. SaxFDM), local institutional RDM support staff, as well as dedicated RDM personnel in research projects. As each of them plays a critical role in implementing RDM practices and fostering data literacy, we then delve into strategies to build fruitful and more elaborate relations between these actors. This can take different complementary forms, such as: clearly dividing responsibilities, collaborating in specific areas, and connecting those seeking support with experts.

We provide an overview of Saxony's RDM landscape, outlining resources and services offered by both Saxon research institutions and the SaxFDM initiative, and sharing our experiences with user involvement and distributed collaboration.

Based on a reflection on the types of users of RDM services and on developing services considering their respective needs, we then make suggestions for cooperation between Base4NFDI, regional initiatives, and other actors. Due to their integration in local structures, regional initiatives can help Base4NFDI to reach end users, and collaboratively develop and disseminate NFDI services according to the users' needs. An example of such cooperation could be jointly organized data literacy trainings in their respective regions.

Keywords: RDM services, regional RDM initiatives, data literacy, training, collaboration

Basic services in integration phase:

IAM4NFDI, TS4NFDI, PID4NFDI

Chair: Raphael Ritz (MPCDF)

IAM4NFDI – The Community Experience

Authors:

Marius Politze^{1*}, Marcel Nellesen², Wolfgang Pempe³, and Thorsten Michelsⁿ

*Marius Politze

¹politze@itc.rwth-aachen.de, RWTH Aachen University

²RWTH Aachen University

³DFN-Verein

ⁿRPTU Kaiserslautern-Landau

Abstract:

The basic service Identity and Access Management, IAM4NFDI, which is funded via Base4NFDI, aims to establish a sustainable, long-term technical and organizational framework for access to research data and the services required to manage it - in other words, it addresses the “Accessible” aspect of the FAIR principles. In addition, interoperability and connectivity to other infrastructures such as the European Open Science Cloud (EOSC) must be guaranteed. An Authentication and Authorization Infrastructure (AAI) serves as the basis for this. Via DFN-AAI and eduGAIN, this NFDI-AAI is connected to more than 5500 universities and research facilities worldwide. Central building blocks of this AAI are the so-called Community AAI, which enable the research communities represented by the NFDI consortia to manage and control access to the resources relevant to them.

The IAM4NFDI project is currently right in the middle of the Integration Phase. It's a good time to talk about our experiences with community engagement. We will talk about our lessons learned regarding the promotion of the concept of Community AAI and related topics among the consortia. Important factors in this regard so far have been the Community-AAI-as-a-Service offering and the so-called Incubators. Incubators are small integration and consulting projects – one cycle lasts 6 months – in which the IAM4NFDI project team supports communities in implementing AAI use cases. These use cases may, for example, involve connecting services to a Community AAI, deeper integration of complex services, the development of new (AAI-)features or implementing the NFDI-AAI policy framework. On this occasion, participants in the first incubator cycle will have the opportunity to report on their experiences so far.

Keywords: IAM4NFDI, Identity & Access Management, AAI, Community AAI, Incubator

Polished Policies for the NFDI-AAI

Authors:

Marcus Hardt¹, Wolfgang Pempe²

¹ hardt@kit.edu, KIT

² pempe@dfn.de, DFN-Verein

Abstract:

Identity and Access Management (IAM) is one key technology for enabling users to access a large variety of resources provided by different infrastructures. For this, users make use of their Home Organisation, i.e. a university, research institute, or social media account.

This is only possible with a small portion of common sense and regulation regarding the operation and security of these infrastructures.

As for the NFDI-AAI, these are captured in a set of policies and templates. The guidelines and documents summarised in this framework address different levels and aspects of Identity and Access Management and are based on internationally established standards and guidelines. Experience of the past two decades has helped to continuously improve them. Detailed explanation is available in the AARC Policy Development Kit (PDK)[PDK].

This talk will provide a brief overview of the structure and the components of the NFDI-AAI Policy Framework, as well as their relevance for the different roles and functions within a research infrastructure.

[PDK] <https://aarc-community.org/policy/policy-development-kit/>.

Keywords: IAM4NFDI, Policies, Identity & Access Management, AAI, Security

Experiences and status quo of IAM / CAAI in DataPLANT

Authors:

Dirk von Suchodoletz*, Jonathan Bauer², and Marcel Tschöpe³

*Lead presenter

¹dsuchod@uni-freiburg.de, Universität Freiburg

²jonathan.bauer@rz.uni-freiburg.de, DataPLANT, Universität Freiburg

³marcel.tschoepe@rz.uni-freiburg.de, DataPLANT, Universität Freiburg

Abstract:

The various tools and services in DataPLANT need to be authenticated to allow access to certain resources. There are primarily two types of authentication and authorization required: users provide credentials to run tools and interact with services. Tools and services need to authenticate to each other e.g. to exchange information and data. Already in the very beginning of the project we identified key objectives to implement proper identity and access management for our service landscape and later on further scientific communities in both centralized cloud and decentralized on-premises installation:

- Provide community members with access to the various resources within DataPLANT without requiring to create a completely new online identity
- Derive a DataPLANT identity from various AAI, primarily Life Sciences AAI or ORCID iD, and provide an SSO environment
- Be flexible and open to include further identity providers to adapt to ongoing developments and future requirements like the integration with other scientific communities within the NFDI
- Build upon existing standards, use Open Source tools and create as few as possible solutions proprietary to DataPLANT
- Be open and flexible to integrate with third party services like Galaxy or Nextflow provided by others
- Provide an easy to use interface to manage DataPLANT identities both by administrators and DataPLANT users, allow for as much self-service as possible

The DataPLANT user management builds upon existing Community AAls (CAAI). Well established services like Life Sciences AAI and ORCID can be combined with local authentication within the central DataPLANT authentication service. The infrastructure is based on Keycloak, a widely used AAI proxy developed by RedHat, that supports modern authentication protocols like OpenID-Connect or SAML. It allows the integration of multiple AAls and identity brokering in an unified environment. Providing an AAI identity management, which can easily be connected with GitLab and other services through either protocols, simplifies the DataPLANT wide user management. The connection of multiple AAls through KeyCloak enables our community to use their existing accounts, for example from the Life Sciences AAI, their home institution or ORCID. During authorization we can assign different roles depending on the account source or on specific attributes. Permissions can be derived from these roles to differentiate between users. These range from privileged users having full access to the data and the ability to create archives/publications, to underprivileged users that have only a reporting function and/or read-only access to raw data. However, authorisation and accounting is still a major challenge. To avoid lengthy negotiations between organizations, it is necessary to ensure that authorized users can only

use the resources to which they are entitled, otherwise a potentially unfair pricing model emerges on both sides. To a large extent, this can be addressed through a well-defined resource request process that takes into account funding and partnerships - which can be integrated into an automated process through user attributes in the respective community AAls. This issue needs to be discussed at all levels: NFDI AAls, other federated AAls (such as the state project bwIDM2) and local institutions. In the last couple of months we started to interact with the IAM4NFDI team (<https://base4nfdi.de/projects/iam4nfdi>) to explore further options of cross-NFDI integration. We started to discuss the following issues with the RegApp team at KIT, one of the four IAM4NFDI solution options presented in the workshops.

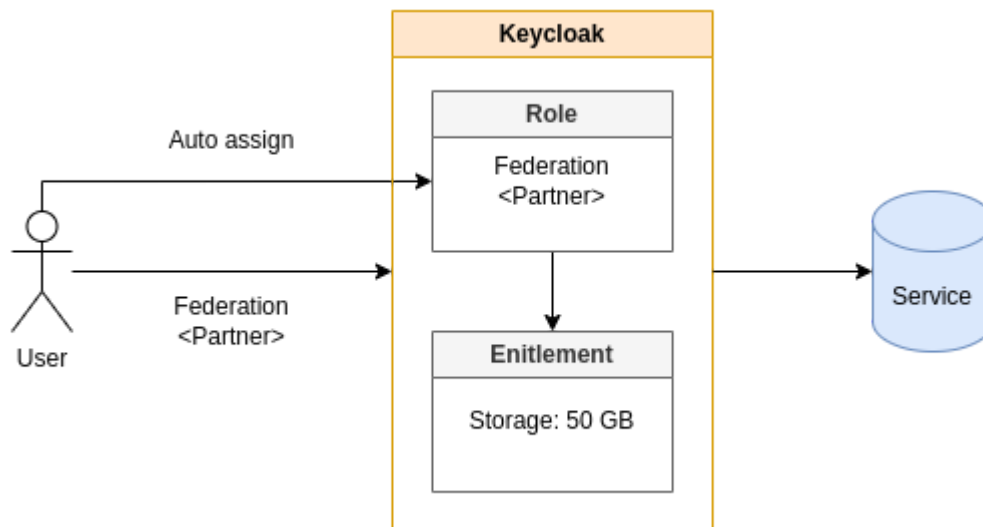


Fig: Assignment of authorisations of a user of a federated cooperation for a service

In this example, a user logs on to the system through a federation partner to access a service. When the user logs in through the federation partner's IDP, the user is automatically added to a pre-defined role that includes all users who log in through the federation partner. All users associated with this role are automatically granted permission to use 50 Gigabytes of storage on the service. To accomplish this, the "Storage: 50 GB" entitlement is added to the aforementioned role. [Fig]

Während Authentication weitflächig gelöst, sind sinnvolle Abbildung von Gruppenzugehörigkeiten oder die Zuweisung von spätere Abrechnung von genutzten Ressourcen eine größere Herausforderung. Das gilt beispielsweise für signifikante Speicher- oder Compute-Bedarfe.

- * Wie kann man individuellen Nutzern bestimmte Ressourcen zuweisen, die sie bei der Nutzung des Dienstes anfordern können

- * Dabei Mapping von verschiedenen Rollen auf verschiedene Ressourcen

- * **Wie stellt man sicher, dass diese Ressourcen nicht überschritten werden (d.h. dass der Anbieter "kein Geld verliert" bzw. der Nutzer bzw. dessen Einrichtung keine überraschend hohen Rechnungen erhält, für die keine Mittel hinterlegt wurden)**

- * **Wie bildet man das sinnvoll ab, so dass hier nicht längliche Aushandlungsprozesse zwischen Nutzern, ihrer Einrichtung und den Diensteanbietern notwendig sind**

Fragen/Überlegungen aus bwIDM2 zu Attributen, Ressourcen-Zuweisungen im Rahmen von Authorization und Gruppenverwaltung mit Blick auf NFDI ...

Lösungsmöglichkeiten zur Kopplung mit Regapp / Community? Hinsichtlich NFDI (aber nicht nur) wäre es sicher sinnvoll, wenn man eine Ressourcenverwaltung hätte, die – ähnlich Jards – mit der RegApp gekoppelt wäre.

So ganz grob:

- * Jards: Ressourcenanträge

- * RegApp: AAI und Community (delegierte PI und Gruppen)

- * ResMan: Verwaltung von Ressourcen/Kontingenten/evtl. in Verbindung mit hinterlegten monetären Mitteln/... Das wäre sicherlich für diverse Services eine interessante Ergänzung.

There will be FAIR. Insights into the PID landscape analysis

Authors:

Sara El-Gebali^{1*}, Stephanie Hagemann-Wilholt², Torsten Kahlert² and Marc Lange³

*Lead presenter

¹sara.elgebal@datacite.org, Datacite

²TIB Hannover

³Helmholtz Open Science Office

Abstract:

This presentation will demonstrate how a strategic, PID-based framework can transform research data management into FAIR, paving the way for enhancements in the reliability and accessibility of scientific research.

One of the main goals of the basic service PID4NFDI is gaining insights from extensive surveys to optimize PID application across varied research outputs, addressing current challenges within NFDI services according to diverse requirements and different maturity levels of PID usage. We will also explore how PIDs do more than ensure persistence and identification; they significantly enhance metadata quality and research integrity.

By integrating Persistent Identifiers (PIDs) with diverse research elements – such as instruments, methodologies, datasets, DMPs and publications – we enhance both the replication of experiments and the reuse of research data. This connectivity enables researchers to access all necessary components for comprehensive studies, thereby increasing the reliability and utility of research findings. Additionally, by making the provenance and production/creation process of diverse research outputs more transparent and by providing accessible licensing information through PIDs, we foster a transparent research environment that maintains high standards of integrity and promotes open science. In addition, PIDs offer context information by providing metadata on protocols and methods, software and input/output datasets, creators and contributors including affiliation organizations. They enable us to trace other related publications e.g. citations and references where an instrument or physical sample is used. Ultimately, leading to reduced duplicate efforts, and allows higher re-use of research data, making it more transparent and impactful.

PID4NFDI is committed to providing clear guidance on choosing and implementing the right PIDs for different types of research outputs. We are developing a comprehensive matrix that identifies suitable PID services for various research entities at different stages of the data lifecycle. This information will help researchers identify the most effective PIDs for their needs. The proposed solution/outcome is in direct response to the challenges identified by NFDI service providers, as highlighted in our survey responses and further feedback from consortia.

We will present the results of our landscape analysis and the resulting outcomes to guide subsequent implementation phases of a PID4NFDI basic service. Additionally we will highlight preliminary results from our cooperation with selected use case partners, where we examine use case-specific challenges, interoperability issues with existing PID infrastructures, and metadata quality insights. Comprehensive, open and linked metadata of research resources and entities is crucial for an ecosystem of well-functioning PIDs offering the above mentioned benefits to support academic collaboration through data sharing and trustworthiness of scientific processes. However, these can vary greatly depending on the discipline, subject of research and methods, and accordingly require different approaches to documentation and referencing.

At the end of the first project phase, a first collection of best practices should be available that can be reused by the consortia for typical use cases in NFDI service infrastructures. Referring

to our evolving training and support concept, we will outline in our presentation how the PIDs4NFDI basic service can support consortia in the future.

Keywords: Persistent Identifier, Basic Service, Metadata, Interoperability

First steps of the basic service Terminology Services 4 NFDI - Step by step through the initialization to the integration phase

Authors:

Roman Baum^{1*}, Jan Fillies², Naouel Karam², Oliver Koepler³, Pooya Oladazimi³, and Julia Sasse¹

*Lead presenter

¹baum@zbmed.de, ZB MED - Information Centre for Life Sciences, Cologne, Germany

²Institute for Applied Informatics (InfAI): Leipzig, Germany

³TIB - Leibniz Information Centre for Science and Technology, Hannover, Germany

Abstract:

Terminologies are crucial for creating semantically rich metadata that convey the full meaning of research data. They establish consensus definitions for entities, ensuring conceptual consistency across various disciplines, despite differing nomenclatures.

Terminology Services provide access to domain-specific collections of ontologies, terminologies, or vocabularies, offering comprehensive functions for human users via GUIs or for machines via APIs.

The basic service Terminology Services 4 NFDI (TS4NFDI) aims to develop a cross-domain, interoperable service to provide, curate, develop, harmonize, and map terminologies for the communities of the National Research Data Infrastructure (NFDI). The service will foster the harmonization and standardization of terminology management within the NFDI, facilitating consensus-building across communities.

We present our work on the initialization of the Basic Terminology Service, addressing the following key objectives:

- Provision of a consortia-agreed IT-Architecture with a harmonized API gateway to access multiple terminology services technologies based on OLS [1], OntoPortal [2] or SKOSMOS [3].
- Development of a Terminology Service Suite (TSS) to support uniform terminology access for provision, management, curation, publication, archiving, and subscription of terminologies.
- Promotion of consensus-finding, harmonization, and alignment by integrating TS4NFDI service components and adopting common practices to harmonize terminologies across disciplines.

We report on the measures taken to achieve the objectives and their respective outcomes. Our comprehensive requirement analysis included a detailed analysis of current terminology service technology stacks, tools, and services used by NFDI consortia to identify needs and gaps. Results were obtained by a survey and extensive interviews with experts from various NFDI consortia representing the diversity of scientific disciplines. We demonstrate reusable JavaScript-based GUI widgets for easy integration into user interfaces and visualization of semantic information as part of Terminology Service Suite. Several small prototypes and demonstrators were developed within the various NFDI consortia, showcasing the desired networking of terminology experts and the TS4NFDI project. We further present a pilot implementation of a service wrapper and API gateway integrating selected backend services of the aforementioned three main terminology service technologies. We further report on the integration of a central Mapping Service to create, manage, and provide cross-domain mappings for terminologies, utilizing the Simple Standard for Sharing Ontology Mappings (SSSOM) [4].

We present collected and derived use case scenarios and their evaluation against the pilot to prove the concept of envisioned architecture.

As an outlook, we will provide future work on the integration phase of the Basic Terminology including necessary steps to incorporate the service into the common NFDI service architecture. We will expand the use case scenarios including further participants from additional communities. The work will also include additional features and functionality to support the measures of ontology harmonization and mapping across the NFDI consortia.

Keywords: terminology service, ontology mappings, ontology, terminology

- [1] S. Jupp, T. Burdett, C. Leroy, and H. Parkinson, 'A new Ontology Lookup Service at EMBL-EBI', in SWAT4LS, 2015.
- [2] 'Welcome to the OntoPortal Alliance', Ontoportal Alliance. <https://ontoportal.org/>
- [3] O. Suominen et al., 'Publishing SKOS vocabularies with Skosmos'.
- [4] N. Matentzoglou et al., 'A Simple Standard for Sharing Ontological Mappings (SSSOM)', Database, vol. 2022, p. baac035, Jan. 2022, doi: 10.1093/database/baac035.

Incoming services

Chair: Reinhard Altenhöner (Staatsbibliothek – PK)

A Platform for the Provision of Teaching and Training Contents in Data Literacy: DALIA4NFDI

Authors:

Jochen Ortmeyer^{1*}, Thomas Stäcker², Gábor Kismihók³, Matthias Müller⁴, Peter F. Pelz⁵, and Sonja Herres-Pawlis⁶

*Lead presenter

¹jochen.ortmeyer@ac.rwth-aachen.de, RWTH Aachen University, Institute of Inorganic Chemistry

²Technische Universität Darmstadt, Universitäts- und Landesbibliothek

³Technische Informationsbibliothek Hannover, Leibniz Informationszentrum Technik und Naturwissenschaften

⁴RWTH Aachen University, IT Center, Lehrstuhl für Informatik

⁵Technische Universität Darmstadt, Institut für Fluidsystemtechnik

⁶RWTH Aachen University, Institute of Inorganic Chemistry

Abstract:

As we navigate an increasingly data-driven world, teaching and training in research data management (RDM) is essential for FAIR (Findable, Accessible, Interoperable, Reuseable) data usage and supply. To have the ability to effectively manage and utilize data, education in data literacy should start from the very beginning. Inspired by this increasing importance of data competencies, members of several NFDI Consortia as well as of the Section Training & Education assembled to establish a platform for data literacy. The formed initiative DALIA (Data Literacy Alliance) is developing this platform as a user-centred knowledge graph based on semantic web technology. These efforts contribute to DALIA's paradigm: "Data literacy for all from the very beginning!"

DALIA promotes learning and teaching of data literacy through the curated provision of open teaching and learning content to empower everyone to collect, analyse, critically evaluate and use data in a FAIR way. The knowledge base can only be successful – i.e., accepted – if quality-assessed informational, teaching, and learning materials are available and if these materials are offered and recommended according to the personal needs of the users. Therefore, the DALIA platform is designed as a single point of entry, which makes suggestions for materials according to individual needs in a heterogeneous and constantly growing landscape of informational, teaching, and learning materials. The users are individuals from academic and adjacent fields, who have different roles (researchers, teachers, learners), various career and RDM experience levels, and come from different scientific disciplines. As a result, the knowledge base will assist users with their questions on the usage (data science) and provision of FAIR data, refer to appropriate teaching and learning materials, and suggest learning paths and best practice documents.

In the early stages of the project (funded by BMBF), the first operating DALIA platform will be provided. After the end of the project, DALIA's operational concept intends the transfer of the platform to the NFDI to keep the knowledge base within the community. For a sustainable operation, the platform has to be established as a basic service in the NFDI (DALIA4NFDI). This will provide a solution for the provision of teaching and training contents in data literacy for all NFDI members and beyond. Therefore, the presentation will highlight and discuss the current stage of the platform development and future steps towards a base service.

The close connection to the NFDI and its sections enables a very strong synergy between the findability, recommendation, and quality assurance of teaching and learning materials, as well as the future development of these materials within the framework of the NFDI consortia. In summary, the project, with its intended results, makes an important contribution to increasing data competency and establishing a data culture in the academic environment.

Keywords: Data literacy, education, training, knowledge base, DALIA

Meet the digital data steward Jarves

Authors:

Tobias Hamann^{1*} and Jonas Werheid²

*Lead presenter

¹ tobias.hamann@wzl-igs.rwth-aachen.de, WZL | RWTH Aachen University | Intelligence in Quality Sensing (IQS)

²WZL | RWTH Aachen University | Intelligence in Quality Sensing (IQS)

Abstract:

Researchers need guidance in their research data management (RDM). While it becomes clear that RDM grows in importance with the amount of data and new methodologies, most researchers are not sure how to manage their data accordingly. Processes, solutions, tools and services are most often not used or even not known. This applies for engineering sciences as well as other disciplines to some extent.

One reason for this is that currently, the “How to” of RDM is either hidden behind abstract frameworks like the data life cycle or accessible through initiatives or helpdesks. However, both solutions lack an immediate nature, meaning that researchers do not receive help at their point of need. Instead, there is effort required to apply RDM and define it beforehand, due to additional tasks that RDM requires and the lack of knowledge researchers have. Therefore, the general structure of the RDM process has to be defined by researchers themselves. Afterwards, RDM has to be applied, resulting in enquiries for solutions or tools to be used in RDM. Lastly, the usage of these solutions or tools has to be figured out by the researchers.

The digital data steward Joined Assistant in Research for Versatile Engineering Sciences (Jarves) aims to address these problems. Jarves supports engineering researchers in their everyday work. It focuses on the reduction of effort and time for RDM. By ordering the RDM activities based on their occurrence in the engineering research process, Jarves provides a structure for RDM in engineering. Based on the research’s general environment, like requirements of funding organisations or institutional boundaries, the next steps, available tools and suitable trainings are recommended. Its partial automation reduces the effort needed in everyday RDM, allowing for seamless (meta)data exchange. While the tool was developed for engineering sciences, researchers in other domains can apply its concepts and possibly transfer them to their needs.

The proposed contribution is set to inform the members of the NFDI about Jarves and its goals, collect feedback and connect to similar solutions in the NFDI. Participants will be introduced to the digital data steward Jarves with a short presentation on the aim and scope of the tool. Afterwards, participants will be able to log in to Jarves, test the tool and collaborate with each other. The participants will be questioned for their possible applications and wishes for Jarves further development. This step aims for the identification of requirements to enhance Jarves' functionality beyond NFDI4Ing, eventually integrating it into Base4NFDI and the EOSC Service portfolio.

Keywords: Engineering, RDM Process, decision support, service interconnection, RDM framework

ASSURED. A training and accreditation service for safe research.

Authors:

Simon Parker^{1*}, Deborah Wiltshire², Vanessa González Ribao¹, Markus Herklotz³ and Wiebke Weber³

*Lead presenter

¹GHGA - German Human Genome-Phenome Archive, Deutsches Krebsforschungszentrum in der Helmholtz-Gemeinschaft, Heidelberg, Germany

²Data Services for the Social Sciences, GESIS - Leibniz-Institut für Sozialwissenschaften, Cologne, Germany

³Social Data Science and AI Lab, Ludwig-Maximilians-Universität München, Munich, Germany

Abstract:

Germany has a large Research Data Centre (RDC) infrastructure that facilitates the analysis of complex and sensitive data. This infrastructure is being further developed through the NFDI to increase the number and range of research communities served. Within this growing infrastructure, there is an increasing need to use sensitive, potentially disclosive, research data responsibly. Researchers wishing to access data and Data Access Professionals working in RDCs require skills to handle these data in an ethical and efficient way.

Many institutions offer generalist data protection training to their staff but this rarely covers key specialist topics such as proper data citation, lawful data sharing, and how to produce non-disclosive research outputs which are essential to achieve the Safe Use¹ of sensitive research data. A survey conducted by the NFDI4Health consortium revealed that over 60% of respondents are interested in learning more about data access and sharing². RDCs may offer dedicated training for Researchers seeking access to the sensitive research data they hold, but a lack of recognition across services means Researchers may have to train multiple times.

To address this situation, we are developing ASSURED³: an adaptable E-Learning training and accreditation system that enables Researchers and Data Access Professionals to achieve a widely-recognised accreditation for working with sensitive research data. This program is being developed collaboratively by institutions across three NFDI consortia; GHGA, KonsortSWD and BERD@NFDI. ASSURED is currently focused upon Researchers and Data Access Professionals in Germany but has the potential to expand in alignment with the objectives of the European Open Science Cloud.

The ASSURED program is divided into two parts; a suite of 'core' modules that explain the key concepts that are indispensable to work with sensitive research data safety. Additional optional 'specialist' modules can then be mandated as required by the RDCs. These 'specialist' modules will cover how to work in a particular secure environment, how to handle different data types, or additional topics related to performing a particular role such as training for Data Access Professionals. The modules are conceived as short and compact learning units. ASSURED will be presented via an online training platform and utilise written material, instructive videos, animations, and interactive quizzes to present the information in a clear and engaging manner. By integrating their 'training status' with an Authentication and Authorisation Infrastructure such as IAM4NFDI, those trained using ASSURED will be able to easily demonstrate their accreditation and have it recognised at multiple RDCs.

As a training and accreditation service, ASSURED is beneficial for the NFDI and the research data landscape as a whole. It supports improving the security models of Research Data Centres and building trust. The ASSURED approach promotes best practices for safe data use and sharing, and fostering career-development within the research data sector.

At the NFDI4Base User Conference, we will introduce our proposal to the NFDI community. We welcome the opportunity to discuss our work with an interested audience to understand

how ASSURED can meet their needs and to open up the project to potential further collaborators.

¹ Restel, K., Depping, R. (2022) Bericht zum 1. NFDI4Health User Survey: Status-quo-Erhebung 2021. Bremen: NFDI4Health, p. 31. <https://doi.org/10.4126/FRL01-006461780>

² Desai, T., Ritchie, F., Welpton, R. (2016). Five Safes: Designing data access for research. Bristol: University of the West of England. <https://uwe-repository.worktribe.com/output/914745>

³ [Assured · Safe research by safe people \(assured-training.org\)](https://assured-training.org/)

Keywords: safe research, adaptable training system, standardised accreditation, sensitive data

Connecting the RDM Helpdesks by a shared Basic Framework

Authors:

Martha Stellmacher^{1*}, Judith Engel^{2*}, Matthias Fingerhuth³, Ralph Müller-Pfefferkorn⁴, Marcus Schmidt⁵, Lukas Weimer⁶, Cord Wiljes⁷, Wolf Zinke⁷

¹martha.stellmacher@slub-dresden.de, SLUB Dresden (NFDI4Culture),

²Universität Bremen - MARUM (NFDI4Biodiversity)

³Landesinitiative für Forschungsdatenmanagement – fdm.nrw

⁴TU Dresden - Center for Interdisciplinary Digital Sciences

⁵ZALF e.V. (FAIRagro)

⁶Niedersächsische Staats- und Universitätsbibliothek Göttingen (TEXT+, Base4NFDI)

⁷Nationale Forschungsdateninfrastruktur (NFDI) e.V.

Abstract

User support is an essential component for the success of an efficient research data infrastructure. Consequently, many consortia of the National Research Data Infrastructure (NFDI) have already established or plan to set up discipline specific helpdesks to support their communities. The technical and organizational foundations of these NFDI helpdesks are diverse and not yet fully interconnected. In addition, the services of generic RDM helpdesks at institutional or state level also need to be included in a helpdesk network that aims to reach researchers at all levels.

Therefore, the working group “RDM Helpdesk Network”[1] proposes a shared framework that efficiently connects the helpdesks within the NFDI as well as with the other RDM helpdesks. In an initial joint workshop of NFDI consortia, RDM state initiatives and local RDM helpdesks[2], we discussed requirements and the potential structure of such a network of helpdesks. In its simplest form, it would provide a central point of contact with a signpost function that refers RDM support requests to the most suitable helpdesk. This is currently being tested in a decentralized approach by close collaboration and joint support cases among NFDI helpdesks of life science and the humanities consortia respectively. The next level would be to standardize the handling of crossdisciplinary requests, e.g. by developing blueprints. Such a decentralized but interconnected network of helpdesks could also efficiently balance workload and may offer first-level generic support for researchers who do not otherwise have access to RDM support. This network could also provide customer support for other NFDI basic services.

These goals require a shared ticketing system and/or interfaces between the existing helpdesks, combined with standardized workflows that build on the experiences of the helpdesks. It includes a shared knowledge base for all helpdesks that is collaboratively created and updated.

At the Base4NFDI User Conference we plan to obtain an overview of support requirements and discuss ideas for a network of helpdesks with users, NFDI consortia and the Base4NFDI basic services.

Keywords: helpdesk, support, ticketing system, consulting

[1] Working Group RDM Helpdesk Network, <https://www.nfdi.de/section-edutrain/working-group-rdm-helpdesk-network/>

[2] Community Workshop: Vernetzung der FDM-Helpdesks, 20./21.11.2023, <https://www.nfdi.de/community-workshop-helpdesk-vernetzung/>

FAIR principles and benchmarking of research software as essential data

Authors:

Jörg F. Unger^{1*}, Tomislav Maric², Jan Linxweiler³ and Bernd Flemisch⁴

*Lead presenter

¹joerg.unger@bam.de, Bundesanstalt für Materialforschung und -prüfung

²maric@mma.tu-darmstadt.de, TU Darmstadt

³j.linxweiler@tu-braunschweig.de, TU Braunschweig

⁴bernd.flemisch@iws.uni-stuttgart.de, Universität Stuttgart

Abstract:

Software is an integral part of research data. The RDA working group 'FAIR for Research Software (FAIR4RS)' has extended the FAIR principles, particularly tailored to research software [1]. The challenge of integrating these general principles into software for numerically solving physics based model is solved by each community and often even each individual researcher with a different approach.

A particular challenge of research software in the simulation domain is its complex dependencies because it is usually built upon many other software contributions. Another challenge for a FAIR implementation is the integration of these modules into simulation workflows that define a sequential order for the execution of individual simulations tasks usually described by a graph, as well as the compute environment. In NFDI4Ing, a comparison of selected workflow tools has been performed [2]. When making research software FAIR, all these levels (individual module, simulation workflow, result of a specific run of the workflow) are to be considered. Several approaches such as RO-Crate with an extension Workflow Run Crate [3], the Citation File Format [<https://citation-file-format.github.io/>], the common workflow language with CWLProv for describing workflow provenance [4], and many more are currently being developed. In NFDI4ING, there are related activities such as a search engine to find research software matching certain attributes [5]. The goal of a working group on the BaseNFDI level would be to establish standards on how FAIR simulation workflows can be developed and shared within the communities, related to both the workflows as well as the results of an individual run. Another requirement is the interoperability with domain-relevant community standards for data exchange. Even though ongoing efforts in our community aim to build standardized file formats, e.g., the VMAP standard for Finite Element Models, implementations are often slightly different, making it a challenge to exchange one interoperable software tool with another one while being able to reproduce the results. From our perspective, benchmarking of tools with well-defined test data is thus mandatory, but a general platform for this purpose is missing. This includes e.g. building an editorial board of experts for defining benchmarks, a platform for hosting codes with appropriate administrative structures to ensure code quality, a tool independent and machine-readable description of benchmark problems, an execution platform for running the benchmarks and tools and resources (e.g. a jupyterhub) to publish/visualize and explore the results within the community. For this purpose, a tool for comparing field data [6] has been developed, and a platform for integrating benchmark tests is planned to be developed. Those challenges are not specific to the engineering domain, and as such, we hope to establish a working group in the base community to build a software benchmarking framework for simulation software to address these challenges in a cross-disciplinary manner.

- [1] N. P. Chue Hong et al., „FAIR Principles for Research Software (FAIR4RS Principles)“, Research Data Alliance, 2021, doi: 10.15497/RDA00068.
- [2] P. Diercks et al., „Evaluation of tools for describing, reproducing and reusing scientific workflows“, Universitäts- und Landesbibliothek Darmstadt, 2023, doi: 10.48694/INGGRID.3726.
- [3] S. Leo u. a., „Recording provenance of workflow runs with RO-Crate“. arXiv, 2023. doi: 10.48550/ARXIV.2312.07852.
- [4] F. Z. Khan et al., „Sharing interoperable workflow provenance: A review of best practices and their practical application in CWLProv“, GigaScience, Bd. 8, Nr. 11. Oxford University Press (OUP), Nov. 01, 2019. doi: 10.1093/gigascience/giz095.
- [5] V. Seibert, A. Rausch, S. Wittek, „Betty’s (Re)Search Engine: A client-based search engine for research software stored in repositories“, Universitäts- und Landesbibliothek Darmstadt, 2023, doi: 10.48694/inggrid.3953.
- [6] D. Gläser et al., „fieldcompare: A Python package for regression testing simulation results“, Journal of Open Source Software, Bd. 8, Nr. 81. The Open Journal, S. 4905, Jan. 31, 2023. doi: 10.21105/joss.04905.

Keywords: [workflows, research software, benchmarking]

RDMTraining4NFDI - Consolidating Training of Research Data Management Skills in the NFDI

Authors:

Rabea Müller^{1*}, Mirjam Blümm², Birte Lindstädt³, Konrad U. Förstner⁴,

*Lead presenter

¹ muellerr@zbmed.de, ZB MED – Information Centre for Life Sciences, Cologne

² mirjam.bluemm@th-koeln.de, Technische Hochschule Köln (TH Köln)

³ lindstaedt@zbmed.de, ZB MED – Information Centre for Life Sciences, Cologne

⁴ foerstner@zbmed.de, ZB MED – Information Centre for Life Sciences, Cologne

Abstract:

The National Research Data Infrastructure (NFDI) aims to foster professional research data management (RDM) skills at universities and research institutions nationwide and support the adoption of the FAIR principles. Achieving this goal requires not only the necessary technical infrastructure, but also individuals who possess broad expertise in both generic and domain-specific aspects of RDM. All of the NFDI consortia meanwhile provide training of RDM skills to their community and are working on further improvement of these offers.

The intended base service "RDMTraining4NFDI" - at the time of the abstract submission in the process of reapplying for the Base4NFDI initiation phase - aims to develop modular teaching concepts and materials that can be used to create training programmes and courses in research data management by testing different training formats. The planned measures will be critically discussed with representatives of all NFDI consortia during a workshop that will take place in July, 2024. As a NFDI basic service, RDMTraining4NFDI would like to make the modular concepts and resources openly available to NFDI consortia and develop certification possibilities to set standards within the NFDI.

The project will support the consortia in transforming the modular concepts and materials into discipline-specific training programmes and running and certifying the corresponding training activities. Aimed at those responsible for training in NFDI consortia, this basic service will provide each training coordinator with the concepts and materials they need. Several formats will be used to prepare, test and evaluate the teaching concepts and materials. These formats are designed to cater to the needs regarding data management in the research community. The learning content and materials will be made freely available through DALIA as open educational resources (OERs) under the CC-BY license and in accordance with FAIR principles. Once established, NFDI certification of RDM training will ensure that the respective courses cover core competencies that are relevant across multiple domains.

Keywords: Training, Data Literacy, RDM

Poster Session

Chair: Lukas Weimer (SUB Göttingen)

The NFDI Data Stewardship Network

Authors:

Daniela Hausen^{1*}, Jens Dierkes^{2*}, and Cord Wiljes³

*Lead presenter

¹hausen@ub.rwth-aachen.de, RWTH Aachen University Library, NFDI4Chem

²dierkes@ub.uni-koeln.de, Cologne University and City Library, NFDI4Health

³cord.wiljes@nfdi.de, German National Research Data Infrastructure (NFDI) Office

Abstract:

The emerging role of *data steward* has become increasingly important for the management of research data (RDM) in the scientific disciplines and large research projects such as Collaborative Research Centres and Clusters of Excellence, in the NFDI consortia, and in infrastructure institutions such as libraries and computing centres. [1,2] Their responsibilities are diverse, including quality control, metadata management, infrastructure development, training, and ensuring compliance with legal and ethical standards. Data stewards often specialise in particular aspects of RDM, depending on the specific needs of their discipline, project, or institutional setting. Despite the diversity of tasks, they often face similar challenges, such as limited resources, a lack of standardised procedures and the need for professional training. [3,4] The DataStew project has identified five prototypical profiles for data stewards, characterised by their competencies, tasks, institutional location, and professional and RDM background. [4]

It is evident that data stewards often have previous experience in research or have acquired some knowledge of RDM. When data stewards are recruited from research, their initial knowledge of RDM and networking experience is mostly rudimentary. This situation highlights the lack of established structures and practices for data stewards to orientate themselves by, which affects their effectiveness and efficiency. Therefore, there is a need to develop and make knowledge accessible, and a strong, evolving community to address important questions about the needs and support required to build such a community. Initial efforts to build a community are being made, for example, by the TU9-UAG FDM. They offer an annual workshop, *Data Stewardship goes Germany* [5], which this year focuses on building a community. Additionally, several entities, such as the section Training & Education and consortia such as NFDI4Health, are working on scaling up the continuing education of data stewards. [6]

In addition, institutionalisation is important to ensure the long-term success of RDM. Most data stewards work on a project-by-project basis, which is not a durable solution and can lead to conflicts. [4] Structures need to be created that provide a stable framework for the evolving roles of data stewards and to meet the challenges of achieving widespread acceptance of NFDI services and standards. With a community of data stewards acting as multipliers, NFDI issues can be disseminated in local settings, and data stewards in universities can be empowered to make effective use of new technologies and disseminate them further locally.

In summary, the current landscape of data stewards in research shows both significant progress and challenges. Establishing clear structures, fostering a vibrant community and institutionalising the role of data stewards are essential steps towards improving the effectiveness of RDM. Therefore, we would like to discuss with you the urgent need that all basic services share: Namely, to increase the awareness and adoption of basic services by the community and to strengthen the commitment of institutions to provide basic services. What actions can we possibly take, e.g. by designing a basic service for data stewardship with a focus on institutionalisation and networking?

[1] Hausen, D., Rosenberg, J., Trautwein-Bruns, U., & Schwarz, A. (2020). Data Stewards an der RWTH Aachen University – Aufbau eines flexiblen Netzwerks. *Bausteine Forschungsdatenmanagement*, (2), 20–28. <https://doi.org/10.17192/bfdm.2020.2.8278>

[2] Curdt, C., Dierkes, J., Helbig, K., Lindstädt, B., Ludwig, J., Neumann, J., & Parmaksiz, U. (2021). Data Stewardship im Forschungsdatenmanagement - Rollen, Aufgabenprofile, Einsatzgebiete: Überblick: 11. DINI/nestor Workshop, 16. und 17.11.2020. *Bausteine Forschungsdatenmanagement*, (3), 70–81. <https://doi.org/10.17192/bfdm.2021.3.8347>

[3] Steinke, B., Hausen, D., Kuberek, M., Hora, M., Kessler, K., Kramer, C., Fuhrmans, M., Grunwald-Eckhardt, L., Hermann, S., Müller-Pfefferkorn, R., Neumann, J., Queitsch, M., & Strötgen, R. (2022). Data Stewards an den TU9-Universitäten – Bestandsaufnahme, Handlungsfelder und Kooperationspotenzial. *Bausteine Forschungsdatenmanagement*, (1), 1–12. <https://doi.org/10.17192/bfdm.2022.1.8364>

[4] Seidlmayer, E.; Hoffmann, F.; Dierkes, J.; Lindstädt, B.; Depping, R.; Förstner, K. U. *Forschung unterstützen - Empfehlungen für Data Stewardship an akademischen Forschungsinstitutionen*; ZB MED Informationszentrum Lebenswissenschaften, Universität zu Köln: Köln, 2023. <https://doi.org/10.4126/FRL01-006441397> (accessed 2023-10-02).

[5] <https://www.dsgg.rwth-aachen.de/cms/~bgnhzs/dsgg/> (last check: 2024-05-24)

[6] <https://indico.cern.ch/event/1408259/timetable/?layout=room#20241021.detailed> (last check: 2024-05-24)

Keywords: Data Steward, Networking, Institutionalisation, Challenges

NFDI4Cat Research Data Management System: A Proposal for Enhanced Data Sharing in Catalysis Research

Authors:

Nadiia Huskova^{1*}, Taras Petrenko¹, Yuliia Dikova¹, Volodymyr Kushnarenko¹, Preston Rodrigues¹, and Thomas Bönisch¹

*nadiia.huskova@hlrs.de

¹High Performance Computing Center Stuttgart (HLRS), Stuttgart, Germany

Abstract:

Efficient collaboration and robust data sharing are important in every research field, particularly in catalysis studies where interdisciplinary collaboration is essential. The National Research Data Infrastructure for Catalysis (NFDI4Cat) addresses these imperatives by developing a centralized research data management (RDM) system. Led by the Fraunhofer Institute For Open Communication Systems in Berlin (FOKUS) and High-Performance Computing Center in Stuttgart (HLRS), one of the NFDI4Cat tasks areas has made a significant effort in developing RDM system for the catalysis-related sciences. However, to fully realize the potential of data sharing between research groups, the central repository must be seamlessly connected with other components.

The central repository, maintained by HLRS, acts as a central hub for storing and organizing a wide range of research data from various domains of catalysis, including biocatalysis, homogeneous catalysis, and heterogeneous catalysis. Beside the data sets themselves, important details about the tools and equipment are also kept. This ensures that the data are well-contextualized which makes it easier for others to reuse and reproduce the results of previous studies.

To increase the findability and understandability of data, NFDI4Cat aims to implement an approach where the storing of data sets in the repository requires only essential metadata, while more comprehensive metadata critical for understanding are prepared within a semantically rich RDF framework. Special attention is paid to the generation and collection of RDF data. In various domains, this task is typically performed manually, which constitutes a significant obstacle for wide adoption of the methodology in the research community. For this purpose, RDF data are produced interactively via an appropriately designed web application. The core of this application is the so-called TRIQ program developed at HLRS, which performs a questionnaire based on specific logical patterns within a reference ontology. The RDF data are stored in a appropriate triple store, and can be explored via a SPARQL endpoint. In addition, the usability of RDF data can be greatly enhanced by converting RDF data sets into HTML pages that can be navigated through a web browser.

This methodology enables high interoperability across diverse data sets and fosters deep insights into the data. In particular, the developed RDM system helps to acquire more detailed information beyond what is usually given by a researcher, as well as describe complex research processes and capture specific knowledge in the field. The software is currently under development and supposed to be a user-friendly solution. Furthermore, this approach integrates data, vocabularies, and ontologies into a unified, interconnected framework.

In conclusion, the infrastructure proposed, not only enhances data management and sharing within the catalysis research community but also holds promise for a broader adoption across various domains. The scalability and interoperability of the system offers a potential approach for improving research data management practices.

Development of Sustainable Aggregating Services for an Improved FAIR Research Software Ecosystem

Authors:

Maxence Azzouz-Thuderoz^{1*}, Madhurima Deb² and Moritz Schubotz²

¹maxence.azzouz-thuderoz@fiz-karlsruhe.de, FIZ Karlsruhe Leibniz Institute for Information Infrastructure, Berlin, Germany.

²FIZ Karlsruhe Leibniz Institute for Information Infrastructure, Berlin, Germany.

Abstract:

Version control systems (VCS), among the various components shaping the research software landscape, have emerged as essential tools in the software development process. Platforms like GitHub and GitLab have gained exponential popularity among scholars by facilitating collaborative development and contributing significantly to the adoption of software engineering practices in academia. These platforms have also made research software more compliant with the FAIR principles (findability, accessibility, interoperability, and reusability). A notable trend is the substantial reliance of academic researchers on VCS platforms, as evidenced by 127,529 Uniform Resource Identifiers (URIs) cited in 385,817 arXiv publications referencing software hosted on platforms like GitHub and GitLab. Similarly, 16,690 of the 44,632 software tools indexed by the mathematical software database swMATH link back to these platforms. However, despite their widespread use, these platforms lack suitable recommender systems for related work, limiting the discoverability of research software.

Moreover, concerns have arisen regarding the inability to host code in desired locations, prompting universities and research institutions to establish their own repository services. This has driven the development and adoption of on-premise solutions. The diverse use of repository platforms—ranging from public GitHub repositories to on-premise GitLab deployments and standalone Git servers—presents additional challenges in monitoring the research software landscape. This diversity hinders the discoverability and reproducibility of research outputs, complicating the curation of metadata and long-term adherence to the FAIR principles.

In response to these challenges, aggregator platforms like swMATH play a crucial role in ensuring the sustainability and accessibility of research software. By curating and organizing software metadata, these platforms enhance the reliability and discoverability of software source code. However, their focus on specific communities often limits their broader adoption, restricting their potential to reach users beyond niche groups.

A promising solution is the establishment of an NFDI Research Software Marketplace, envisioned as an interoperable platform integrated with standardized software metadata catalogs like swMATH. This marketplace could serve as a centralized hub for discovering, accessing, and collaborating on research software projects, regardless of the repository services used. Crucially, it would also support advanced recommender system services tailored to the specific needs of research software developers, closing the gap in discoverability addressed by the FAIR criteria.

Keywords: Research software, FAIR, aggregator, NFDI, recommender systems

Overlaps between NFDI Knowledge Graphs

Authors:

Daniel Mietchen¹

¹daniel.mietchen@mardi4.de, FIZ Karlsruhe — Leibniz Institute for Information Infrastructure

Abstract:

Knowledge graphs are playing various roles within the research infrastructure landscapes, both within NFDI and beyond. This contribution will cover the MaRDI Portal and its overlaps with other knowledge graphs used in NFDI contexts, particularly as part of the Basic Service KGI4NFDI. The potential scope of these overlaps includes multiple dimensions, such as the graph's purpose, content, tech stack, licensing, community, governance, workflows and integration with other Basic Services. Besides reporting on the present state and some of the developments that led to it, the focus of the presentation will be on potential future developments, with special attention to opportunities and challenges for meaningful cross-graph integration.

These opportunities and challenges will initially be discussed from a mathematical perspective and then generalized a bit. For instance, lots of non-mathematical fields use mathematical modeling in some form or another. Should the information about those mathematical models reside (1) primarily in the (non-mathematical) domain-specific knowledge graphs or (2) primarily in mathematical knowledge graphs or (3) in some combination of both? In case (1), the domain-specific graphs would need to be made aware of some key concepts relating to mathematical models, ideally in a way that is broadly consistent across graphs. Conversely, in case (2), the mathematical knowledge graphs would need to be made aware of some key concepts relating to each of these non-mathematical fields, again ideally in a way that is broadly consistent across graphs. Finally, in case (3), the precise tuning might depend on the field, with implications for the overall landscape of NFDI knowledge graphs. Besides mathematical modeling, another case of interest is the history of mathematics, which comes with similar questions and options in that the knowledge could in principle reside primarily on either the mathematical or the historical end (e.g. in NFDI4Memory) but it would need to be connected in a way that makes sense across research fields, i.e. that the history of mathematics is handled in some manner that is consistent with the respective historical knowledge about other fields like medicine, musicology or materials science. Basic Services for terminologies or identifiers might play a key role here.

Zooming out, it is an interesting exercise to consider a matrix of overlaps between different NFDI projects and existing or potential knowledge graphs in their respective domains, and to ponder how information at the intersections should be handled. For instance, information regarding natural products - i.e. chemicals produced by living systems - would fall within scope of NFDI4Chem and NFDI4Biodiversity (and potentially other NFDI projects), of which only the former currently has a knowledge graph, though non-NFDI knowledge graphs already exist for both domains. Similarly, NFDI4Objects and NFDI-MatWerk might share an interest in - and thus perhaps collaborate on knowledge graph representations of - certain types of objects (e.g. coins or tools). For many pairs (or even groups) of consortia, such overlaps can be readily identified, and this contribution aims to stimulate discussions around that, both in general terms and with an eye on how current or future Basic Services could assist.

Keywords: knowledge graphs, transdisciplinarity, mathematics, collaboration

A terminology service via TS4NFDI for BERD@NFDI knowledge graph infrastructure

Authors:

Renat Shigapov^{1*}, Oliver Koepler², and Roman Baum³

*Lead presenter

¹ renat.shigapov@uni-mannheim.de, University of Mannheim, Mannheim, Germany

² TIB - Leibniz Information Centre for Science and Technology, Germany, Hannover

³ ZB MED - Information Centre for Life Sciences, Cologne, Germany

Abstract:

The use of standardised terminologies facilitates interoperability and data integration, making research data FAIR (Findable, Accessible, Interoperable, Reusable). Terminologies are often made available via terminology services through graphical user interfaces (GUIs) and APIs. Several domain-specific terminology services already exist within the NFDI consortia such as SemLookP [1, 2], BiodivPortal [3], and NFDI4Chem Terminology Service [4, 5]. In the social sciences, services such as the GESIS Controlled Vocabulary Service [6] which is based on SKOSMOS and the STW Thesaurus for Economics [7] hosted by the ZBW are notable examples. However, a common terminology service for BERD@NFDI does not yet exist. BERD@NFDI is developing a knowledge graph infrastructure (KGI) for German company data, transforming analogue books into multiple knowledge graphs through optical character recognition, automatic structuring, and semantification [8, 9, 10]. These knowledge graphs are a part of distributed KGI4NFDI [11, 12]. A common terminology service for BERD@NFDI would significantly enhance the efficiency of semantification and linking processes.

One of the aims of BERD@NFDI is to ensure that data management adheres to the highest standards of integrity and accessibility. To achieve this, we are setting up a terminology service for BERD@NFDI using the Terminology Service Suite (TSS) [13]. The widgets of the TSS could handle multiple backends such as SKOSMOS, OLS and OntoPortal. This ensures that the newly developed terminology service for the BERD community meets the standards of other terminology services, promoting interoperability with services such as the GESIS Controlled Vocabulary Service, which utilises SKOSMOS as a backend system. The new terminology service can be seamlessly integrated into the BERD knowledge graph infrastructure. The GUI of the terminology service, constructed using TSS widgets, allows for straightforward embedding of terminology service data into other applications or services within the Business, Economic, and Related Data domains. Additionally, terminologies can be accessed via the provided APIs, ensuring wide usability and integration.

By implementing a standardised terminology service, BERD@NFDI not only enhances the semantification and linking of knowledge graphs but also ensures that these processes are aligned with the broader NFDI goals. The terminology service supports the overall mission of the NFDI by promoting data interoperability, thereby advancing research capabilities across various domains.

Keywords: terminology service, knowledge graph, FAIR research data

[1] Baum, R. (2024, Januar 18). SemLookP widgets - Embed ontologies in service applications. Zenodo. doi.org/10.5281/zenodo.10529181

[2] <https://semanticlookup.zbmed.de/>

[3] <https://biodivportal.gfbio.org/>

[4] Strömert, P., Limbachia, V., Oladazimi, P., Hunold, J., Koepler, O., Towards a versatile Terminology Service for empowering FAIR research data: Enabling ontology discovery,

design, curation, and utilization across scientific communities. Studies on the Semantic Web, Vol. 56 Knowledge Graphs: Semantics, Machine Learning, and Languages. IOS Press; 2023. [doi:10.3233/ssw230005](https://doi.org/10.3233/ssw230005)

[5] <https://terminology.nfdi4chem.de>

[6] <https://lod.gesis.org/en/>

[7] <https://zbw.eu/stw>

[8] Shigapov, R. (2022, November 28). Knowledge graphs in BERD and in NFDI. Focused Tutorial on Capturing, Enriching, Disseminating Research Data Objects. Use Cases from Text+, NFDI4Culture and BERD@NFDI, Mannheim and online. Zenodo. doi.org/10.5281/zenodo.7373258

[9] Kamlah, J., Schmidt, T., & Shigapov, R. (2022, November 28). Extracting research data from historical documents with eScriptorium and Python. Focused Tutorial on Capturing, Enriching, Disseminating Research Data Objects. Use Cases from Text+, NFDI4Culture and BERD@NFDI, Mannheim and online. Zenodo. <https://doi.org/10.5281/zenodo.7373135>

[10] Kamlah, J., & Shigapov, R. (2023, Mai 5). The German Production Pipeline: Mannheim - OCR & Knowledge Graphs. Zenodo. <https://doi.org/10.5281/zenodo.7900133>

[11] Rossenova, L., Schubotz, M., & Shigapov, R. (2023). The Case for a Common, Reusable Knowledge Graph Infrastructure for NFDI. Proceedings of the Conference on Research Data Infrastructure , 1. <https://doi.org/10.52825/cordi.v1i.266>

[12] Rossenova, L., Shigapov, R., Schubotz, M., Limani, F., Zapilko, B., & Förstner, K. U. (2024). KGI4NFDI - Knowledge Graph Infrastructure for the German National Research Data Infrastructure. Zenodo. <https://doi.org/10.5281/zenodo.13118749>

[13] <https://github.com/ts4nfdi/terminology-service-suite>

Facilitating Comprehensive Metadata Capture and Validation in Data Repositories (nmrXiv) through Terminologies and Terminology Service Suite Widgets

Authors:

[Venkata Nainala](#)^{1*}, Noura Rayya², Christoph Steinbeck², Oliver Koepler³

¹ chandu.nainala@uni-jena.de, Friedrich Schiller University, Jena, Germany

² Friedrich Schiller University, Jena, Germany

³ TIB - Leibniz Information Centre for Science and Technology, Hannover, Germany

Abstract:

Analytical chemistry, one of the oldest scientific disciplines, integrates methods from physical, inorganic, and organic chemistry. The standard workflow begins with formulating a research question and devising experiments, methodologies, and surveys to assess the hypotheses. Experiments are conducted, and samples are treated using established or novel methods, with the resulting data being meticulously recorded. However, the metadata captured during these experiments often needs more details for reproduction and is frequently not machine-readable. A survey conducted by Herres-Pawlis et al. revealed that only 42% of participants described their collected data with metadata [1]. This partial or nonexistent annotation leads to misinterpretation and the loss of time and resources, ultimately hindering future research progress.

Data repositories face the challenge of enabling user-friendly yet high-quality metadata capture. Ensuring that metadata is both comprehensive and easy for researchers to input requires robust terminologies and seamless integration of annotation tools. In this era of large language models (LLMs), the quality of data is paramount, as LLMs rely heavily on high-quality, well-annotated data to generate accurate and reliable insights. Currently, analytical data repositories encounter significant limitations in data annotations, primarily due to insufficient or incorrect terms available in existing terminologies. This results in missing or inaccurate annotations, compromising the quality and reliability of the metadata. For instance, the ontology nmrCV [2], essential for NMR metadata, has several issues that require attention. Domain-independent problems include structural inconsistencies, incorrect annotations, incomplete definitions, and outdated documentation. Domain-specific problems pertain to the NMR field and include inaccuracies in definitions, misaligned hierarchies, and the inclusion of non-NMR entities. To mitigate these issues, collaborative efforts with NFDI4Chem TA6 have been undertaken to enhance the terminologies. This work focuses on improving properties, annotations, and documentation, specifically concerning NMR solvents, calibration compounds, and instrument manufacturers. The goal is to ensure that the ontology accurately represents NMR metadata and aligns with other specialised ontologies such as ChEBI.

On the other end, the lack of effective integration of rich, terminology-driven form widgets within data repositories exacerbates the issue. The nmrXiv data repository, developed as part of the NFDI4Chem initiative, employs the ontology data of the NFDI4Chem terminology service [3, 4] during the data provisioning phase via widgets provided by the Terminology Service Suite [5] from the Terminology Services 4 NFDI (TS4NFDI) base service. This integration allows

submitters to annotate their data with metadata in a semi-automatic manner, significantly enhancing both metadata quality and user experience. Additionally, this enables nmrXiv to actively learn from these new annotations by engaging with data submitters. This iterative process will refine the knowledge graph, with the ultimate goal of achieving fully automated metadata annotation.

Keywords: Metadata Annotation, Analytical Chemistry, NMR Ontology, Terminology Services, Data Repositories

[1] Herres-Pawlis, S., Koepler, O., & Steinbeck, C. (2019). NFDI4Chem: Shaping a Digital and Cultural Change in Chemistry. *Angewandte Chemie International Edition*, 58(32), 10766-10768. <https://doi.org/10.1002/anie.201907260>

[2] FAIRsharing.org: nmrCV; Nuclear Magnetic Resonance Controlled Vocabulary, DOI: [10.25504/FAIRsharing.xm7tkj](https://doi.org/10.25504/FAIRsharing.xm7tkj)

[3] Strömert, P., Limbachia, V., Oladazimi, P., Hunold, J., Koepler, O., Towards a versatile Terminology Service for empowering FAIR research data: Enabling ontology discovery, design, curation, and utilization across scientific communities. *Studies on the Semantic Web, Vol. 56 Knowledge Graphs: Semantics, Machine Learning, and Languages*. IOS Press; 2023. [doi:10.3233/ssw230005](https://doi.org/10.3233/ssw230005)

[4] <https://terminology.nfdi4chem.de>

[5] <https://github.com/ts4nfdi/terminology-service-suite>

Basic services to the community – facilitating the application of basic services through interactive case study textbooks

Authors:

Melanie Seltmann^{1*}, Evgenia Samoilova², Hannes Schnaitter¹, Philipp Schneider¹, Maria Chlastak³, Ulrike Lucke², Sonja Schimmler^{4, 5}

*Lead presenter

¹melanie.seltmann.1@hu-berlin.de, Humboldt-Universität zu Berlin

²Universität Potsdam

³Gesellschaft für Informatik

⁴Technische Universität Berlin

⁵Fraunhofer FOKUS

Abstract:

Data and software literacy are essential for researchers, data management professionals, and data stewards across various disciplines.¹ Base4NFDI addresses these aspects by providing support in targeted areas, such as for example creating and applying data and software management plans. In this contribution, we present an instructional design based on discipline-specific case studies and interactive textbooks to train users in tools and solutions of the basic services. This approach aims to facilitate both easier adoption within as well as greater involvement from subject-specific communities. The design is relevant for Base4NFDI, as researchers and practitioners need to see how these basic services are applied to concrete problems and research questions to bridge the gap between theoretical knowledge and practical application. The instruction design (Samoilova et al., 2024) was developed for the Berlin-Brandenburg Competence Centre for Data Literacy QUADRIGA² (Buchholz et al., 2024) for researchers in the area of Digital Humanities and Public Administration, yet it can also be adapted for use in other disciplines.

The instructional design focuses on transforming real research practices into online interactive educational resources available on-demand and asynchronously. This approach aims to make learning more relevant, engaging, and applicable to real-world research scenarios. While scalability and standardization of basic services are crucial for cross-disciplinary applicability, case studies (Foran, 2001) coalesced with a problem-based learning approach (Kay et al., 2000) can present solutions/tools at the center of basic services within carefully designed problems/research questions, enabling users to learn about the tool based on real-life authentic tasks. Such case studies can illustrate the use of services in concrete communities, highlighting specific demands, challenges and potential solutions at various levels, including research-specific, organizational, and technical aspects. For example, one such case study illustrating a coherent data flow across multiple Base4NFDI services for humanities, historians, information scientists, etc. could be: identifying letters in a collection (PID4NFDI), classifying

¹ Contributor roles according to CRediT: Melanie Seltmann (Writing – original draft), Evgenia Samoilova (Methodology/Writing – original draft), Hannes Schnaitter (Methodology/Writing – original draft), Philipp Schneider (Writing – original draft), Maria Chlastak (Writing – review & editing), Ulrike Lucke (Conceptualization/Methodology/Writing – review & editing), Sonja Schimmler (Conceptualization/Writing – review & editing).

² QUADRIGA is funded as part of the German Federal Ministry of Education and Research's guideline for funding projects to establish data competence centers in science under the references 16DKZ2034A, 16DKZ2034D, 16DKZ2034F, 16DKZ2034G and 16DKZ2034H.

the letters (TS4NFDI), creating OCRs of the letters and transforming the data and metadata (Jupyter4NFDI) and finally publishing the results within the knowledge graph (KGI4NFDI).

The training format is delivered using interactive textbooks, that can be implemented as Jupyter books (Chen & Asta, 2022). The hierarchical structure, where the smallest unit is a chapter section, enables users—both as learners and as instructors using these materials for further trainings—to easily navigate and have a high level of control over the learning experience. Embedded interactive exercises with feedback ensure engagement and learning on a concrete task. Additionally, Jupyter books allow to easily modify or re-structure the content, including when collaborating. For instance, each chapter or section of the textbook could be developed by the specific Base4-project in collaboration with various partner-projects. Discipline-specific communities can also incorporate materials from Base4NFDI into their own training programs.

References

- Buchholz, B., Lucke, U., Schimmler, S., Bakels, J.-H., Bernoth, J., Fischer, F., Gieseke, L., Grotkopp, M., Hiltmann, T., Krupka, D., Loist, S., Meisner, J., Mischke, D., Möser, R., Neuroth, H., Petras, V., Schmeling, J., Söring, S., & Trilcke, P. (2024). Umsetzungskonzept QUADRIGA: Berlin-Brandenburgisches Datenkompetenzzentrum für Digital Humanities, Verwaltungswissenschaft, Informatik und Informationswissenschaft. *Zenodo*. <https://doi.org/10.5281/zenodo.10805016>
- Chen, E., & Asta, M. (2022). Using Jupyter Tools to Design an Interactive Textbook to Guide Undergraduate Research in Materials Informatics. *Journal of Chemical Education*, 99(10), 3601–3606. <https://doi.org/10.1021/acs.jchemed.2c00640>
- Foran, J. (2001). The Case Method and the Interactive Classroom. *Thought & Action*, 17(1), 41–50.
- Kay, J., Barg, M., Fekete, A., Greening, T., Hollands, O., Kingston, J. H., & Crawford, K. (2000). Problem-Based Learning for Foundation Computer Science Courses. *Computer Science Education*, 10(2), 109–128. [https://doi.org/10.1076/0899-3408\(200008\)10:2;1-C;FT109](https://doi.org/10.1076/0899-3408(200008)10:2;1-C;FT109)
- Samoilova, E., Sluyter-Gäthje, H., Skorinkin, D., Schnaitter, H., Trilcke, P., & Lucke, Ulrike. (accepted). Forschungsgeleitete Vermittlung von Datenkompetenz. Mediendidaktische Aufbereitung von Fallstudien zu Bildungsangeboten. *Hochschule 2034*.

Keywords: Data Literacy, Case Studies, Interactive textbooks, Jupyter Books

<https://events.gwdg.de/event/658/abstracts/> -> deadline 29th May 24

Might best fit into that one: Basic Services, Basic Service incubators and networking activities (e.g. How are existing Basic Services developing in Initialisation and Integration stages? How can interoperability of Basic Services be ensured?)

Authors/contributors: DataPLANT infrastructure team, ...

Software repository for the NFDI

Authors:

Dirk von Suchodoletz^{1*}, Timo Mühlhaus, Christoph Garth², and Björn Usadel³

***Lead presenter**

¹dsuchod@uni-freiburg.de, DataPLANT, University of Freiburg

²RPTU Kaiserslautern

³HHU Düsseldorf

Abstract:

Software is tightly linked to data management and its development constitutes a fundamental aspect within most NFDI consortia. The collaboration among contributors, whether local or global, spanning from individual developers and researchers to interdisciplinary and cross-institutional cooperation, necessitates appropriate services and working environments. Thus, a key objective is to establish common public services to facilitate collaborative endeavors across the research community. Furthermore, beside software many digital objects utilized in research stand to benefit significantly from versioning and traceability throughout the duration of a project. Provenance tracking plays a critical role in ensuring the FAIR (Findable, Accessible, Interoperable, and Reusable) principles are upheld in both code and data management practices.

We suggest to utilize GitLab for this purpose as it is widely used in many institutions albeit not as a public service available to all NFDI participants and member institutions. Services like OpenCode (<https://opencode.de/en>) already exist for the public sector and provide a showcase for a well visible common initiative. The NFDI as a major driver of RDM in Germany should provide such a common service and well visible platform or at least moderate the creation of such to fulfill its mission.

For the NFDI to effectively control the software development processes and ensure its long-term sustainability, it is essential to thoroughly test the fitness of the available offerings, business and operation models and make informed decisions regarding the most suitable license. Linking to the NFDI AAI will streamline access and facilitate user authentication, thereby enhancing ease of use. Identifying suitable service providers (within or outside the NFDI) is crucial for obtaining the necessary infrastructure support which includes basic provisioning of storage and computational resources, as well as allocating resources for continuous integration and development processes. Thus, developing a sustainable business and refinancing model is paramount to ensure the long-term viability of such a service. Further, establishing a robust support and training infrastructure is essential for providing assistance to users and ensuring they can effectively utilize the resources.

Various points should be discussed, defined to create such a service inside or outside the Base4NFDI initiative:

- Suggest a proper working group in the Common Infrastructure section to advance this topic - most probably the well established RSE is the best fit
- Define the role of various types of software within the NFDI
- Create a new service from scratch modeled after the NFDI requirements
- Extend some existing service for the whole NFDI community (and beyond)
- Define the operation and business models to allow both broad participation and solid and long-term operation
- Deploy the Base4NFDI service proposal procedures to achieve this goal (or modify, extend them to

GitLab4NFDI (naming suggestion for such a service) can become a solution designed to establish a centralized platform for versioning and sharing research software and datasets across the NFDI and beyond. It would foster better collaboration and the support of the OneNFDI spirit. In the end, establishing a robust support and training infrastructure is essential for providing assistance to users and ensuring they can effectively utilize the resources.

Cloud based flexible service infrastructure stack for the NFDI

Authors:

Jonathan Bauer^{1*}, Marcel Tschöpe¹, Paul Chr. Schnürle¹, Julian Weidhase¹, and Christian Garth², and Timo Mühlhaus²

*Lead presenter, jonathan.bauer@rz.uni-freiburg.de

¹DataPLANT, University of Freiburg

²RPTU, DataPLANT

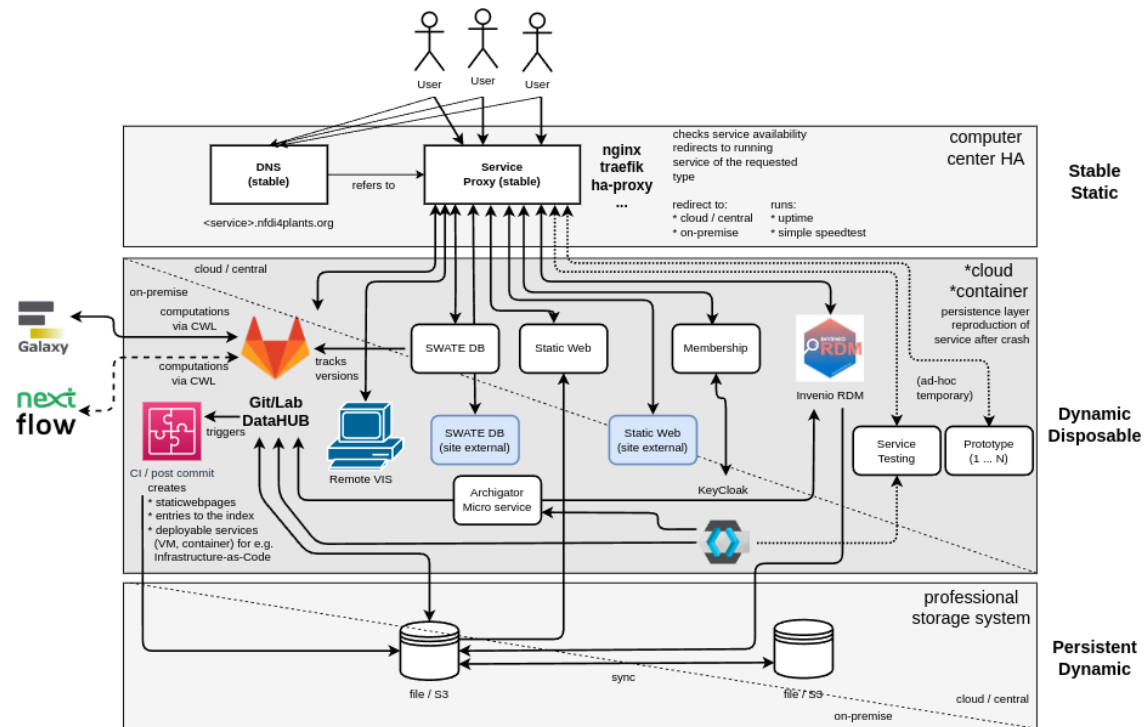
Abstract:

A core objective of the DataPLANT consortium is to provide a range of tools and services to foster its community in research data management. The offered science gateway-base for data management and publication workflow-as well as the various tool backends are designed as a set of flexible cloud-based (micro) services. The setup aims at both on-premises installations and future integration into a shared NFDI infrastructure. We will present the DataPLANT DataHUB, which provides various RDM workflows to support research data scientists at different stages of the data lifecycle - from development to publication of the results obtained. In such setups storage is usually the most local resource to consider because of the amounts of data, data protection and security considerations. Computations on the data should be done near to the storage, to achieve this remote desktop infrastructure and remote visualization would be helpful. [Fig. 1]

The aim is to exchange ideas with other NFDI consortia on the services required for RDM and the principles for service development and deployment. These considerations can be used as input for joint infrastructure development, e.g. in the context of the NFDI Common Infrastructure section or Base4NFDI. Over the past two and a half years, the DataPLANT team has developed a set of software and system components that provide services to the basic plant research community. The set of tools and microservices that have been developed and evolved to date have focused on extending the existing digital landscape of the typical plant scientist. The core services focus on data management, versioning, sharing and publishing. All services are designed as portable modules that can be integrated in the larger base infrastructure. A reverse proxy automatically manages SSL certificates for new services. It also greatly increases the flexibility of the entire infrastructure by being able to quickly change the backend services that respond to the static user-facing service hostnames. Other base services include a central authentication service and monitoring services to detect service outage and to gather performance data.

The development of applications and tools to support bottom-up community-driven research data management requires the involvement of several parties. During the development of the services, we agreed on design principles to provide high-level guidance and a set of criteria for creating desirable and maintainable applications. In DataPLANT, tool development is always motivated by community requirements, conveyed by researchers, e.g., through data stewards, to developers. Developments in DataPLANT follow an incremental and iterative approach, ensuring commitment and alignment of expectations of all stakeholders. Another aim of the service development is to enable both central and local

installation of services without divergent implementations. Thus, we hope to encourage adoption by other communities and integration into a future NFDI service infrastructure.



Version 2024-05-29

[Fig. 1] The actual service stack of DataPLANT in layer oriented representation.

IAM4NFDI – JARDS Incubator

Authors:

Marius Politze^{1*}, Marcel Nellesen² and Camilla Lummerzheim³

*Marius Politze

¹politze@itc.rwth-aachen.de, RWTH Aachen University

²RWTH Aachen University

³RWTH Aachen University

Abstract:

IAM4NFDI supports a variety of different Incubator projects, these are small integration and consulting projects, in which the IAM4NFDI team supports communities in implementing AAI use cases. In this talk we would like to showcase one of these incubator projects “JARDS”.

JARDS is a widely used platform where researchers can apply for resources such as storage space or computation time in the context of NFDI4Ing, NFDIMatWerk and NHR. The applications are reviewed in a scientific-led, multi-stage review process. After the approval, the researchers gain access to the resource and can manage the project members within the RegApp. To achieve this, information about approved projects and users is required within the RegApp. This role can be fulfilled by an external Attribute Authority, which allows the RegApp to gain additional information from a secondary system, like JARDS.

Within the Incubator project the RegApp, as one of four CAAI solutions, is setup for logging in to JARDS. Additionally a result of the project is, that it is now possible to allow attribute queries from the community AAI solution to JARDS through a well-defined interface using the SAML protocol. This thereby allows the enrichment of the user information within the RegApp or connected services. JARDS thereby serves as a demonstrator on how to integrate external applications as Attribute Authorities into the community AAI.

The results of the incubator proposal can directly be used within the production system to improve the connection between the systems while at the same time increasing the security of the solution. Similar solutions could also easily be adapted in other services within the NFDI consortia, which could benefit from enriching the data that is provided by the RegApp.

Keywords: IAM4NFDI, Identity & Access Management, AAI, Community AAI, Incubator, JARDS, HPC, NHR, NFDI4Ing

BITS - a Use Case for Terminologies in Earth System Sciences

Authors:

Claudia Martens^{1*}, Anette Ganske², and Alexander Wolodkin³

*Lead presenter

¹martens@dkrz.de, German Climate Computing Center [DKRZ]

²Leibniz Information Centre for Science and Technology University Library [TIB]

³Senckenberg – Leibniz Institution for Biodiversity and Earth System Research [SGN]

Abstract:

The exponential growth of data due to technological developments along with an increased recognition of research data as relevant research output during the last decades substantiates fundamental challenges in terms of interoperability, reproducibility and reuse of scientific information. Being cross-disciplinary at its core, research in Earth System Science comprises divergent domains such as Paleontology, Marine Science, Biodiversity Research, Atmospheric Sciences and Molecular Biology. Furthermore, different types of data such as observation and simulation data and a huge variety in data sizes (from Climate Model Output to Images of Specimen) exemplifies the challenges of implementing particularly the I and R in FAIR data principles. Within the various disciplines, distinct methods and terms for indexing, cataloguing, describing and finding scientific data have been developed, resulting in a large amount of controlled Vocabularies, Taxonomies and Thesauri. However, given the semantic heterogeneity across scientific domains (even within the Earth System Sciences), effective utilisation and (re)use of data is impeded while the importance of enhanced and improved interoperability across research areas will increase even further.

The BITS Project¹ (BluePrints for the Integration of Terminology Services in Earth System Sciences) aims to address the inadequate implementation of encoding semantics by establishing a Terminology Service that may serve the whole Earth System Science Community on national, european and international level. This TS will be developed based on the existing Terminology Service² of the TIB, supplemented by an ESS Collection³ that already contains relevant terminologies for Earth and Environmental Sciences and to which further relevant terminologies will be added. The implementation of this TS within two data repositories (World Data Center for Climate⁴ at the German Climate Computing Center⁵ and a Data Collection at Senckenberg - Leibniz Institution for Biodiversity and Earth System Research⁶) will showcase the benefits regarding e.g. enhanced and improved discoverability of research products. As all BITS Project Partners are involved in several NFDI activities (on section level as well as within NFDI consortia), we aim to foster collaboration activities across services and infrastructures: by extending the already existing cooperation of the TIB TS and TS4NFDI, or by advocating output of NFDI (BASE and Consortia) services within the wider Earth System Science Community. Within NFDI4Earth the Interest Group on FAIR Earth System Science Terminologies (FAIR ESST) was established, a.o. to monitor the BITS Project from a domain specific perspective. To sum up: we believe that the BITS Project may serve as a Use Case for the interaction of basic and thematic services within the NFDI (and beyond). How exactly this interaction is done (by using widgets offered by TS4NFDI), how

¹ <https://projects.tib.eu/bits/home>

² <https://terminology.tib.eu/ts>

³ <https://terminology.tib.eu/ts/collections?col=ESS>

⁴ <https://www.wdc-climate.de/ui/>

⁵ <https://www.dkrz.de/en>

⁶ <https://www.senckenberg.de/en/>

this will enhance the discoverability of research output (within the Earth System Sciences) and how this may contribute to further standardisation (e.g. for CF conventions) will be part of our proposed presentation/contribution.

Keywords: Terminologies, Earth System Sciences, interoperability, metadata, discoverability

Semantic linked data and metadata storage made easy with LARAsuite.

Authors:

Mark Doerr^{1*} [ORCID: 0000-0003-3270-6895]

*Lead presenter

¹mark.doerr@uni-greifswald.de, Inst. of Biochemistry, University Greifswald

Abstract:

The LARAsuite (<https://gitlab.com/larasuite>) is modern scalable generic open source **Research-Data-Management (RDMS)** system, that combines features of a **Labor-Information-Management-System (LIMS)** system and an **Electronic Lab Notebook (ELN)** in a new fashion, intended to dramatically reduce the input of users and shielding a lot of the the complexity of linked data generation through automation.

Its is based on a dockerised microservice architecture that is easy deployable on a kubernetes cluster or docker compose environment. The LARA suite is designed to dynamically scale horizontally to capture changing demands of users. A very important feature of the LARA suites is its support for automated linked metadata annotation of all generated data and its extensive API, making it suitable for AI and Machine Learning applications that support automated model generation.

All metadata contain links to related domain ontologies (using concepts from EMMO as top-level ontology) and is captured in a triple/quadstore (OpenLink Virtuoso), forming an automatically generated knowledge graph. This Service exposes a SPARQL endpoint, which can be used for extensive (federated) queries and logic reasoning to derive implicit knowledge.

The LARAsuite is a generic framework that addresses most natural sciences, like physics, chemistry, biology and related disciplines. It is developed as a world-wide open source project of a small group of enthusiasts. The development is also interacting with NFDI4Cat (esp. for complex Biocatalysis tasks) and NFDI4Chem (reaction planning) – its power is the integration with SiLA based lab automation systems.

A short demonstration of some selected features is planned.

Keywords: Research-Datata-Management-Sytem, Electronic-Lab-Notebook (ELN), Labor-Information-Management-System (LIMS), Semantic Metadata, Lab Automation

Base4NFDI in the German National Research Data Infrastructure (NFDI) e.V. Association

Authors:

Cord Wiljes^{1*}, Stefanie Fuchsloch², Lukas Jansen², Anne Lehmann², and Hendrik Seitz-Moskaliuk²

*Lead presenter

¹cord.wiljes@nfdi.de, Nationale Forschungsdateninfrastruktur (NFDI) e.V.

²Nationale Forschungsdateninfrastruktur (NFDI) e.V.

Abstract: The goal of the German National Research Data Infrastructure (NFDI) e.V. association is to create and advance a comprehensive and sustainable infrastructure for research data management that supports researchers of all disciplines in the efficient management of their research data. NFDI's 26 consortia, which are collaboration of institutions within a research area, create services tailored to their communities. In NFDI's five sections, the consortia and additional partners work together on cross-disciplinary solutions that are relevant to all disciplines.

The basic services initiative Base4NFDI is a joint initiative of all 26 consortia with the aim of creating basic services relevant to all scientific disciplines. As a division of the NFDI association, Base4NFDI benefits from the strong governance and long-term sustainability of the NFDI association. The 295 member organisations of the NFDI, including most German universities and many research institutes, are committed to applying the solutions offered by NFDI, and to provide valuable feedback.

Base4NFDI plays a crucial role in NFDI. Building on the broad interconnected network of NFDI, Base4NFDI is ideally suited to meet the basic RDM needs of researchers in an efficient manner and to provide a sustainable backbone for the disciplinary solutions created by the consortia.

Keywords: NFDI, Base4NFDI, infrastructure, research data management, rdm

A Cloud-Native Virtual Bioimage Analysis Research Desktop (BARD) for Deployment of Containerised Bioimage Tools on Kubernetes

Authors:

Yi Sun^{1*}, Christian Tischer², Jean-Karim Heriche³

¹yi.sun@embl.de , ²christian.tischer@embl.de , ³heriche@embl.de

^{1,2,3} European Molecular Biology Laboratory (EMBL), Heidelberg, Germany

Abstract:

Bioimage analysis research workflows typically require the integration of various software tools and demand substantial computational power along with high levels of interactivity. These workflows often create inconsistent results due to specific software and operating system dependencies, a challenge that is especially obvious as computationally intensive methods like deep learning become increasingly prominent. Traditional high-performance-computing (HPC) clusters have significant compute power but often fall short in supporting interactive tools, which limits their suitability for bioimage analysis. To overcome these limitations, researchers frequently opt for large cloud-based virtual machines with remote desktop functionality. While this solution addresses some challenges, it faces obstacles in deploying and maintaining the reproducibility of image analysis workflows, particularly when dealing with conflicting software requirements and when configuring systems to utilise specialised hardware like GPUs.

To address these issues, we develop *BARD*, a kubernetes based virtual bioimage analysis research desktop service, leveraging the abcdesktop.io [<https://www.abcdesktop.io/>] project. *BARD* is a unified, containerized, cloud-native platform designed for both intensive computational demands and interactive workflows. It allows bioimage analysts to quickly use a personal cloud-based desktop capable of handling complex computations and image processing tasks without local hardware limitations.

By containerizing all software along with its specific versions and necessary dependencies, *BARD* creates a self-contained computational environment for software that addresses common issues of software incompatibility. This ensures that the environment remains consistent across various computing platforms, allowing researchers to reproduce identical experimental results from any computer, anywhere. Additionally, by leveraging Kubernetes, *BARD* offers a resource-efficient alternative to virtual machines. Through Kubernetes, *BARD* not only reduces the consumption of physical resources, such as CPU and memory, but also streamlines the software deployment process.

Keywords: kubernetes, bioimage analysis, cloud, container

Enhancing FAIR Research Data Management: advancing fine-grained data citation by PIDs for dataset elements.

Authors:

Janete Saldanha Bach^{1*} and Peter Mutschke²

*Lead presenter

¹ janete.saldanhabach@gesis.org, GESIS – Leibniz Institute for the Social Sciences

² GESIS – Leibniz Institute for the Social Sciences

Abstract:

Persistent Identifiers (PIDs) at the study or dataset level are insufficient for addressing the complexity of data management in research. The lack of granularity citation at the level of inline data objects, such as individual survey variables, qualitative data files, and even smaller data points, leads to ambiguities in data citation, inadequate metadata, and challenges in data discovery and reusability.

The PID registration service introduced by KonsortSWD [1], part of the German National Research Data Infrastructure (NFDI) [2], significantly advances the granularity of PIDs. This service supports the assignment of PIDs to these finer dataset elements and ensures accurate data citation since researchers usually use only a subset of the elements in a dataset. Data citation practices vary widely [3], and researchers often do not follow a standard such as the Data Citation Principles (Joint Declaration of Data Citation Principles [4]). The service boosts data referencing practices and improves research transparency and integrity. It also ensures adherence to FAIR principles by enabling precise referencing of individual data elements.

Beyond benefits for human access, such as how researchers can effortlessly find and cite data, Research Data Centres (RDCs) also benefit from PIDs in data governance activities and their services. PIDs' machine-actionable features have potential advantages, including citation tracking and impact measurement, linking articles using the same dataset elements, increasing data findability, and interoperability enhancement. It empowers the RDC's authority by demonstrating a commitment to best practices, extending its reputation within the research community by adopting PIDs at multiple granularity levels, such as recommended by the European Open Science Cloud (EOSC) PID policy [5]. The service facilitates streamlined FAIR research data management by integrating with established metadata standards [6] [7], fostering reproducibility. A single data point, the PID, enhances data citation, reuse, and direct access for automated access (i.e., by a computer program and under some requirements) [8], [9].

In terms of implementation, the technical solution [10] [11] employs the ePic API [12] and relies on the Handle standard [13]. Tests PIDs have been successfully applied to diverse datasets, including survey variables from GESIS and the German Institute for Economic Research (DIW). Tests are also currently taking place at the German Center for University and Science Research (DZHW) in 2024 and are planned at Qualiservice in 2025.

The PID service's design and operational framework align with broader NFDI initiatives such as PID4NFDI. PID4NFDI sets the foundation for a unified and FAIR-enabled PID harmonization towards NFDI-wide operation research data infrastructure. The PID registration service is necessary for harmonized services within the research data infrastructure across the NFDI consortia. It addresses critical gaps in data management by providing a stable and permanent way to reference and cite fine-grained data elements. This capability is crucial for ensuring that researchers can directly access specific data points without navigating entire datasets, facilitating more effective data reuse and verifying research findings. Moreover, by

serving as a base service under the umbrella of PID4NFDI, KonsortSWD's PID service provides a scalable framework that can adapt to many domains across the NFDI.

Keywords: Persistent Identifiers (PIDs), Granular Data Citation, Research Data Services - Technical Infrastructure, Social Sciences - Survey Variables, Dataset elements.

-
- [1] KonsortSWD (Consortium for the Social, Behavioural, Educational and Economic Sciences) is funded by the National Research Data Infrastructure (NFDI) <https://www.konsortswd.de/>
 - [2] German National Research Data Infrastructure (NFDI): <https://www.nfdi.de/>
 - [3] Gregory, K., Ninkov, A. B., Ripp, C., Roblin, E., Peters, I., & Haustein, S. (2023). Tracing data: A survey investigating disciplinary differences in data citation (Version v1). Zenodo. <https://doi.org/10.5281/zenodo.7555266>
 - [4] Data Citation Synthesis Group: Joint Declaration of Data Citation Principles. Martone M. (ed.) San Diego CA: FORCE11; 2014 <https://doi.org/10.25490/a97f-egy>
 - [5] European Commission, Directorate-General for Research and Innovation, Hellström, M., Heughebaert, A., Kotarski, R. et al., A Persistent Identifier (PID) policy for the European Open Science Cloud (EOSC), Publications Office, 2020, <https://data.europa.eu/doi/10.2777/926037>
 - [6] Bach, Janete Saldanha, Klas, Claus-Peter, & Mutschke, Peter. (2023). *KonsortSWD Measure 5.1: Metadata schema extended report* (1.0). Zenodo. <https://doi.org/10.5281/ZENODO.7588902>
 - [7] Klas, Claus-Peter, Zloch, Matthäus, Bach, Janete Saldanha, Baran, Erdal, & Mutschke, Peter. (2022). *KonsortSWD Measure 5.1: PID Service for variables report* (1.0). Zenodo. <https://doi.org/10.5281/zenodo.6397367>
 - [8] Klas, Claus-Peter, & Hopt, Oliver. (2022, December 7). DDI Variable Documentation and data access using R. EDDI - European User Conference on DDI, Paris, France. Zenodo. <https://doi.org/10.5281/zenodo.7408629>
 - [9] Klas, Claus-Peter, Bach, Janete Saldanha, & Mutschke, Peter. (2023). *GESIS Use case Variable publication and citation & Fine granular access to research data*. <https://doi.org/10.5281/ZENODO.7750031>
 - [10] Zhang, Y., Baran, E., Zloch, M., Mühlbauer, A., Klas, C.-P., & Mutschke, P. (2022). *KonsortSWD PID Registrator* (v0.1.0). Zenodo. <https://doi.org/10.5281/zenodo.5914219>
 - [11] Baran, E., Zloch, M., Saldanha Bach Estevao, J., Klas, C.-P., Zhang, Y., & Mutschke, P. (2023). *KonsortSWD Measure 5.1: extended report on the architecture of the PID registration service*. In *KonsortSWD Measure 5.1: extended report on the architecture of the PID registration service* (1.0, pp. 1–18). Zenodo. <https://doi.org/10.5281/zenodo.10401641>
 - [12] PIC is an international consortium that provides a reliable Handle-based PID infrastructure for research data. ePIC currently has nine members and is open to any centre that stores scientific/research data. ePic Homepage: <http://www.pidconsortium.net/>.
 - [13] Handle Homepage: <https://www.handle.net/>

NFDI4Objects and the Base4NFDI Services: An Overview

Authors:

Fabian Fricke^{1*}, and Florian Thiery^{2*}

*Lead presenter

¹ fabian.fricke@dainist.de, Deutsches Archäologisches Institut

² florian.thiery@leiza.de, Leibniz-Zentrum für Archäologie

Abstract:

NFDI4Objects (N4O) represents a broad community dealing with material remains of human history from around 3 million years and involves numerous disciplines from the humanities, cultural studies and natural sciences with an archaeological and historical focus [1]. The objects examined include serially produced objects such as coins and organic remains such as wood, bones, or pollen. The objects and their relations are constantly changing, causing individual biographies. Therefore, the N4O central tasks include (a) comprehending representations of physical objects as research data, (b) relating them to individual contexts, (c) transforming them adequately into the digital space, and (d) curating them according to domain-specific requirements. Consequently, the NFDI4Objects Research Data Lifecycle goes along with the Object Biography. This leads to a paradigm derived from the FAIRification workflow: (i) capturing, (ii) semantic (meta) data modelling and qualifying, (iii) applying (scientific) analysis, (iv) sharing the data by visualisation and FAIRified publication, (v) interlinking and integrating data into knowledge graphs [2].

Within the NFDI infrastructure and to apply the FAIR principles, research (meta)data must be shared, interlinked, and modelled in a common and community-accepted way. To implement this, *N4O Fundamental Services* (N4O FS) and *N4O Community Services* (N4O CS) will be provided. N4O FS are, e.g., DANTE, cocoda [3], BARTOC, the N4O Knowledge Graph [4], the FOSS N4O Management Hub [5], and the N4O Commons Inkubator [6]. N4O CS are FAIRification Tools and Services such as Alligator, Academic Meta Tool, re3dragon, SPARQLing Unicorn QGIS Plugin [7], archaeology.link, and Wikibase / Semantic Media Wiki / Triplestore infrastructures.

N4O contains many terminologies [8], such as ontologies, authority files, and community-driven vocabularies, that must be published and indexed (via DANTE and BARTOC) and semantically aligned (via Cocoda [3]) to contribute to TS4NFDI. Building a Knowledge Graph Infrastructure within NFDI requires the establishment of consortia-driven Knowledge Graphs that make data interchangeable within the consortium itself. This is done in N4O by a *N4O Minimal Metadata-Set* (N4O MMDS) and its derived *N4O Objects Ontology* (N4O OO). The N4O MMDS comprises information that is absolutely necessary to describe object data on its way through the object life cycle (capture, qualify, analyse, share and interlink). Its content will be developed in close cooperation with the (humanities and cultural studies) MoU group in order to establish interoperability within these domains. The N4O minimal metadata set will be provided in a technical description (e.g. JSON). It will be transformed into various 'dialects', e.g. LIDO/XML, for the individual data exchange of heterogeneous systems. The N4O MMDS and the N4O OO as its RDF representation (based on CIDOC-CRM, schema.org, PROV-O, SKOS, BFO) map the object cycle, which is successively enriched with information through the task areas with the tasks of documenting, collecting, analysing and protecting. These structures are used because the data flow in the Knowledge Graph provides for LIDO/XML and RDF.

This paper will showcase a selection of *N4O Fundamental Services* and *N4O Community Services* and explain how NFDI4Objects can contribute to TS4NFDI and KGI4NFDI.

References

- [1] <https://zenodo.org/doi/10.5281/zenodo.10409227>
- [2] <https://doi.org/10.52825/cordi.v1i.326>
- [3] <https://zenodo.org/doi/10.5281/zenodo.10820783>
- [4] <https://nfdi4objects.github.io/n4o-property-graph/>
- [5] <https://zenodo.org/doi/10.5281/zenodo.10756823>
- [6] <https://zenodo.org/doi/10.5281/zenodo.10973050>
- [7] <https://zenodo.org/doi/10.5281/zenodo.10774878>
- [8] <https://zenodo.org/doi/10.5281/zenodo.10906911>

Keywords: NFDI4Objects, Open Source, Knowledge Graphs, Terminologies, FAIR

Persistent Identifier (PID) a journey of making data machine-actionable

Authors:

Preston Rodrigues^{1*}, David Linke² and Thomas Bönisch²

*Lead presenter

¹preston.rodrigues@hlrs.de, High-Performance Computing Center Stuttgart (HLRS), University of Stuttgart, Germany

²Leibniz-Institut für Katalyse e.V. (LIKAT), Rostock, Germany

Abstract:

We are in an age of digital convergence. This has resulted in a growing demand for data and services that can be accessed from any device at any time. However, finding and linking the data reliably has been a continuous challenge. FAIR principles states findability as one of key pillar in the journey of making data machine-actionable.

This journey is not new. A good example is a published book with its International Standard Book Number (ISBN). However, in this age of rapid digitization it is also imperative to uniquely identify digital object. In order to distinctly identify a digital object, it needs to have an universally unique identifier. Some of the well known techniques used in industry to uniquely identify digital object is Digital Object Identifiers (DOI)¹ and Handle System².

The Handle System includes an open set of protocols, an identifier space and way to modify the associated digital data. The protocol enables a distributed computer system to store identifiers of digital resources and resolve those identifiers into the information necessary to locate and access the resources. The associated information can be changed as needed to reflect the current state of the identified resource without changing the identifier, thus allowing the ID of the item to persist over changes to its location or content.

In NFDI4Cat project, we plan to leverage the flexibility of the Handle system to not only provide PID to digital object but enable linking different digital objects that use the PID. We propose a unique solution combining group and permission based access control and domain specific digital object subscription. This combination ensures strict data privacy, a must for both academia and industry. Making sure only those PID that are made explicitly public viewable can be accessed. Furthermore, to realizing our journey towards machine-actionable data we are leveraging on linkML³. LinkML allows to define a standardized schema and express data and its semantics in a machine-readable format. This ensures data consistency, integrity, and interoperability across different systems, tools, and domains.

As a proof-of-concept we have setup a standalone Handle server⁴ on public cloud infrastructure. The handle prefix is register with Corporation for National Research Initiatives (CNRI) hence, any publicly readable values stored in the handle system will be accessible to all. NFDI consortia members can request for a domain specific space (4cat, 4chem, 4earth, 4ds, 4hum). Using REST API they can mint and manage their PIDs. The REST API provides secure access to the domain specific space enabling flexible PID integration into your existing of future solutions. This Poster will present our current progress towards our journey of making data machine-actionable.

References:

- [1] DOI Homepage. URL: <https://www.doi.org/>, last access on 15.05.2024.
- [2] Handle Homepage. URL: <https://www.handle.net/>, last access on 15.05.2024.
- [3] linkML Homepage. URL: <https://linkml.io/>, last access on 15.05.2024.
- [4] NFDI4Cat API Homepage. URL: <https://api.nfdi4cat.org> ,last access on 15.05.2024.

Keywords: PID, LinkML, API, Handle.net, permission based access, subscription

Open source Virtual Desktop Infrastructure solution for remote visualization and training in the NFDI

Authors:

Armin Saur^{1*}, Michael Scherle², Dirk von Suchodoletz²,
Rafael Gieschke², Yi Sun³, Jean-Karim Hériché³

*Lead presenter

¹armin.saur@rz.uni-freiburg.de, Albert Ludwigs University of Freiburg

²Albert Ludwigs University of Freiburg

³European Molecular Biology Laboratory, Heidelberg

Abstract:

The evolving landscape of IT services is witnessing a significant shift from traditional desktop computer software environments to centralized, cloud-like infrastructures. This transition offers a compelling alternative to the labor-intensive administration associated with distributed machines deployment and updating. Accompanying advantages encompass:

- highly flexible provisioning of diverse software environments, hardware components, or networks,
- scalability for large amounts of data,
- streamlined administration and sharing of work environments,
- support for environmentally sustainable practices, aligning with green IT initiatives, and
- enhanced access and security measures allowing the division of software, data visualization, and (sensitive) data.

Termed as Virtual Desktop Infrastructure (VDI), this approach to machine operation is gaining widespread acceptance. Within the realm of VDI, various use cases emerge, including:

- Scalable desktop infrastructures for education and training of (junior) researchers in various fields on their data management workflows, accessible across a broad spectrum of end-user systems.
- Remote visualization capabilities integrated on top of different High-Performance Computing setups or imaging analysis workstations, with efficient WAN remote access to resources. This reduces the need for data replication to achieve their visualization.
- Secure remote access solutions for sensitive data.
- Establishment of custom (permanent) desktops as alternatives to local desktop installations or terminal servers.

To address these use cases, BAND [1], a first generation OpenStack-based VDI was developed during the EOSC-Life project [2] with a focus on bioimage analysis and as bwLehrpool Remote for general desktops [3]. Both are publicly accessible. To construct a further developed generation, five core areas necessitate implementation, leveraging existing open-source components:

1. Development of an access gateway facilitating unified access to all VDI resources in the cloud. End-users without technical background but advanced research requirements are enabled, by this web page, to use VMs customized for their needs.
2. Construction of a service which allows service steward or administrators the creation, upload and sharing of new VM templates, customized for their unique workflow.
3. Implementation of efficient VM framebuffer grabbing mechanisms, running on the host, and hardware-accelerated video encoding for the remote access protocol SPICE. SPICE enables features like bidirectional audio, printer sharing and USB forwarding.

4. Integration of hardware-accelerated rendering, encompassing 3D-graphics and video processing capabilities within VMs, using passedthrough GPU partitions. Hardware with good open-source drivers is favored over proprietary drivers.
5. Establishment of long-term resource scheduling within the OpenStack framework, catering to the substantial demands of large-scale remote desktop sessions.

As this project is continuing, the first milestones for an open source VDI have been completed [4]. Our research indicated that the Bumblebee project is not modular enough, and a new frontend is developed following previous experience and feedback from [5]. Using a modified QEMU/SPICE server, the VM's frame buffer is accessible and encoded into a video stream, using GPU features. With a latency of less than 100 ms, a first large improvement is done. Using the VDI, the end-user's browser, instructed by the access gateway, opens a SPICE-Client, to establish a connection to the VM.

[1] BAND <https://band.embl.de>

[2] EOSC-Life project <https://www.eosc-life.eu>

[3] bwLehrpool Remote <https://bwlehrpool.ruf.uni-freiburg.de>

[4] Open-source VDI <https://github.com/OpenSourceVDIhttps://github.com/OpenSourceVDI>

[5] bwLehrpool <https://www.bwlehrpool.de/en>

Keywords: VDI, OpenStack, video stream, DaaS

Infrastructure for Research Based Learning and the NFDI Basic Service Jupyter: Searching for Opportunities for Cooperation and Areas of Differentiation

Authors:

Jan F. Kraemer^{1*} and Lilian Löwenau²

*Lead presenter

¹jan.kraemer@hu-berlin.de, Humboldt-Universität zu Berlin, Germany

²Humboldt-Universität zu Berlin, Germany

Abstract:

The appeal of using hosted infrastructure is seen not only in research practice but also in many educational settings. Minimizing individual configuration effort allows focusing on the task at hand while also improving the reproducibility of results. In the educational context, centrally provisioned environments enable teaching staff to integrate the use of advanced computational methods in their curriculum, while focusing on the subject domain rather than the implementation details. Reproducible research practices are equally much easier to implement if existing environments contain the required computational tools and interfaces to data infrastructure. In both contexts, this presents an inherent tradeoff between ease-of-use and the immediacy of usability, and the concomitant explosion of the number of variations that would ideally be offered, but cannot realistically be centrally supported and maintained.

In the specific context of Jupyter based computational notebooks, a specific view heavily influenced the default tools to configure and deploy JupyterHub: a concentration on smaller numbers of similar users, i.e. individual research groups and specific classes. While this allows a local deployment under most circumstances, it also leads to a highly distributed set of “administrative users”, with a high opportunity cost regarding coordination and a multitude of “works for me” solutions to (probably) shared problems.

In the context of larger research projects with established structures, including those with regards to funding as is the case with the NFDI Basic Service Jupyter, this can be approached in a structured and organized manner. In education, requirements for the integration into a set of specific and highly idiosyncratic services for course management as well as varying computational infrastructures at universities, provide an additional layer of complications to coordination.

Inspired by our experience while planning the centrally provided jupyterhub for teaching support at Humboldt-Universität zu Berlin, we seek ways to improve this coordination bottleneck. The announcement of the Jupyter Basic Service initiative of the NFDI in April of 2024 may provide one avenue towards this goal, if sufficient overlap between the respective requirements can be found. Integrating NFDI basic services and local educational infrastructure initiatives could allow the inclusion of research based learning into earlier stages of a students educational journey and thus help to prepare future researchers for the emerging best practices in reproducible research and FAIR data practices.

AT UC4B2024 we aim to present our experiences with and requirements for a centrally managed Jupyter Environment, as well as the results of discussions at (potentially, currently in review) two workshops on the topic: one at the annual meeting of DINI (Deutsche Initiative für Netzwerkinformation) the other at the TURN Conference. We hope to initiate a constructive dialog with the NFDI community about shared needs and possible avenues for joint efforts of improvement of the jupyter stack.

Keywords: Higher Education, Research Based Learning, Infrastructure for Education in Computational (Research-)Methods, Jupyter, JupyterHub

NFDI4Cat Web Application for User-Friendly Generation and Browsing of Catalysis Metadata within a Semantically Rich RDF Framework

Authors:

Taras Petrenko^{1*}, Yuliia Dikova¹, Nadiia Huskova¹, Volodymyr Kushnarenko¹, Preston Rodrigues¹, and Thomas Bönisch¹

*taras.petrenko@hlrs.de

¹High Performance Computing Center Stuttgart (HLRS), Stuttgart, Germany

Abstract:

The key to the success of the NFDI initiative lies in the creation of a common data infrastructure, the development of standardized ontologies, efficient data integration techniques, user-friendly generation of RDF data, and the use of advanced tools and technologies, including AI. Generating RDF data based on a common ontology for a semantic metadata framework, particularly in a specialized field like catalysis, involves several significant challenges::

- Representation of complex concepts and relationships for the domain of catalysis
- Lack of standardized ontologies which hinders interoperability of data
- Heterogeneous data sources
- Ensuring consistency, accuracy and completeness of data
- Generation and collection of RDF data is too complicated for the most researchers to be done manually

The latter aspect poses a major barrier to the wide adoption of the semantic methodologies in the research community. In addition, the exploration of RDF data by means of conventional SPARQL queries is also not very user-friendly. To address the problems systematically, we have developed the following approach. The reference Metadata4Cat ontology is based on the middle-level Metadata4Ing ontology, which provides a reasonable framework for the semantic description of research data. Within the Metadata4Cat ontology, the required metadata fields are indicated by appropriate restriction classes. Such constructs are then used by the questionnaire program TRIQ, developed in our group, for reasoning and querying the corresponding metadata via user-friendly dynamic web forms. This leads to a chain-like instantiation of various resources representing a data context which is required to make data FAIR. The generated RDF metadata are immediately available in a browseable HTML form. In addition, the TRIQ program will be coupled with a triplestore, PID handle service, and the terminology services of NFDI. The proposed methodology can be easily extended to other domains of knowledge. We believe that our application can play an important role in the broad adoption of the semantic data technologies within NFDI.

Keywords: NFDI4Cat, TRIQ, generation of RDF data, semantically rich metadata metadata enrichment, Metadata4Cat ontology, Metadata4Ing, catalysis

Program Committee



Reinhard Altenhöner
Staatsbibliothek zu Berlin – PK



Raphael Ritz
MPCDF



Sonja Schimmler
Fraunhofer FOKUS



Lukas Weimer
SUB Göttingen



User Conference 4 Base4NFDI 2024

Funded by DFG as part of NFDI.

Grant Numbers: 521453681, 521460392, 521462155, 521463400, 521466146, 521471126,
521473512, 521474032, 521475185, 521476232