HELMHOLTZ
Metadata Collaboration

HMC Conference

28.-30.04.26, Heidelberg

"Metadata in Action"

Book of Abstracts

Overview

Intro

Welcome to HMC Conference 2026!

This Book of Abstracts brings together the diverse and inspiring contributions that shape the HMC Conference 2026 – *Metadata in Action*. It reflects not only the breadth of current developments in the field, but also the strong engagement and shared commitment of a growing community dedicated to advancing metadata practices in research.

The Helmholtz Metadata Collaboration (HMC) Conference serves as a unique platform for exchange across disciplines, communities, and areas of expertise. Researchers, data stewards, developers, and infrastructure experts come together to discuss how metadata can enable better science – from improving data quality and interoperability to supporting reproducibility, transparency, and innovative, data-driven research approaches.

The high number of submissions to this year's conference, along with the diversity of topics submitted, underlines the growing importance of metadata across disciplines and research infrastructures. The contributions collected in this volume highlight both conceptual advances and practical implementations. They demonstrate how metadata is increasingly embedded in scientific workflows, infrastructures, and tools, and how it plays a crucial role in making research data more accessible, reusable, and ultimately more impactful. At the same time, they reflect the challenges that remain and the collaborative efforts required to address them.

The theme *Metadata in Action* is not only a title, but clearly reflected in the contributions presented at our conference. Many contributions move beyond theoretical considerations and showcase concrete applications, tools, and services that bring metadata to life in a real-world research context. Together, they illustrate how metadata is not only managed, but actively used to support discovery, integration, and innovation.

We would like to sincerely thank all authors for their valuable submissions and their willingness to share their work with the community. The high number and quality of contributions clearly demonstrate the relevance of the topic and the strength of the community that continues to grow and connect across disciplines. The contributions presented in this book form the foundation of a rich and engaging conference programme.

We also extend our sincere thanks to our keynote speakers for their contributions, as well as to all members of the programme committee and everyone involved in the review and organisation process. Their dedication and efforts ensure a high-quality and well-balanced programme.

We hope that this Book of Abstracts will serve as a valuable companion throughout the conference and beyond – providing orientation, inspiring discussions, and fostering new connections. Above all, we hope it encourages continued collaboration and exchange within and across communities working on metadata and FAIR data practices.

On behalf of the HMC Conference 2026 committees, we wish you an inspiring and productive conference.

Sören Lorenz
Speaker of HMC

Volker Hofmann
Chair of Scientific Programme Committee
HMC Conference 2026

Conference Programme

Overview of the HMC Conference Programme 2026

Here you can find our conference programme:
https://events.geomar.de/event/884/timetable/#all

Tuesday, 28.04.2026

Tuesday, 09:00-12:30

Pre-Conference Workshops

Room Marsilius Kolleg - Lecture Hall
09:00 - 10:30

ID W01

Building Connected Data Ecosystems - How to Facilitate FAIR Data Workflows across Tools and Services

Hosts: Emanuel Söding[1], Rory Macneil[2], Tilo Mathes[2]

[1] GEOMAR Helmholtz Centre for Ocean Research Kiel, [2] Research Space

Most research centers maintain dedicated infrastructures to capture, curate, and store research data produced by their personnel. The employed solutions, however, are often run independently of one another and therefore lack connectivity, creating gaps in the data workflows. An integrated data ecosystem, however, would manage information, provide workflows, and support data documentation as data is produced.

Lab and field notebooks are essential tools for documenting structured information during measurement campaigns or field and laboratory work. Modern Electronic Lab Notebooks (ELNs) and data collection tools offer advanced features to support this documentation process and can enrich records with additional metadata—such as instrumentation details, personnel involved, sample registration, and more. They are often positioned in sections of the data workflow, where critical information is generated and possibly merged and thus could operate as data workflow orchestrators. On the other hand, this task could also be assumed by other tools, depending on the architecture of the envisioned data ecosystem. However, in practice, many centers and laboratories face significant barriers: ELNs and other services are not readily available, may require costly licenses, don't integrate sufficiently into existing workflows and infrastructure. They also often lack institutional support or training opportunities. As a result, their use is not yet widespread.

In this workshop, which builds on the results of a workshop taking place in summer 2025. We would like to discuss potential architecture models within research centers, and invite participants to explore the potential of ELNs and other tools within scientific workflows. Together, we'll discuss desirable features, briefly review a few existing solutions, adoption challenges and consider whether centrally provided ELN services across Helmholtz could be a sustainable way forward. The aim is, to form a working group, developing interoperability standards for data ecosystems."

Corresponding Authors: Emanuel Söding, esoeding@geomar.de;
Rory Macneil, rmacneil@researchspace.com; Tilo Mathes, tilo.mathes@researchspace.com

Room Marsilius Kolleg – Lecture Hall
11:00 – 12:30

ID W08

Semantics Hidden in the Dark –Make Datasets Shine (Practical Integration of Terminology Services for FAIR Data)

Authors: Claudia Martens[1], Alexander Wolodkin[2], Anette Ganske[3]

Co-Authors: Claus Weiland[2], Andrea Lammert[1], Angelina Kraft[3]

[1] German Climate Computing Centre, [2] Senckenberg - Leibniz Institution for Biodiversity and Earth System Research, [3] Leibniz Information Centre for Science and Technology and University Library

Semantic technologies and terminology services are a cornerstone for implementing the FAIR principles, as they make the meaning of data explicit, machine-actionable, and reusable beyond their original context. While data may be technically accessible, a lack of shared semantics often limits interoperability and hinders reuse across disciplines, infrastructures, and research communities. Terminology services address this challenge by providing controlled concepts, semantic relationships, and persistent identifiers that enable consistent interpretation and integration of data.

This workshop focuses on the practical adoption of terminology services in research data infrastructures, moving beyond conceptual discussions toward concrete, transferable implementations. It presents key results of the BITS project (Blueprint for the Integration of Terminology Services in Earth System Science) and demonstrates how terminology services on the example of the ESS TS (https://terminology.nfdi4earth.de) can be embedded into research data infrastructures to improve discovery, interoperability, and semantic enrichment of datasets. Designed as an interactive forum, the workshop combines short inputs, live demonstrations, and participatory elements to make the added value of semantics tangible for different stakeholder groups. Participants will engage with real-world implementations of terminology services integrated into repository interfaces (via API usage) and metadata pipelines, supported by interactive elements such as live polling, search challenges, and guided discussion. By bringing together repository managers, researchers, and data stewards, the workshop fosters exchange between technical and conceptual perspectives and supports community-driven learning. Overall, the workshop aims to lower barriers to adopting terminology services, strengthen awareness of their strategic importance for FAIR data, and stimulate discussion on scalable and sustainable implementations across research infrastructures.

Corresponding Hosts: Claudia Martens, martens@dkrz.de; Anette Ganske, anette.ganske@tib.eu

Room Marsilius Kolleg - Seminar Room 1
09:00 - 10:30

ID W07

Making Helmholtz Data Assets Visible via the Helmholtz Knowledge Graph

Host: Volker Hofmann[1]

Co-Hosts: Anand Deshpande[2], Fiona DMello[1], Gabriel Preuß[3], Lucas Kulla[2], Lucas Lamparter[1], Marco Nolden[2], Mustafa Soylu[1], Oonagh Brendike-Mannix[3], Said Fathalla[1], Stefan Sandfeld[1]

[1] Forschungszentrum Jülich, [2] German Cancer Research Center, [3] Helmholtz-Zentrum Berlin

This workshop aims to advance the Helmholtz Knowledge Graph (HKG) as a shared metadata backbone by identifying new data providers and sources, extending and refining the HKG data model, and jointly evaluating practical onboarding processes for data providers across Helmholtz. The Helmholtz Knowledge Graph is a federated metadata infrastructure that makes digital assets—such as datasets, publications, software, and instruments—discoverable, comparable, and queryable across the Helmholtz Association. While the HKG already integrates metadata from multiple infrastructures, its continued value depends on active collaboration with data providers, domain experts, and metadata professionals.

The workshop provides a structured, interactive setting to work on three closely connected themes. First, participants will identify novel data providers and metadata sources, including domain-specific repositories, institutional services, and emerging infrastructures, that could meaningfully extend the coverage of the HKG. This includes discussing when data sources can be considered authoritative and how they may be used to validate, enrich, or contextualize other metadata in the graph. Second, the workshop will explore metadata schemas and domain-specific structures that are currently not, or only partially, represented. Participants will review limitations of the existing HKG data model and discuss extensions that improve expressiveness for search, discovery, and cross-domain analysis. Finally, participants will discuss how a structured onboarding process for data providers can be established, identifying challenges, best practices, and opportunities to better align technical pipelines with real-world metadata creation and maintenance.

Outcomes include a curated list of candidate data providers, shared criteria for authoritative metadata, concrete proposals for extending the HKG data model, and initial milestones for onboarding new data providers. The workshop will be organized into parallel and successive discussion tables, followed by joint synthesis sessions to consolidate results across perspectives.

Corresponding Host: Volker Hofmann, v.hofmann@fz-juelich.de

Room Marsilius Kolleg – Seminar Room 1
11:00 – 12:30

ID W05

From Shared Challenges to Shared Action: Metadata Harmonization in Practice

Hosts: Andreas Pfeil[1], Anis Koubaa[1], Christine Lemster[2], Fabia Martens[1], Hamideh Haghiri[3], Lucas Lamparter[4], Markus Kubin[5], Oonagh Brendike-Mannix[5], Santiago Casas[6]

[1] Karlsruhe Institute of Technology, [2] GEOMAR Helmholtz Centre for Ocean Research Kiel, [3] German Cancer Research Center, [4] Forschungszentrum Jülich, [5] Helmholtz-Zentrum Berlin, [6] Deutsches Zentrum für Luft- und Raumfahrt (DLR)

Metadata harmonisation is a collective action problem. In this workshop our goal is to bring together data stewards, infrastructure providers, and researchers to share practical experiences in improving metadata quality, and co-identify actionable next steps toward harmonized metadata practices.

The workshop builds on our analysis of metadata provided, previous workshops, and one-on-one counselling sessions.

Intended Outcomes:

The workshop will:

- present a summary of insights gathered from community workshops and one-on-one provider counseling,
- provide short provider case reflections illustrating practical harmonization efforts (successes and challenges),
- facilitate an interactive group exchange on lessons learned, remaining obstacles, and community-identified priorities,
- synthesize outcomes into a joint set of next steps for HMC and providers, and shared recommendations.

Expected results:

(1) Shared understanding of practical paths to improve metadata in provider contexts,

(2) A curated list of next steps and recommendations for provider networks and HMC,

(3) Strengthened network of practitioners engaged in metadata harmonization.

Corresponding Host: Oonagh Brendike-Mannix, oonagh.mannix@helmholtz-berlin.de

Room Marsilius Kolleg – Seminar Room 2
09:00 – 10:30

ID W03

Creating RDF-Compliant Metadata Templates with the AIMS Metadata Profile Service

Hosts: Jürgen Windeck[1], Kseniia Dukkart[2], Marc Fuhrmans[1], Max Schröder[3], Moritz Kern[2], Sebastian Schick[3]

[1] Technical University Darmstadt, [2] RWTH Aachen University, [3] University of Rostock

Generating FAIR research data and enabling its reuse is the overall goal of research data management. However, establishing machine-readable knowledge representation - the "I" in FAIR - as the foundation for FAIR data and metadata remains a major challenge for many research communities. We have developed an approach to create subject-specific, RDF-compliant metadata profiles (i.e., SHACL shapes) that enable precise and flexible documentation of research processes and data. Our modelling approach supports inheritance between profiles: communities can create and share modular profiles as building blocks, which others can adopt and extend, so that metadata remains community-specific and interoperable at the same time.

To facilitate the modelling process and make it accessible to users with limited ontology expertise, we have developed a web service that provides a graphical user interface for creating metadata profiles [1]. It allows users to add suitable terms from existing terminologies together with constraints on permitted value nodes (e.g. expected data types, classes, or node shapes) and attribute cardinalities. Based on those profiles, metadata forms can be automatically generated for entering profile-compliant metadata [2] as well as search interfaces to explore profile-based metadata via faceted search [3].

In this workshop, participants will learn how to use the AIMS editor to create and extend metadata profiles and discuss the challenges of creating RDF-compliant metadata for research data. We will also present the new user interface prototype and conduct a hands-on user test. By gathering feedback from metadata experts, data stewards, and domain experts, we aim to improve the current user interface and discuss how RDF-based metadata can be embedded into everyday research workflows.

References:

- [1] NFDI4ING Metadata Profile Service. https://profiles.nfdi4ing.de
- [2] Shacl-form. https://github.com/ULB-Darmstadt/shacl-form
- [3] RDF-Store. https://github.com/ULB-Darmstadt/rdf-store

Corresponding Hosts: Jürgen Windeck, juergen.windeck@tu-darmstadt.de; Kseniia Dukkart, dukkart@itc.rwth-aachen.de; Marc Fuhrmans, marc.fuhrmans@tu-darmstadt.de; Moritz Kern, kern@itc.rwth-aachen.de

Room Marsilius Kolleg - Seminar Room 2
11:00 - 12:30

ID W04

From Chaos to Clarity: Smart Sample Management with LinkAhead & O2A SAMPLES

Hosts: Maren Rebke[1], Florian Spreckelsen[2]

[1] Alfred Wegener Institute for Polar and Marine Research, [2] IndiScale GmbH

LinkAhead is a flexible open source toolbox for research data that adapts easily when workflows or requirements change. It offers a clear web interface, programmatic access and a semantic structure that can be extended for many different research contexts.

In this workshop we will introduce LinkAhead and demonstrate how it supports O2A SAMPLES, a sustainable and interoperable platform for transparent, FAIR compliant and AI ready sample metadata. O2A SAMPLES enables reliable sample registration, storage tracking and Nagoya documentation, and connects smoothly with Helmholtz infrastructures. With well-defined workflows, QR based tracking and fully documented procedures, it provides an efficient and collaborative approach to managing samples from field collection to digital archive. This unified framework strengthens reproducibility, accessibility and discoverability, enabling efficient digitization and collaboration across the entire sample lifecycle.

In this workshop, participants will first get to know the open-source research data management software LinkAhead [1, 2] which is the basis of the O2A SAMPLES platform at AWI. We will introduce LinkAhead's datamodel and webinterface including hands-on examples of how to query for, insert, and edit data entries in LinkAhead. We will then continue with an introduction of the O2A Samples platform with its sample and storage management workflows. Participants will learn how samples are registered, and how to export and update their metadata. An outlook will be given on configuring and adapting the sample management module [3] to the participants' (or their institutions') needs.

References:

- [1] https://doi.org/10.3390/data4020083
- [2] https://gitlab.com/linkahead
- [3] https://gitlab.com/linkahead/linkahead-sample-management

Corresponding Hosts: Maren Rebke, maren.rebke@awi.de;
Florian Spreckelsen, f.spreckelsen@indiscale.com

Room ATV 106
09:00 - 10:30

ID W02

Creating and Inspecting Research Object Crates - The Interactive Way

Hosts: Andreas Pfeil[1], Christopher Raquet[1]

[1] Karlsruhe Institute of Technology

NovaCrate [1] is a web-based interactive editor for creating, editing, and visualizing Research Object Crates [2] (RO-Crates). Built for inspecting, validating, and manipulating RO-Crates, it enables getting a deeper understanding of an RO-Crate's content and structure.

In our workshop, we aim to provide training in NovaCrate and RO-Crate. We also hope to extend our understanding of the requirements of researchers, data stewards, and any other roles that may come in contact with RO-Crates or NovaCrate [1] for further investigation for potential improvement.

During the hands-on session, we will guide and encourage participants to work together in small groups and package some prepared research data as an RO-Crate with the help of NovaCrate. To do so, participants will describe the research data with metadata created through NovaCrate [1]. Here we see a close connection to track topic No. 4, "From Harmonisation to Action(ability)".

In this process, teams are encouraged to take notes on challenges, blockers, and ideas for improvement. At the end of the workshop, we will discuss the experience with the participants, guided by the notes the participants have taken.

The discussion will be centered around these questions:

- In which scenarios are RO-Crates useful?
- How to approach reuse or consumption of RO-Crates?
- How can you incorporate RO-Crates into your research?

We hope to have an interesting discussion not only providing us with crucial input for the development of our services, but also to offer our participants with the opportunity for discourse on the applications of RO-Crates in their research area.

References:

- [1] https://novacrate.datamanager.kit.edu
- [2] https://www.researchobject.org/ro-crate

Corresponding Host: Christopher Raquet, Christopher.raquet@kit.edu

Room ATV 106
11:00 – 12:30

ID W06

Make Your Own FAIR Digital Objects – The Graphical Way

Hosts: Andreas Pfeil[1], Christopher Raquet[1]

Co-Host: Thomas Jejkal[1]

[1] Karlsruhe Institute of Technology

To accelerate the adoption of FAIR Digital Objects (FDOs), their creation and usage needs to be implemented in software. Our work targets the task of creating and maintaining FDO records. We introduce an application to build designs for FDO records in an intuitive and visual way, targeting non-experts and experts in the field alike. From a design, code and FDOs can be generated to automatically create FDO records from given information.

In this workshop, we aim to provide the skill to create FAIR Digital Objects in smaller and larger scales with minimal resources. We encourage the participants to bring JSON-encoded metadata of the objects they would like to publish as FAIR DOs. For those who do not, we will provide examples to work with. We also hope to get some feedback for the further development of the FAIR DO Designer and insights into deeper requirements of the target group. The workshop will have the following shape:

- Introduction to the FAIR DO Designer (10 min)
- Demonstration and guidance through the basic concepts (interactive, 20 min)
- Working session, so participants can build their own FDOs (guided, 45 min)
- Discussion and Feedback (15 min)

References:

- [1] FAIR DO Designer Code Repository: https://github.com/kit-data-manager/fair-do-designer
- [2] FAIR DO Designer Online Demonstrator: https://kit-data-manager.github.io/fair-do-designer

Corresponding Host: Andreas Pfeil, andreas.pfeil@kit.edu

Room H1.00.028
09:00 - 12:30

ID W10

From Ontology to ELN: Create Your Made-to-Measure Semantic Metadata Platform

Host: Fabian Kirchner[1]

Co-Hosts: Anahita Nafissi[1], Mihir Rambhia[1]

[1] Helmholtz-Zentrum Hereon

This workshop will teach you how to use Herbie for setting up a bespoke semantic electronic laboratory notebook or research metadata platform which is customized to your concrete scientific needs.

We will start with an ontology of your scientific domain, pick a typical metadata record you might want to collect, and end with a set of (re)usable web forms for entering such a record in a fully semantically annotated way.

A typical and cumbersome approach would be creating spreadsheets and a set of transformation scripts to facilitate easy data entry for non-technical users. In the workshop you will get to know an alternative approach using Herbie: You will learn to write validation schemas in the standardized SHACL Shapes Constraint Language, upload these alongside your ontology to Herbie, and obtain a platform with easily usable web forms which automatically persist all entered data into a semantically annotated RDF knowledge graph.

After entering a few exemplary records, we will explore how you can query the created RDF knowledge graph using SPARQL to extract the data you need in downstream projects.

This workshop is intended for those who have an application ontology and want to start collecting (small) (meta)data that is properly semantically annotated. There are no restrictions on the domain. Herbie works best for data entered manually in an append-only approach, like it is typically done in laboratory notebooks.

We assume basic understanding of RDF and OWL, in particular you should be able to understand RDF graphs in the turtle format. You should bring your own computer and have Python and Node.js installed to be able to run some development tools.

Corresponding Host: Fabian Kirchner, fabian.kirchner@hereon.de

Room A0.225
09:00 – 12:30

ID W11

Semantic x-Lab: Bridging Laboratory Metadata and Semantic Knowledge Discovery

Hosts: David Pape[1], Felix Mühlbauer[2], Manja Luzi-Helbing[2], Martin Voigt[1], Oliver Knodel[1]

[1] Helmholtz Center Dresden-Rossendorf, [2] GFZ Helmholtz Centre for Geosciences

The Semantic x-Lab project addresses a fundamental challenge in modern research data ecosystems: the fragmentation of laboratory metadata across heterogeneous systems and disciplinary silos. Funded within the Helmholtz Metadata Collaboration (HMC) and co-led by HZDR, GFZ, and GSI, the project aims to interlink ontology-based descriptions of workflows, instruments, resources, and experimental data to make them discoverable, interoperable, and semantically rich. By building a distributed knowledge graph through a user-centered co-design process with laboratory partners and large-scale facility stakeholders, Semantic x-Lab fosters cross-domain insights that were previously inaccessible due to isolated metadata landscapes.

Building on our 2025 Kick-Off Workshop where we introduced the project scope, collaborative use cases, and the foundational vision for FAIR semantic integration of lab information, this workshop will advance hands-on discussions on concrete integration strategies and community engagement practices. Participants will explore how semantic search interfaces, ontology alignment, and co-design methodologies can support FAIR metadata workflows across research domains.

The workshop aligns with key HMC Conference 2026 track topics, by showcasing ontology-based harmonisation efforts, Human-Machine Collaboration in (Meta)data Acquisition through discussions on digital tools and workflows and Domain and Application-specific Ontologies via real use cases from laboratory contexts. Based on this, we will develop and discuss exemplary knowledge graphs in groups during the workshop in order to introduce researchers to the field, but also to show infrastructure providers and central data stewards how knowledge graphs can support their work and the scientists they serve. We will also take these insights into account as our project progresses and incorporate them into further work.

This workshop invites researchers, data stewards, and infrastructure developers to contribute to shaping Semantic x-Lab's next phases and to collectively envision semantic metadata as a cornerstone for future-ready, cross-disciplinary research discovery.

Corresponding Hosts: David Pape, d.pape@hzdr.de; Felix Mühlbauer, felix.muehlbauer@gfz.de; Manja Luzi-Helbing, manja.luzi-helbing@gfz.de; Martin Voigt, m.voigt@hzdr.de; Oliver Knodel, o.knodel@hzdr.de

Lecture Hall, 14:00–15:00

Keynote

Making FAIR Happen: Culture Change Across the Research Ecosystem

Dr. Marta Teparek[1], Dr. Dani Metilli[2]

[1] Programme Leader for FAIR Data at Open Science NL, Dutch Research Council, NOW, [2] TU Delft

In her keynote, Marta will address how research cultures, infrastructures, and policy frameworks can evolve to strengthen metadata quality and enable FAIR data. Drawing on her international work in professionalising data stewardship and coordinating community-driven initiatives, she will discuss how funders, research organisations, and multidisciplinary teams can collaborate to embed FAIR workflows, improve interoperability, and foster more transparent and trustworthy research.

This will be complemented by, Dr Dani Metilli, ontology engineer at the Thematic Digital Competence Centre for Natural and Engineering Sciences, showing examples of concrete research data interoperability support.

Altogether, the talk will offer a forward-looking perspective on how the research ecosystem can harness metadata as a driver for both scientific excellence and cultural change.

Lecture Hall, 15:30-17:00

Talk Session

Track 1

Metadata in Action: Embedding Quality and Context into Research Infrastructures

Chair: Christine Lemster[1]

[1] HMC Office - GEOMAR Helmholtz Centre for Ocean Research Kiel

Metadata is increasingly expected to do more than document research objects—it must actively shape decisions, workflows, automation, and interoperability. This session explores how metadata becomes operational: enriched through quality assessment, structured via semantic models, and embedded into research infrastructures to guide both human and machine action.

The contributions span decision support, repository workflows, physical samples and sensors, laboratory environments, and open-access publishing. Together, they show how automated validation, persistent identifiers, and interoperable frameworks transform metadata from minimal compliance into FAIR- and AI-ready research infrastructure.

ID T01

Data Quality Assessment Results as Decision-Relevant Enriched Metadata for Decision Support Systems

Authors: Carsten Oliver Schmidt[1], Elena Salogni, Stephan Struckmann

[1] University Medicine Greifswald

Trustworthy and impactful decision support systems (DSS) critically depend on data that are fit for their intended use. Here, data quality assessments should be understood not merely as a validation step, but as a systematic metadata enrichment process that produces structured, machine-actionable descriptors of data plausibility, and correctness.

Building on established data quality concepts in medical research, this talk generically conceptualises automated data quality assessment as a crucial pre-assessment layer that is analytically distinct from decision logic. Results from such assessments inform decision logic on appropriate data use.

To implement such assessment in an efficient and transparent manner, we have implemented a layered metadata model comprising: (i) metadata describing core data quality dimensions such as missing data, inadmissible data, contradictions, or unexpected associations; (ii) metadata capturing the formal rules and expectations used to detect data quality issues; (iii) metadata that classifies the severity and decision relevance of the detected issues within a specific application context - this classification is inherently normative and context dependent, it reflects explicit assumptions about how particular data limitations may affect decisions; and from the first three layers finally (iv) the derived fitness-for-decision metadata that explicitly characterises the suitability of a dataset for downstream decision processes.

Using the tools dataquieR (R) and dqrep (Stata) as concrete implementations, based on an example from the medical domain, we demonstrate how this separation of assessment and decision layers enables context-specific, automated generation of interoperable data quality metadata with minimal programming effort. We illustrate how such metadata can shape DSS behaviour – for example by suppressing unreliable alerts, or down-weighting model inputs.

We conclude by discussing implications for metadata standards, interoperability of data quality assessment results. Treating such results as decision-relevant metadata provides a structured mechanism to potentially increase the value of DSS.

Corresponding Author: Carsten Oliver Schmidt, carsten.schmidt@uni-greifswald.de

ID T02

Self-Assessment for FAIR Data Publication: Empowering Researchers to Improve Dataset Quality Before Submission

Author: Marleen Marynissen[1]

Co-Author: Dieuwertje Bloemen[1]

[1] KU Leuven

The CoreTrustSeal certified institutional data repository RDR, built on Dataverse, is central to KU Leuven's efforts to support FAIR data publication. Since its launch in 2022, the growing number of dataset submissions has highlighted the need for an efficient, transparent, and consistent curation workflow. To address this, the RDR team developed an open-source review dashboard that integrates with Dataverse and streamlines the curation process.

Initially designed to optimize the review workflow, the dashboard's second iteration introduced Python-based automated checks for systematic quality assessment. These checks validate metadata completeness and consistency while flagging issues such as missing PIDs, unclear licensing, insufficient metadata, or absent README files. Crucially, automation complements rather than replaces human judgement: curators can override or contextualize outcomes, ensuring nuanced interpretation remains part of the process.

Recurring metadata issues often surface only during curation, causing delays and additional review rounds. Building on the insights from the automated checks in the review dashboard, the RDR team is developing a self-assessment tool for researchers. This tool enables more complex pre-submission validation of draft datasets than is possible in the Dataverse UI and embeds FAIR-oriented guidance, including PID requirements, licensing clarity, consistent metadata, and documentation completeness. By providing concrete, and actionable feedback, it helps prevent common issues before formal review and supports the creation of more complete datasets.

The presentation will introduce the design principles and implementation of this self-assessment tool, highlighting the metadata checks and how feedback is presented to users. We will discuss how automated assessment assists researchers in fulfilling essential requirements while encouraging more complete metadata. Furthermore, we will reflect on key insights and challenges, offering guidance for institutions aiming to strengthen research support and enhance metadata quality for FAIR-aligned data publication.

Corresponding Author: Marleen Marynissen, marleen.marynissen@kuleuven.be

ID T03

Ontology-Driven FAIR Sensor Maintenance Metadata for Environmental Observations

Author: Smruthishree Srinivasa[1]

[1] Helmholtz-Zentrum Hereon

Environmental research relies on a wide variety of sensors deployed across oceanic, terrestrial, and atmospheric platforms, yet maintenance metadata remains poorly standardized and difficult to integrate into FAIR-compliant workflows. Metadata is crucial for interpreting sensor data, particularly on research vessels equipped with diverse instruments operating globally. Contextual details including sensor specifications, timestamp, location, etc. are essential for ensuring interoperability and reproducibility of the data.

The MOIN4Herbie project funded by HMC addresses this challenge by integrating maintenance-specific semantic models into Herbie, an electronic laboratory notebook built on an RDF-based collaborative knowledge graph. To support the recording of maintenance activities, we developed the MOIN and MOIN4BoknisEck ontologies. MOIN ontology captures maintenance tasks such as calibration and cleaning, along with the instruments, platforms, and personnel involved, building on established ontologies including SSN/SOSA and PROV-O. Because maintenance procedures vary substantially across sensor types, manufacturers, environmental conditions, and operational protocols, the MOIN4BoknisEck ontology captures use-case-specific requirements.

Using these ontologies in combination with SHACL, we implemented backend and frontend protocols that generate ontology-driven web forms in Herbie. Users enter maintenance information through user-friendly interface, after which Herbie validates the input, applies semantic annotations, and stores the enriched metadata directly in the knowledge graph.

Core device metadata about sensors and platforms are retrieved from O2A registry and ingested into MOIN4Herbie via Registry2RDF module and it is stored as structured RDF records capturing key details such as involved instruments, and responsible personnel. Correspondingly maintenance metadata recorded in MOIN4Herbie, including calibration and cleaning, these records are queried via SPARQL and automatically exported to the O2A Registry through API, reducing manual metadata handling and improving interoperability. Additionally, integration with the OGC SensorThings API enables linkage between maintenance metadata and sensor data streams, further enhancing accessibility across systems.

Corresponding Author: Smruthishree Srinivasa, smruthishree.srinivasa@hereon.de

ID T04

FAIR AIMS – Bringing Rich Metadata for Physical Samples into the Digital World

Author: Kirsten Elger[1]

Co-Authors: Alexander Brauser[1], Birgit Heim[2], Heike Görzig[3], Katherine Rial[3], Mareike Wieczorek[2], Mojeeb Rahman Sedeqi[3], Rolf Krahl[3], Tatjana Antipanova[1]

[1] GFZ Helmholtz Centre for Geosciences, [2] Alfred Wegener Institute for Polar and Marine Research, [3] Helmholtz-Zentrum Berlin

Persistent identifiers (PID) are critical elements of digital research data infrastructures, enabling the unambiguous identification, location and citation of digital representations of a growing range of entities, such as publications and data. Physical samples form the basis for many research results and data. The International Generic Sample Number (IGSN) provides a globally unique, persistent, and web-resolvable identifier for physical objects, allowing them to be found, cited and reanalysed. IGSNs facilitate direct links between data, publications, originating samples, and records of their creation. This closes one of the final gaps in the provenance of research results.

FAIR AIMS builds upon the successful HMC project FAIR WISH. The main outcome of FAIR WISH was the FAIR SAMPLES Template – a modular template, developed for EaE researchers that allows users to select sample-type-specific metadata properties and create customized, rich sample descriptions that comply with the IGSN schema - regardless of the level of digitisation of sample metadata and the individual researcher's metadata training. The template includes a number of linked data vocabularies and forms the basis for the semi-automated generation of IGSN metadata XMLs and subsequent IGSN registration. IGSN metadata submitted via the FAIR SAMPLES TEMPLATE have already proven to be much richer and more complete than previous submissions.

FAIR AIMS will develop an online version of the FAIR SAMPLES Template with automated workflows for IGSN registration and the integration of linked-data vocabularies as dropdown lists, as well as automated metadata quality checks during uploads. FAIR AIMS is also the first IGSN-related project to actively reach beyond the geosciences. Our partner HZB will contribute to FAIR AIMS by developing IGSN metadata profiles for material science samples and integrate the online template into their sample database SEPIA.

Corresponding Author: Kirsten Elger, kelger@gfz.de

ID T05

From Siloed Experiments to Collective Intelligence: Operationalizing the Post-FAIR Laboratory

Author: Nick Garabedian[1]

[1] datin GmbH

While the theoretical benefits of FAIR data are well-established, the operational reality of integrating these principles into active R&D environments reveals a distinct set of challenges and opportunities. This presentation moves beyond the "why" of digitalization to the "how," based on observations from scaling semantic architectures in tribology and materials engineering.

We identify three critical pillars for the "Lab of the Future." First, the Democratization of Context: enabling PhD students to leverage advanced data structures without requiring extensive training in data science. Second, the Contextualization of Automation: ensuring that data streams from continuous robotic testing are automatically enriched with metadata to prevent high-speed resource waste. Third, the Realization of AI Utility: moving past vague promises to a concrete understanding of AI's role: from automating routine analysis to powering high-level predictive models. By addressing these pillars, we show how isolated data points can be woven into a unified, queryable asset, transforming a collection of individual projects into a robust, self-improving research ecosystem.

Corresponding Author: Nick Garabedian, nick.garabedian@datin.io

ID T06

Open Data, Open Access, Open Source: Liberating Metadata for OA Books and Chapters

Author: Tobias Steiner1

[1] Thoth Open Metadata

As open access (OA) becomes the dominant model for scholarly book publishing, the integration of open, standard-compliant metadata into publishing workflows, library systems, and preservation infrastructures has become increasingly urgent. This presentation reports on key findings from a recent metadata study ([Steiner et. al. 2026] [1]) that reviewed international standards and requirements for OA books and chapters, with a particular focus on the needs of small-to-medium-sized, scholar-led and institutional Diamond OA publishers. The study identifies persistent challenges in discoverability, interoperability, and sustainability, and outlines practical approaches to improving open metadata management across the long-form publishing lifecycle.

Building on this analysis, the talk introduces an extended, format-agnostic metadata framework aligned with established regional and national recommendations for OA publishing (e.g. NISO, NAG, AG Univerlage Quality Criteria, Diamond OA Standard). The framework is also compatible with open data principles such as that of the Barcelona Declaration on Open Research Information, and incorporates the requirements of major metadata aggregators in the scholarly book supply chain. Designed to be both robust and adaptable, it supports wide dissemination while remaining responsive to future policy and infrastructure developments.

The presentation then demonstrates how [Thoth Open Metadata] [2] operationalises this framework through a freely available, open-source metadata management platform. Drawing on examples from independent, library-based, and university presses across Europe, North America, Latin America, and Africa, we show how publishers retain control over fully open (CC0) and FAIR metadata, manage it centrally, and automatically export it in multiple industry-standard formats (including MARC, ONIX, KBART, and Crossref XML) via open APIs. Finally, we illustrate how Thoth enables seamless dissemination to major OA platforms, automated DOI registration, library integration, and transparent open archiving, while maintaining a format-agnostic upstream source for high-quality open metadata for OA books and chapters.

References:

- [1] https://doi.org/10.5281/zenodo.18173982
- [2] https://thoth.pub

Corresponding Author: Tobias Steiner, toby@thoth.pub

Foyer & 1st Floor, 17:00 - 19:00

Poster & Demo Session

Track 4

Human-Machine Collaboration in (Meta)data Acquisition

Chair: Marta Dembska[1]

[1] HMC Project LabFriend - Deutsches Zentrum für Luft- und Raumfahrt (DLR)

Scientific progress increasingly depends on effective collaboration between humans and machines. Central to this is the digital, standardized acquisition and management of (meta)data in laboratories and fieldwork alike.

This session explores how digital tools (ELNs, LIMS), formalized workflows using ontologies and controlled vocabularies, and automation through robotics or AI agents can enhance (meta)data capture. Emphasis will be placed on metadata standardization, intuitive data entry, and user support to ensure interoperability and reuse across research domains.

ID D05

Building Knowledge Graphs with the ELN Herbie

Author: Fabian Kirchner[1]

Co-Authors: Catriona Eschke[1], Martin Held[1], Anahita Nafissi[1], Andrews Mensah[1], Severin Gescher[1], Linda Baldewein[1], Smruthishree Srinivasa[1], Mihir Rambhia[1], Katharina Zwosta[1], Carsten Schirnick[2], Claas Faber[2]

[1] Helmholtz-Zentrum Hereon, [2] GEOMAR Helmholtz Centre for Ocean Research Kiel

There is an increasing effort in scientific communities to create shared vocabularies and ontologies. These build the foundation of a semantically annotated knowledge graph which can surface all research data and enable holistic data analysis across various data sources and research domains.

Making machine-generated data available in such a knowledge graph is typically done by setting up scripts and data transformation pipelines which automatically add semantic annotations. Unfortunately, a good solution for capturing manually recorded (meta)data in such a knowledge graph is still lacking.

Herbie, the semantic electronic lab notebook and research database developed at Hereon, fills this gap. In Herbie, users can enter all (meta)data on their experiments in customized web forms. And once submitted, Herbie automatically adds semantic annotations and stores everything directly in the knowledge graph. So, it is as easy to use as a spreadsheet but produces FAIR data without any additional post-processing work. Herbie is configured using the standardized SHACL Shapes Constraint Language and furthermore builds on well-established frameworks in the RDF ecosystem like RDFS, OWL, or RO-Crate.

We will showcase this approach through a typical example of a production and analysis chain as can be found in many scientific domains.

Corresponding Author: Fabian Kirchner, fabian.kirchner@hereon.de

ID D06

Easy RDM Access in Shepard

Authors: Felix Lettowksy[1], Patrick Kaufmann[1], Roland Glück[1]

[1] Deutsches Zentrum für Luft- und Raumfahrt (DLR)

At the German Aerospace Center in Augsburg, the data management system shepard (Storage for Heterogeneous Production and Research Data) is being developed as a tool for research data management. It organizes data and brings metadata to the data by organizational elements, starting from Collections as top level elements above Data Objects which in turn contain Data References. The latter point to the actual data and measurements, stored in dedicated containers. Shepard supports timeseries data, custom-tailored structured data, raw files and spatial data as data types. The use of semantic annotations allows for further enrichment with metadata.

Technologically, the core of shepard consists of an API, provided by a Quarkus server. A web based user interface allows for quickly viewing and browsing the content. However, in order to collect data automatically and general purpose scripting, the API has to be used. As a consequence, the hurdles for its usage should be as low as possible so that researchers do not have to spend too much time and effort to tap into the full potential of shepard.

In this session, we aim to showcase the low-threshold access to shepard. In particular, we present a one shot configuration which allows the user to build and run a local instance in short time. A script generating an example collection describing the inner solar system gives an introduction how to structure and to work with data in shepard. Furthermore, we show how the API can be used via Python clients also by rather inexperienced programmers. Particular focus is being given towards the integration of Jupyter notebooks as an easy possibility for visualization and first steps toward data evaluation. Also, the frontend will be demonstrated where all kinds of objects can be created, related to each other, and semantically annotated in an easy intuitive way.

Corresponding Author: Roland Glück, roland.glueck@dlr.de

ID D07

Large-Scale Extraction and Annotation of Quantitative Information on Energy Technologies from Scientific Literature

Author: Maxime Gorres[1]

Co-Authors: Jan Göpfert[1], Jann M. Weinand[1], Patrick Kuckertz[1], Titan Hartono[1]

[1] Forschungszentrum Jülich

Systematic literature reviews are fundamental to energy system analysis, yet are often time-consuming, incomplete, and inconsistent. While manually curated datasets provide valuable structured information for specific subdomains of energy research, extending such efforts to the entire field remains challenging. At the same time, easily accessible and extensible quantitative evidence would substantially benefit the research community.

In this contribution, we present a large-scale, automatically compiled dataset of quantitative information extracted from 15 years of energy systems literature using Quinex, an LLM-based information extraction tool. Quinex identifies quantitative statements and transforms them into structured data containing numerical values, units, quantified properties, entities, and contextual metadata such as spatial and temporal scope. The literature corpus was com-piled using advanced searches in Scopus and Web of Science, covering a broad range of keywords. It comprises approximately 76,000 abstracts, of which around 31,000 include full texts.

Applying Quinex to this corpus yielded roughly three million quantitative datapoints. As the tool is domain-agnostic, the extracted information includes values unrelated to energy systems. To enable meaningful analysis, we implemented a filtering and normalization workflow based on regular expressions, resulting in a dataset tailored to energy system research.

A preliminary analysis demonstrates the dataset's potential applications. Photovoltaic and wind technologies constitute the largest share, with cost and efficiency being the most frequently reported properties. The distribution of technologies exhibits strong regional patterns, reflecting differences in research focus across countries. Normalized data and metadata further enable temporal analyses, revealing trends in key techno-economic parameters such as efficiency, lifetime, and capacity factor.

The processed data are made available through an interactive dashboard that allows users to filter, visualize, and download customized subsets. Future work will map extracted metadata to the Open Energy Ontology and integrate the dataset into a collaborative infrastructure to support community-driven data sharing.

Corresponding Author: Maxime Gorres, m.gorres@fz-juelich.de

ID P01

Automating Cryo-EM Metadata Management with ELN

Author: Chaitali Suhas Bagwe[1]

Co-Authors: Carsten Sachse, Daniel Mann, Florian Rhiem, Thomas Heidler

[1] Forschungszentrum Jülich

Modern cryo-electron microscopy (cryo-EM) experiments generate large imaging datasets along with extensive metadata. This metadata describes microscope configurations, acquisition parameters, and experimental conditions. At shared microscopy facilities, metadata is often spread across multiple vendor-specific acquisition tools. It is stored in different file formats and locations. As a result, metadata is frequently incomplete, inconsistently documented, and difficult to retrieve. This limits transparency, traceability, and the effective use of experimental context during data processing and interpretation.

We present an automated workflow for managing cryo-EM metadata at the Ernst-Ruska Centre (Forschungszentrum Jülich). The workflow captures metadata directly from microscope acquisition software and integrates it into SampleDB, an institutional metadata repository. Custom Python-based tools extract relevant metadata at the time of data generation. The metadata is transformed into structured records and linked to the corresponding raw imaging data. This shifts metadata handling from a manual, retrospective task to a systematic process embedded in routine data acquisition. By standardizing metadata capture and integration, the workflow improves completeness, reduces manual entry errors, and strengthens data traceability across the facility.

Corresponding Author: Chaitali Suhas Bagwe, c.bagwe@fz-juelich.de

ID P03

From Isolated Catalogues to Integrated Ecosystems: Integrating UFZ Spatial Data Infrastructure with Global Research Ecosystems

Author: Johann Wurz[1]

Co-Authors: Andrea Pörsch[2], Dorothee Kottmeier[3], Emanuel Söding[4], Stanislav Malinovschii[4], Sören Lorenz [4]

[1] UFZ - Helmholtz Centre for Environmental Research, [2] GFZ Helmholtz Centre for Geosciences, [3] Alfred Wegener Institute for Polar and Marine Research, [4] GEOMAR Helmholtz Centre for Ocean Research Kiel

Building a FAIR (Findable, Accessible, Interoperable, Reusable) data space requires more than individual good practice; it demands coordinated measures across infrastructures, facilities, and research associations. Within the Helmholtz Association, implementing the FAIR principles in practice involves integrating heterogeneous data infrastructures into a cohesive ecosystem. This work aims to contribute to that effort by connecting the distributed Helmholtz data landscape and creating a sustainable, distributed, and semantically enriched (meta)data space.

To support this goal, we implemented interoperable interfaces for the spatial data infrastructure of the Helmholtz Centre for Environmental Research (UFZ), enabling seamless integration with external digital research ecosystems. As a first step, we connected the UFZ GeoNetwork metadata catalogue to the Ocean Data and Information System (ODIS) using schema.org as a common semantic layer. This approach enables metadata exchange following patterns similar to those used by web search engines, improving discoverability and interoperability across domains.

The resulting connection highlights gaps and strengths within the UFZ data infrastructure, providing a replicable pathway for linking additional Helmholtz data facilities to national and international ecosystems. Ultimately, this work supports the creation of a truly FAIR data space across Helmholtz, opening avenues for new scientific applications, such as cross-domain data products and integrated research workflows.

Corresponding Author: Johann Wurz, jwurz@geomar.de

Track 6

Empowering Research Communities: Turning Metadata into Action

Chair: Emanuel Söding[1]

[1] HMC Hub Earth and Environment - GEOMAR Helmholtz Centre for Ocean Research Kiel

Turning metadata recommendations into practice requires engagement across all levels of the research data ecosystem. Overcoming differences in knowledge, communication, and technical skills requires coordinated efforts in training, devising shared guidelines and establishing communities of practice. This session explores formats, initiatives, and practical building blocks that enable communities to improve their (meta)data handling and foster sustainable change in research practices.

ID D08

Come as You Are – Umwelt.info's Way of Sampling Environmental and Nature Protection Data Across Platforms and Standards

Author: Maximilian Berthold[1]

Co-Authors: Anja Reineke[1], Johannes Vogel[1]

[1] Umweltbundesamt

We, the National Centre for Environmental and Nature Conservation Information, develop the portal umwelt.info (https://umwelt.info) which acts as a central access point to all of Germany's knowledge on the environment and nature protection. We integrate all openly accessible sources from municipalities to federal states, civil society, economy and sciences into one flexible catalogue. We follow an agnostic approach and neither require nor enforce a specific technical standard to add data and information into our catalogue. This approach is inclusive, but due to the nature of this flexible catalogue no fixed metadata standard is preferred.

Here, we want to present our approach on how to combine this diverse data ecosystem into one searchable catalogue. We develop an open-source software, where everybody can contribute to the development. Thus, we would like to give insights into our development process, how you can contribute, or give feedback on useful features.

We offer a native API representing our own flexible metadata schema, as well as an emulated CKAN interface. Furthermore, we create editorials and download scripts that provide access to openly available data in Germany using our catalogue. These products aim to help scientists to gain easier access, as well as information on reusability. Our current product can be found at https://umwelt.info and our current development stage at https://gitlab.opencode.de/umwelt-info.

Corresponding Author Maximilian Berthold, maximilian.berthold@uba.de

ID D09

HZB's Tailored Digital Lab Workflows Towards AI-ready Datasets

Author: Carla Terboven[1]

Co-Authors: Daniel Amkreutz[1], Michael Götte[1], Marcel Risch[1], Lucas Thum[1], Ana Karen Velazquez Sanchez[1], Maddalena Zoli[1]

[1] Helmholtz Zentrum Berlin

Efficient data management and structured digital workflows are essential for transforming experimental science toward FAIR datasets. We implement a comprehensive digital lab infrastructure that links experimental data across the full sample lifecycle—from synthesis to advanced characterization—ensuring machine-readable, metadata-rich datasets.

At HZB, Data Stewards and Laboratory Scientists collaborate closely to integrate software, metadata, and experimental data. Data Stewards develop the open-source platform NOMAD Oasis to structure data according to FAIR principles and harmonize metadata using collaboratively created vocabularies (e.g., voc4Cat, TFSCO). Laboratory Scientists generate and document heterogeneous experimental data, which are captured through digital laboratory workflows and subsequently analyzed using jointly developed, customized Jupyter notebooks. Through this close partnership, both groups produce interoperable, reusable datasets that support automated analyses and AI-driven applications.

In this contribution, we present two workflows developed for Thin Film Catalyst laboratories focusing on Thermocatalysis and Electrocatalysis. These workflows systematically capture and structure research data across multiple lab processes, enabling high-throughput analysis, AI-driven insights, and efficient reuse. By highlighting our design decisions for optimizing experimental parameters using Bayesian optimization and for creating linked datasets to efficiently build combinatorial libraries, we aim to share practical strategies for implementing FAIR-aligned, metadata-rich digital lab workflows in heterogeneous experimental environments.

Corresponding Author: Carla Terboven, carla.terboven@helmholtz-berlin.de

ID D10

Key Approaches and Insights from the HMC Survey Task Force

Author: Markus Kubin[1]

Co-Authors: Lucas Kulla[2], Christine Lemster[3], Stanislav Malinovschii[3], Andreas Schmidt[4], Sangeetha Shankar[5], Karl-Uwe Stucky[4]

[1] Helmholtz-Zentrum Berlin, [2] German Cancer Research Center, [3] GEOMAR Helmholtz Centre for Ocean Research Kiel, [4] Karlsruhe Institute of Technology, [5] Deutsches Zentrum für Luft- und Raumfahrt (DLR)

HMC's Survey Task Force has been actively engaged since 2020 in gathering information on practices, gaps, and needs related to (meta)data management within the Helmholtz Association. Through its surveys, the Task Force has identified key strengths and weaknesses in research data management (RDM) practices across the Association. This poster and live demonstration present a review of the Task Force's most recent survey approach and deliverables, along with selected key findings. It will include an interactive visualization of the survey data, allowing conference participants to explore the results from the perspective of their target groups. With this review, we aim to foster discussion and gather feedback on future directions.

Corresponding Author: Markus Kubin, markus.kubin@helmholtz-berlin.de

ID D11

ManGO & FriGO: A Growing and Open-Source Platform for Active Research Data Management and Long-Term Archiving

Authors: Mariana Montes[1], Paul Borgermans[1]

Co-Authors: Danai Kafetzaki[1], Ingrid Barcena Roig[1], Jef Scheepers[1], Joachim Bovin[1], Mustafa Dikmen[1]

[1] KU Leuven

Following the introduction of the metadata driven ManGO RDM platform at HMC 2025, we present the expansion of this open-source RDM (eco) system. Originally developed at the University of Leuven to facilitate our researcher's metadata-driven storage and workflows during the active research phase, ManGO is now starting to see a broader adoption across European institutions and universities. And of course, as needs and insights evolve, also the feature set grows accordingly.

As primarily a system aimed at active research data management, ManGO offers tools and services to discover, use and organize data along metadata while it is produced and/or analyzed. In other words, before (part of it) is to become part of established and curated datasets.

However, not all data during the active research phase is meant to be published in public archives, but nevertheless must be kept in long term archives. This can be imposed either through legal requirements or policies at the institution level with in any case the option to possibly retrieve it sometime in the future. At the University of Leuven the standard RDM policy is to keep relevant research data for at least 10 years after a research project finishes.

To address this, we introduce FriGO, a long-term archiving solution fully integrated into the ManGO architecture. FriGO utilizes a combination of BagIt and RO-Crate specifications to ensure data integrity and standardized metadata packaging. While the bulk data is moved to cost-effective cold storage, the (normally) rich, explicitly defined metadata remains indexed within ManGO's active storage. This hybrid approach ensures that archived datasets remain searchable and interpretable over many years, directly supporting the "Findable" and "Reusable" pillars of FAIR data management.

Corresponding Author: Paul Borgermans, paul.borgermans@kuleuven.be

ID D12

SEPIA: A Metadata-Driven Infrastructure for Persistent and Traceable Sample Management

Authors: Mojeeb Rahman Sedeqi[1], Katherine Rial[1], Rolf Krahl[1], Heike Goerzig[1]

[1] Helmholtz-Zentrum Berlin

Scientific data collected at large-scale research infrastructures is only as reusable and reproducible as the metadata describing the samples under investigation. In practice, sample metadata is often incomplete, fragmented across systems, or insufficiently linked to experiments, datasets, and people. The **SEPIA (Sample Essentials, Persistent Identifiers & Attributes) system** addresses this challenge by providing a metadata-driven infrastructure that enables comprehensive, persistent, and traceable sample descriptions throughout the entire research lifecycle.

SEPIA is built around a robust backend system and Open REST API that uniquely identifies samples and systematically captures their essential attributes, provenance, and relationships. By supporting persistent identifiers such as IGSN, structured metadata aligned with the DataCite Metadata Schema, and tight integration with ICAT, SEPIA enables unambiguous referencing of samples and seamless linkage to investigations, datasets, and users. These capabilities are complemented by a modern web-based frontend that provides user-friendly access to the SEPIA API, allowing researchers, beamline scientists, and administrators to register, update, and explore sample metadata through an intuitive interface. The system supports flexible sample registration via structured JSON payloads, existing published identifiers, or DataCite XML uploads, allowing integration into diverse research workflows.

Currently implemented as a pilot at the Helmholtz-Zentrum Berlin für Materialien und Energie (HZB), SEPIA streamlines sample management across beamlines, enhances collaboration between researchers and facilities, and ensures that sample-related metadata remains accessible, interoperable, and FAIR. A modern web-based frontend complements the backend API, enabling researchers, beamline scientists, and administrators to register, track, search, and explore samples efficiently.

At the HMC Conference 2026, SEPIA will be presented as an interactive poster and live demonstration, showcasing how persistent identifiers and well-defined metadata attributes transform sample metadata into an active component of research infrastructure, enabling better data traceability, interoperability, and reuse.

Corresponding Author: Mojeeb Rahman Sedeqi mojeeb.sedeqi@helmholtz-berlin.de

ID P06

CDIF-4-XAS: OSCARS Project on Interoperability

Authors: Abraham Nieva de la Hidalga[1], Arofan Gregory[2], Heike Goerzig[5], Leandro Liborio[3], Markus Kubin[5], Patrick Austin[3], Rolf Krahl[5], Simon Hodson[2], Vyacheslav Tykhonov[2]

[1] Cardiff University, [2] The Committee on Data of the International Science Council, [3] Rutherford Appleton Laboratory, [4] Helmholtz-Zentrum Berlin

X-ray Absorption Spectroscopy (XAS) research has expanded to become a set of widely used scientific methods with applications across Physics, Chemistry, Surface Science, Nanoscale Science, Biology, and Environmental and Earth Sciences. Over time, various scientific communities, research facilities, device providers, and software developers have created different formats and applications to store XAS data and describe it with metadata. These custom data formats are very often developed ad-hoc they serve their immediate needs and they are integrated with local workflows but they are not easily integrated or interoperable outside their immediate usage. While some standards such as XAFS Data Interchange (XDI) format and NeXus HDF5 format have been developed, add-hoc formats are a widely implemented and practical interoperability across the full range of systems has yet to be realized. This project looks at using the WorldFAIR Cross-Domain Interoperability Framework (CDIF), in combination with existing domain standards and ontologies, to show that real interoperability can be achieved. CDIF utilizes common cross-domain standards such as Schema.org, SKOS, and DDI-CDI to support FAIR exchange of data. The project includes drafting the CDIF profile and mappings from existing standards, and implementation of CDIF tools and workflows inside the EOSC Galaxy infrastructure, including the use of generative AI to enhance metadata capture. This presentation will present the profile and mappings and discuss how this approach may be applied to other use cases. We will also discuss and demo the implementation in Galaxy and the wider applicability of these solutions.

Corresponding Author: Rolf Krahl, rolf.krahl@helmholtz-berlin.de

ID P07

Easy-to-Use RDF-Compliant Metadata Templates for Subject-Specific Metadata with the AIMS Metadata Profile Service

Authors: Ilona Lang[1], Jürgen Windeck[2], Kseniia Dukkart[1], Marc Fuhrmans[3], Marius Politze[1], Max Schröder[4], Sebastian Schick[4]

[1] IT Center RWTH Aachen, [2] Universitäts- und Landesbibliothek Darmstadt, [3] Technical University Darmstadt, [4] University of Rostock

High-quality subject-specific metadata is essential for FAIR research data and for ensuring the comprehensibility and reproducibility of research. Within NFDI4ING and the Applying Interoperable Metadata Standards (AIMS) project, we developed an approach for creating subject-specific RDF-compliant metadata profiles in the form of SHACL shapes. These profiles enable precise, flexible and machine-interpretable documentation of research processes and data.

The approach allows domain experts to benefit from linked data technologies without requiring deep expertise in semantic web standards. It supports semantic encoding and validation of metadata and data through a hierarchical inheritance model combined with modular profiles. This enables the representation of complex research setups while keeping individual profiles reusable across different contexts, thereby increasing interoperability. The approach relies on low- to medium-specificity ontology terms that are widely available across disciplines, achieving sufficient precision through context-specific combinations.

To make profile modelling accessible to users with limited ontology knowledge, we developed an open-source web service with a graphical user interface for creating metadata profiles. The service supports searching existing terminologies, adding terms to profiles, and defining constrains such as data types, classes, node types, and cardinalities. This allows users without SHACL expertise to create profiles that can be used to create, display and browse RDF metadata.

The service also supports community-driven curation. Profiles can be submitted for review and, once approved, are tagged as community-recommended, helping users identify peer-reviewed profiles.

In the follow-up project AIMS 2, the approach is applied to Materials Science and Surface Physics. Metadata profiles will be integrated into electronic laboratory notebooks, using the open-source ELN eLabFTW as an example. This enables semantic metadata capture and RDF export without disrupting existing research workflows. Additionally, profile exploration, visualization, and the creation of interdependent profiles will be further improved to support modular and reusable modelling practicies.

Corresponding Authors: Jürgen Windeck, juergen.windeck@tu-darmstadt.de; Kseniia Dukkart, dukkart@itc.rwth-aachen.de; Marc Fuhrmans, marc.fuhrmans@tu-darmstadt.de

ID P08

FAIR Assessment Tools as Supporters for Interdisciplinary Research?

Author: Anna Lehmann-Gnirck[1]

[1] Fachhochschule Potsdam

Metadata recommendations only unfold their effect when translated into concrete research practice. Within the context of the FAIR principles (Wilkinson et al. 2016), FAIR assessment tools increasingly assume this translational role by interpreting metadata into measurable indicators for FAIRness. They are frequently understood as neutral instruments for assessing data quality (Pellegrino & Tuozzo 2025, p. 2317). However, initial studies from infrastructure and community contexts suggest that FAIR is operationalised differently across existing assessment metrics, leading to inconsistent assessment outcomes (Gehlen et. al 2022, p. 12; Devaraju & Huber 2021 p. 9f). These inconsistencies are mainly expressions of various communities of practice in which FAIR becomes actionable. The poster conceptualises FAIR assessment tools as analytical translation mechanisms situated between discipline-specific data practices and technical information infrastructure (Wilkinson et al. 2016, p. 3). The key concern is on how these tools can be used across disciplinary boundaries to support heterogeneous data aggregation, bridging diverse (meta-)data practices. Based on an understanding of FAIR as a domain-independent information model, this contribution examines how individual assessment tools interpret and operationalize its principles differently.

Methodologically, this study draws on a comparative analysis of selected FAIR assessment tools (including FUJ-I, FAIR Assessment Tool by TKFDM, ACME-Fair) along with systematically developed criteria related to 15 principles and to 41 indicators (RDA FAIR Data Maturity WG 2020, p. 6) to evaluate these tools. To make the tools' implicit requirements explicit, a reference dataset is outlined as a representation of key features to emphasize aspects for analysis. The heuristic guides dataset selection for subsequent comparative tool analysis.

The poster contributes to the discussion by showing how a transparent and reflective use of assessment metrics can help to frame metadata not merely as an unpopular requirement, but as a configurable means for supporting interdisciplinary data aggregation.

Corresponding Author: Anna Lehmann, anna.lehmann@fh-potsdam.de

ID P09

Highlights of the 2nd Interdisciplinary NFDI Metadata Workshop

Authors: Andreas Czerniak[1], Christin Henzen[2], Cord Wiljes[3], Emanuel Söding[4], Heike Fliegl[5], Josh Moore[6], Leyla Jael Castro[7], Oliver Koepler[8], Thorsten Trippel[9], Ulrik Stervbo[10]

[1] Bielefeld University, [2] Dresden University of Technology, [3] German National Research Data Infrastructure, [4] GEOMAR Helmholtz Centre for Ocean Research Kiel [5] FIZ Karlsruhe – Leibniz Institute for Information Infrastructure, [6] German BioImaging – Society for Microscopy and Image Analysis. [7] ZB MED Information Centre for Life Sciences, [8] Leibniz Information Centre for Science and Technology and University Library, [9] Leibniz Institute for the German Language & University of Tübingen, [10] Ruhr-University Bochum

The German National Research Data Infrastructure (NFDI) unifies activities across all scientific disciplines, represented by NFDI consortia, to realise an overarching research data management. Interdisciplinary domain analysis and agreement on common metadata standards for the description of research outputs are key elements for a successful FAIR implementation. The Taskforce Metadata within the NFDI Section (Meta)data, Terminologies, Provenance coordinates the community process of assessing existing cross-disciplinary terminologies and metadata schemas as suitable candidates for NFDI-agreed recommendations. A first workshop was held in Dresden in January 2025 to begin alignment on metadata [1].

Here, we present the results of the second interactive workshop organized by the Taskforce at the TIB Leibniz Information Centre for Science and Technology in June 2025. The goal was to identify and evaluate suitable existing generic metadata schemas and provide recommendations on how to apply these schemas in the individual disciplines. Altogether, 61 participants from 25 NFDI consortia discussed common metadata schemas, disciplinary needs, and standardisation approaches. The recommendation from the first workshop supporting the registration of repositories in re3data was reaffirmed. Thus, the consensus is now being translated into the official NFDI recommendation document. Additionally, breakout sessions facilitated cross-disciplinary discussions on using the metadata schemas identified in the first workshop – DCAT-AP, DataCite and schema.org – and how to connect these to disciplinary schemas. The Taskforce Metadata recommends that NFDI consortia adopt at least one of these three standards, thereby advancing interoperability on an international level, e.g. in the European Open Science Cloud (EOSC) context.

Reference:

- [1] A. Czerniak, L. J. Castro, H. Fliegl, J. Grieb, C. Henzen, O. Koepler, J. Moore, E. Söding, U. Sterbo, T. Trippel, C. Wiljes, Building Bridges through Metadata within the NFDI, https://doi.org/10.5281/zenodo.17195542 and https://doi.org/10.5281/zenodo.16735955

Corresponding Authors: Emanuel Söding, esoeding@geomar.de; Heike Fliegl, heike.fliegl@fiz-karlsruhe.de

ID P10

Improving NetCDF Metadata Quality: A Collaborative Approach for Data Producers

Author: Romy Fösig[1]

Co-Authors: Andrea Pörsch[2], Björn Saß[3], Christof Lorenz[1], Corinna Rebmann[1], Dorothee Kottmeier[4], Emanuel Söding[5], Florian Obersteiner[1], Katharina Loewe[1], Klaus Getzlaff[5], Philipp S. Sommer[3], Sabine Barthlott[1], Tobias Kerzenmacher[1], Ulrich Loup[6]

[1] Karlsruhe Institute of Technology, [2] GFZ Helmholtz Centre for Geosciences,
[3] Helmholtz-Zentrum Hereon, [4] Alfred Wegener Institute for Polar and Marine Research,
[5] GEOMAR Helmholtz Centre for Ocean Research Kiel, [6] Forschungszentrum Jülich

To ensure FAIR data (Wilkinson et al., 2016: https://doi.org/10.1038/sdata.2016.18), well-described datasets with rich metadata are essential for interoperability and reusability. In Earth System Science, NetCDF is the quasi-standard for storing multidimensional data, supported by metadata conventions such as Climate and Forecast (CF, https://cfconventions.org) and Attribute Convention for Data Discovery (ACDD, https://wiki.esipfed.org/Attribute_Convention_for_Data_Discovery_1-3).

While NetCDF can be self-describing, metadata often lacks compatibility and completeness needed by repositories and data portals. The Helmholtz Metadata Guideline for NetCDF (HMG NetCDF) Initiative addresses these issues by establishing a standardized NetCDF workflow. This ensures seamless metadata integration into downstream processes and enhances AI-readiness.

The HMG NetCDF Initiative is a collaborative effort across German research centers, supported by the Helmholtz DataHub. It contributes to broader Helmholtz activities (e.g., HMC) to improve research data management, discoverability, and interoperability.

This presentation will outline the key challenges and solutions and their anticipated impact on the geoscientific community. We will present version 1.0 of the NetCDF metadata attribute guidelines, which have been released, as well as the status of our custom NetCDF checker tool for data producers.

Corresponding Author: Romy Fösig, romy.foesig@kit.edu

ID P11

INSTRUGRAM

Author: Sylvia Reißmann[1]

[1] GEOMAR Helmholtz Centre for Ocean Research Kiel

INSTRUGRAM aims to simplify and standardize the registration of scientific instruments in metadata systems by providing reusable templates based on the PIDINST schema that are linked to controlled vocabularies. The project will develop a modular tool that supports researchers and technical staff to generate high-quality, interoperable metadata with minimal effort, improving metadata quality, scope, and interoperability across research domains. Integration with systems like the O2A Registry and Sensor Management System ensures relevance within Helmholtz infrastructures. The tool will support automated registration via an open API, enable community-driven template development, and contribute to ongoing standardization efforts. All results, including software, templates, and documentation, will be openly available for long-term reuse. This proposal directly supports the HMC mission to foster FAIR data through better provenance metadata.

Corresponding Author: Sylvia Reißmann, sreissmann@geomar.de

ID P12

FID Physik – The New Information Service and Metadata Hub for Physics

Authors: Lina Bockhorn[1], Holger Israel[1], Esther Tobschall[1], Julia Hoffmann[1], Eva Seidlmayer[1], Markus M. Becker[2], Johanna Hickmann[3]

[1] TIB - Leibniz-Informationszentrum Technik und Naturwissenschaften und Universitätsbibliothek, [2] Leibniz-Institut für Plasmaforschung und Technologie e.V., [3] Physikalisch-Technische Bundesanstalt

Starting in February 2026, FID Physik, the discipline-specific information service for physics, is going to provide a free, open science -minded information portfolio tailored to meet the needs of cutting-edge physics research.

FID Physik will offer a suite of services to facilitate discovery and retrieval of relevant subject-specific information, most notably a literature search engine based on quality-curated corpus. Complementary tools include support of analysis and comparison of publications via visualisation powered by knowledge graphs and large language models (LLMs) and a consolidated conference overview.

We have set ourselves the goal to improve the state of metadata for physics research since high-quality metadata form the necessary basis of the advanced, semantic services that we are envisioning.

Crucially, FID Physik is going to set up a Physics Metadata Forum, to raise awareness for metadata in the research community, and to address researchers' metadata needs. The metadata forum ties into a more extensive community engagement that will ensure alignment of our services with needs from the community.

FID Physik is funded by DFG under project number 558452925, via the "Discipline-Specific Information Services" programme, for an initial period of three years.

Corresponding Author: Holger Israel, holger.israel@tib.eu

ID P13

Making PIDs Work in Organizations: A Practical Implementation Plan

Authors: Andrea Pörsch[1], Dorothee Kottmeier[2], Emanuel Söding[3], Johann Wurz[4], Stanislav Malinovschii[3], Sören Lorenz[3]

[1] GFZ Helmholtz Centre for Geosciences, [2] Alfred Wegener Institute for Polar and Marine Research, [3] GEOMAR Helmholtz Centre for Ocean Research Kiel, [3] UFZ - Helmholtz Centre for Environmental Research

At the Helmholtz Association, we aim to establish a well-structured, harmonized data space that integrates information across distributed data infrastructures. Achieving this goal requires standardizing dataset descriptions using appropriate metadata and defining a single source of truth for much of this metadata, from which different systems can draw. Persistent Identifiers (PIDs) in metadata enable the reuse of common information from shared sources. Broad adoption of PID types enhances interoperability and supports machine-actionable data. As a first step, we recommend implementing ROR, ORCID, IGSN, PIDINST, DataCite DOI, and Crossref DOI in our data systems.

However, to practically record and integrate this information into our repositories, we must first identify the specific locations and stakeholders within institutions where this data is generated and maintained. We must also assess which tools and services the Association needs to provide to support seamless data management for its users.

In this presentation, we propose and highlight several tools and services to implement across the organization, based on envisioned workflows, to form a coherent data ecosystem. These include, for example, repository software, electronic lab notebooks (ELNs), terminology services, and other infrastructure components. Implementing these tools will support the various stakeholder groups in fulfilling their roles. It will contribute to a fully integrated, FAIR-compliant data collection and publication ecosystem within the Helmholtz Association.

Corresponding Author: Emanuel Söding, esoeding@geomar.de

ID P14

Navigating the Text+ Data Space: Metadata & the Search Experience

Author: Thomas Eckart[1]

Co-Authors: Leon Fruth[2], Maximilian Hebeis[2], Stefan Buddenbohm[3], Timm Lehmberg[4], Tobias Gradl[2], Uwe Kretschmer[1]

[1] Saxon Academy of Sciences and Humanities in Leipzig, [2] University of Bamberg,
[3] Göttingen State and University Library, [4] Academy of Sciences and Humanities in Hamburg

Research infrastructures serve as vital interfaces between data providers and user groups. Within the German National Research Data Infrastructure (NFDI), the Text+ consortium focuses on text- and language-based research data across three core domains: collections, scholarly editions, and lexical resources.

Since 2021, Text+ has been developing a federated data space that integrates geographically and organizationally distributed data centers to ensure the FAIR (Findable, Accessible, Interoperable, Reusable) and transparent provision of resources.

A central pillar of the **Text+ Data Space** is the consistent application of metadata, including controlled vocabularies and authority files. These elements are essential for enhancing resource discovery and findability. By connecting an extensive resource inventory based on shared vocabulary and referenced authority file entities the researchers are able to navigate a complex digital landscape. The use of authority files and persistent identifiers (PIDs) offers additional advantages: they provide stable references to disambiguated entities and ensure interoperability within a heterogeneous and evolving environment.

However, establishing such a large-scale infrastructure involves significant challenges. These range from the structural heterogeneity of data formats to the semantic diversity of metadata—which must balance global applicability with project-specific requirements. Furthermore, a primary goal is the user-friendly design of services that cater to a wide range of researchers with varying levels of technical expertise.

This poster presents the search experience within the **Text+ Data Space**. Through practical examples, we demonstrate how researchers can effectively navigate the infrastructure and how the integration of metadata and authority files facilitates the discovery of text- and language-based research data.

Corresponding Authors: Stefan Buddenbohm, buddenbohm@sub.uni-goettingen.de; Thomas Eckart, eckart@saw-leipzig.de

ID P15

Supporting Multiple Metadata Schemas and Standards in the Research Data Management Platform BEXIS2

Author: Franziska Zander[1]

Co-Authors: Birgitta König-Ries[1], Cornelia Fürstenau, David Schöne[1], Sven Thiel[1]

[1] Friedrich Schiller University Jena

BEXIS2 (https://bexis2.uni-jena.de) is a flexible research data management platform designed to support heterogeneous scientific domains with diverse requirements for metadata, data handling, and publication workflows. Unlike domain-specific repositories with fixed metadata schemas, BEXIS2 enables both the simultaneous use of multiple metadata schemas and user-defined metadata schemas that can be adapted to disciplinary standards and project-specific needs.

Different research communities require distinct approaches to editing, validating, exporting, and publishing metadata and data. BEXIS2 addresses this diversity by supporting dynamic metadata structures and, more recently, by allowing metadata values to be selected or linked from external terminologies and controlled vocabularies rather than relying on static predefined lists. In combination with a metadata mapping system, this enables interoperability with established standards and infrastructures such as DataCite for DOI minting, GBIF using Darwin Core Archives, and Bioschemas for web-based discovery. In addition, the platform supports versioning of both data and metadata, ensuring traceability and reproducibility.

A core challenge of this approach is the absence of a single or fixed set of metadata schemas. We present how BEXIS2 currently addresses these challenges and outline planned developments, including improved metadata mapping, terminology integration, and export mechanisms. The ongoing development of BEXIS2 is driven by close collaboration between the BEXIS2 developer team, data managers, and curators to ensure that user needs are effectively met.

Corresponding Author: Franziska Zander, franziska.zander@uni-jena.de

ID P16

The Data Hub: An Open-Source Software Framework for Data Collaboration

Authors: Jonathan Ströbele[1], Juliane Boenecke[1]

[1] Bernhard Nocht Institute for Tropical Medicine

Implementing (meta)data standards in the research ecosystem remains challenging, especially when operating across disciplines. Beyond technical barriers, varying semantic interpretations, and heterogeneous skill levels, a critical gap persists: how to transform principles into operational workflows?

This contribution presents the Data Hub (https://datasnack.org), a self-hostable, open-source framework addressing this implementation gap. Built in Python on top of the Django web framework and PostGIS/PostgreSQL database, it provides interdisciplinary teams with an information infrastructure for reproducible data harmonisation and collaborative metadata management while maintaining institutional sovereignty.

Development follows participatory practice with content matter experts and computer scientists, identifying barriers across domains and iteratively refining technical solutions. This process revealed three core design principles: (i) Open-source code is essential for responding to different (meta)data practices; (ii) Code-based data integration ensures transparency, traceability and reproducibility; (iii) Integrated metadata management incorporates quality assessment into data workflows at each transformation step. In addition, the modular architecture supports incremental adoption: teams may begin with basic harmonization and minimal metadata requirements, expanding capabilities as expertise develops.

Currently piloted in global health research projects (but also applicable to different domains), the Data Hub contributes practical infrastructure for translating (meta)data standards into collaborative research practice, with ongoing development informed by community feedback.

Corresponding Authors: Jonathan Ströbele, jonathan.stroebele@bnitm.de; Juliane Boenecke, juliane.boenecke@bnitm.de

ID P17

The MareHub Initiative: Metadata First - A Basis for Joint Marine Data Infrastructures for Earth System Sciences

Author: Angela Schäfer[1]

Co-Authors: The entire MareHub Consortium

[1] Alfred Wegener Institute for Polar and Marine Research

MareHub, as part of DataHub [1], is a joint initiative of the marine research centers in the Helmholtz Association's Earth and Environment research field. Together with the Deutsche Allianz Meeresforschung (DAM) [2] MareHub integrates research data infrastructures to digitally capture and secure heterogeneous research data right from the outset and publish it as joint data products in accordance with the FAIR principles.

At an operational level, experts from the center's research and digital infrastructure teams establish cross-centre working groups to develop data workflows from the field to publication in repositories. They cover topics ranging from the implementation of common metadata standards and terminologies to schemas and SDI interfaces, while providing the HMC [3] with recommendations based on practical feasibility, and vice versa.

(1) From the outset, the possibility of a standardised recording of all necessary metadata, together with recommended PIDs and a handle system, plays a decisive role in ensuring permanent transparency. Therefore, the web-based O2A REGISTRY [4] was developed to manage metadata of measurement platforms, devices, sensors and product-related information. It supports all scientists, device managers, data managers and curators involved in the whole process.

(2) In addition to ongoing coordination efforts, the joint publication of Standard Operating Procedures (SOPs) [5] was identified as a unifying measure to successfully create FAIR data and metadata. Agreed workflows, data and metadata formats, terminologies and schemas for the entire process are published as SOPs in a dedicated DAM community space on Zenodo [6].

(3) Finally, the data products are also processed within the data workflow in accordance with the guidelines of a common SDI [7] enabling them to be visualized and explored in thematic and featured viewers within the jointly developed Marine Data Portal [8] as well as the Earth Data Portal [9].

References:

- [1] https://earth-data.de/about
- [2] https://www.allianz-meeresforschung.de
- [3] https://helmholtz-metadaten.de
- [4] https://registry.o2a-data.de
- [5] https://zenodo.org/communities/dam/records?q=&f=subject%3ASOP&f=subject%3Adata%20processing&l=list&p=1&s=10&sort=newest
- [6] https://zenodo.org/communities/dam
- [7] https://hdl.handle.net/10013/927187e9-9866-4811-bc25-e7ea641d333e
- [8] https://marine-data.de
- [9] https://earth-data.de

Corresponding Author: Angela Schäfer, angela.schaefer@awi.de

ID P18

Treasure Map of Semantic Artefacts in Computationally Intensive Fundamental Physics

Author: Victoria Tokareva[1]

[1] Karlsruhe Institute of Technology

The PUNCH4NFDI (Particles, Universe, NuClei and Hadrons for German National Research Data Infrastructure) consortium brings together researchers and data-infrastructure specialists from fundamental physics communities such as astro-, astroparticle, particle, and nuclear physics. As pioneers of computationally intensive, data-driven research, these communities have decades of experience in developing research data management strategies and knowledge infrastructures. Consequently, a wide variety of semantic artefacts (also referred to in the literature as knowledge artefacts), defined as machine-actionable and machine-readable formalisations of domain knowledge enabling discovery, interpretation, and reuse of research data by humans and machines, has emerged. Encompassing thesauri, vocabularies, ontologies, metadata schemas and standards, these discipline-specific artefacts have been developed to support diverse levels of abstraction, research data lifecycles and scientific use cases. Examples include the international metadata standards developed by the LQCD (Lattice Quantum Chromodynamics) and IVOA (International Virtual Observatory Alliance) communities, metadata schemas used by platforms such as the CERN Open Data portal and the KASCADE Cosmic-Ray Data Centre (KCDC), IVOA keywords and many more. While these artefacts are well suited to their respective communities and many of them are widely adopted internationally, they are based on heterogeneous data and metadata models, representation formalisms, and encoding strategies. Alongside domain-specific artefacts, a wide range of general-purpose, cross-domain standards - such as DCMI Metadata Terms, DataCite, as well as shared controlled vocabularies and ontologies - are employed within the underlying knowledge infrastructures of fundamental physics, acting as integrative elements across disciplines.

This contribution aims to provide a systematic cross-disciplinary overview of knowledge artefacts in computationally intensive fundamental physics and to facilitate informed discussion on practical approaches to harmonisation and interoperability, promoting FAIR and machine-actionable reuse of complex research data.

Corresponding Author: Victoria Tokareva, victoria.tokareva@kit.edu

Track 7

Enriched Metadata for Decision Support Systems

Chair: Hamideh Haghiri[1]

[1] HMC Hub Health - German Cancer Research Center

This track focuses on how metadata enrichment fuels advanced decision support systems by enhancing data quality, context, and interoperability. Submissions are welcome from both conceptual and applied perspectives. We invite contributions that propose new concepts, frameworks, or models for enriching metadata to improve the foundations of decision support systems, as well as implementations or case studies that demonstrate how enriched metadata can be effectively used within decision support tools and infrastructures.

ID P19

Enriching Weather Index Metadata for Agricultural Decision Support: Standardized Quality, Uncertainty, and Fitness-for-Purpose Assessment

Author: Markus Möller[1]

Co-Authors: Mahdi Hedayat Mahmoudi[1], Paul Peschel[1]

[1] Julius Kühn Institute

Agricultural decision support systems - from crop insurance to yield forecasting—depend critically on trustworthy geodata metadata. However, quality information is typically scattered across documentation, leaving uncertainty and fitness-for-purpose reasoning opaque to users.

We present a best practice example for enriching weather index metadata that operationalizes quality, uncertainty, and decision context as machine-actionable FAIR Digital Objects. Using Germany-wide phenology and precipitation data (1 km resolution, 1993-2022), we integrate three metadata components:

(1) Standardized quality metrics: ISO 19157-1-compliant accuracy measures (R^2, $RMSE$) for each crop, phenological phase, and year.

(2) Spatial uncertainty quantification: Uncertainty layers quantifying local interpolation error for all phenological predictions, enabling site-specific accuracy assessment.

(3) Fitness-for-purpose matrices: Structured metadata capturing validated use contexts, limitations, and application-specific quality requirements - extracted and formalized using LLM-assisted workflows.

Rather than treating metadata as static compliance artifacts, enriched metadata travels with data products as ARC containers, enabling users to query "where are data sufficiently accurate and validated for this specific decision context?"

By embedding ISO-compliant quality metrics, uncertainty products, and formalized fitness-for-purpose knowledge into operational workflows, enriched metadata transforms research data into trusted decision support infrastructure. This framework provides a transferable template for domains where metadata quality directly influences decision reliability and stakeholder trust.

Corresponding Author: Markus Möller, markus.moeller@julius-kuehn.de

ID P20

Software CaRD - Decentralized (Software) Metadata Curation

Authors: David Pape[1], Oliver Bertuch[2]

[1] Helmholtz Center Dresden-Rossendorf, [2] Forschungszentrum Jülich

As baseline for a satisfaction of the FAIR4RS principles, research software must be published with metadata in publication repositories that assign persistent identifiers and make the metadata accessible. Additionally, published software metadata must be correct, and rich enough to further improve findability, accessibility, interoperability and reusability. Metadata curation for software publication not only safeguards the respective metadata quality, but also assesses compliance with relevant policies in the Helmholtz Association, its centers, and beyond.

Furthermore, software metadata can cumulatively be enriched with dynamic metadata (e.g., usage, citations, development) and can thus be used for evaluation and academic reporting, e.g., to contribute to software-related indicators currently developed within the Helmholtz Association. While software publication can be automated, metadata curation, publication approval and evaluation processes usually require human involvement and should be supported by user interfaces that build on automation tools.

In our project, we create "Software CaRD" (Software Curation and Reporting Dashboard), an application that presents software publication metadata for curation. Preprocessed metadata from automated pipelines are made accessible in a structured graphical view. Issues and conflicts are highlighted to allow for easy resolution. Software CaRD also assesses metadata for compliance with configurable policies. For evaluation and reporting, relevant metadata from applicable sources is tracked and visualized.

On this poster, we explain the technical solutions employed by the decentralized curation workflow.

Corresponding Author: David Pape, d.pape@hzdr.de

Track 9

From Minimum Requirements to FAIR and AI-Ready: Assessing Metadata Quality

Chair: Volker Hofmann[1]

[1] HMC Hub Information & HMC FAIR Data Commons - Forschungszentrum Jülich

Assessing metadata quality can follow many paths — from institutional policies and FAIR principles to domain-specific needs like AI-readiness. Yet, current formalised assessment approaches remain limited in producing objective, actionable results.

This session invites researchers and infrastructure developers to share and discuss methods for evaluating metadata quality, with a focus on strengthening assessment frameworks that can effectively guide implementation towards interoperability, and data reuse.

ID D19

PID4NFDI: Practical Approaches to PID Metadata Quality and Interoperability

Authors: Sara El-Gebali[1], Stephanie Hagemann-Wilholt[2], Robin Kraus[2], Frederik Springer[2]

[1] DataCite & PID4NFDI, [2] Leibniz Information Centre for Science and Technology and University Library

This poster introduces the PID4NFDI project and the work of the PID Coordination Hub, with a focus on PID metadata and practical approaches to assessing its quality across research infrastructures. Using DataCite DOI metadata as a reference, it highlights how structured metadata can be evaluated to support discovery, reuse, trust, and informed decision-making.

Moving beyond static views of metadata, the poster demonstrates how common institutional questions, such as metadata completeness, identifier adoption (e.g., ORCID, ROR), or funding traceability, can be translated into systematic checks and indicators of metadata quality. It introduces the concept of a "metadata health check" as a lightweight, actionable approach to identifying gaps, inconsistencies, and opportunities for improvement across repositories and tools.

Attendees will be invited to engage with a hands-on metadata health check, offering a practical entry point into assessing and improving their own metadata. The poster also promotes an upcoming hands-on workshop focused on using the DataCite API to query and analyse DOI metadata at scale, enabling participants to turn metadata into evidence for curation, reporting, and strategic decision-making.

Corresponding Host: Sara El-Gebali, sara.elgebali@datacite.org

Track 11

Semantics in Practice: Domain & Application Ontologies

Chair: Said Fathalla[1]

[1] HMC Hub Information - Forschungszentrum Jülich

This track invites contributions on the design, development, and practical application of ontologies at the domain and application levels that support the design and implementation of robust, interoperable data infrastructure. Submissions may showcase novel approaches, reuse of existing ontologies, semantic alignment strategies, and solutions that enable automated discovery, integration, and meaningful use of data across disciplines.

ID D18

Demonstration of Continuous RDF and OWL Vocabulary Quality Checks with LintedData

Author: Jan Martin Keil[1]

Co-Author: Niklas Berndt[1]

[1] Deutsches Zentrum für Luft- und Raumfahrt (DLR)

RDF and OWL vocabularies are an important prerequisite for the FAIR representation of metadata. These vocabularies itself must adhere to certain quality standards to be useful. Over the last years, the Semantic Web community has come up with many recommendations as well as anti-patterns for the development of vocabularies. But, checks for compliance to these best practices can still only be automated partially and require considerable effort. However, learning from software development, quality checks must be performed regularly and completely automated to fully come to effect.

To address this gap, we develop LintedData, a command line tool for automated quality checks of RDF and OWL-based vocabularies. Currently, LintedData is able to perform more than 60 different quality checks. It covers a large part of the best practices that have been broadly accepted in the Semantic Web community as listed in the Ontology Pitfall Catalog or the OBO Foundry Principles. Due to the command line interface and an available Docker image, LintedData can easily be used in context of Continuous Integration (CI) pipelines to be automatically executed each time changes are pushed to an ontology development repository. Using JUnit XML or Markdown files as output formats enables the direct result presentation in the interface of platforms like GitLab or GitHub. An optionally provided configuration file allows to parameterize individual checks and define which checks to execute for a particular pipeline.

During the demonstration we showcase the automated use of LintedData in a CI pipeline of an ontology development repository. Visitors will be able to trigger changes to an ontology, introducing either improvements or new issues, and to experience the automated timely response of LintedData on these changes, pointing to contained problems.

LintedData is publicly available under a permissive license on: https://gitlab.com/dlr-dw/linteddata.

Corresponding Author: Jan Martin Keil, jan_martin.keil@dlr.de

ID P41

Data Integration and Semantic Alignment Across Disciplines with NFDIcore

Authors: Tabea Tietz[1], Shiva Nahani[1], Alsayed Algergawy[1], Dilek Yargan[1], Jörg Waitelonis[1], Harald Sack[1]

[1] FIZ Karlsruhe – Leibniz Institute for Information Infrastructure

Core ontologies play a fundamental role in data interoperability, particularly when data and metadata originate from heterogeneous sources across disciplines, such as in the National Research Data Infrastructure (NFDI). They provide a shared, minimal, and stable semantic backbone that enables disciplines to communicate, understand, and integrate data consistently.

To support interoperability between NFDI domains, NFDIcore has been developed as a mid-level ontology and represents metadata about NFDI resources, including research data, agents, projects, services, and guidelines. The ontology provides a structured, unified framework to streamline the management, organisation, and integration of research data across scientific fields, facilitates sharing and promotes reuse, fostering more efficient collaboration and sustainable data practices within the broader research community. With its modular architecture (cf. Fig. 1), NFDIcore also provides the basis for various application and domain ontologies, allowing for a focused approach to domain-specific research questions through flexible extensions:

NFDI4Culture Ontology (CTO): Research data of the culture community within a research data index, i.e. a single point of access to decentralised cultural heritage research resources.

NFDI4Memory Ontology (MEMO): Concepts from the historical sciences, including the harmonization of metadata and the detailed representation of provenance.

NFDI4DataScience Ontology (DSAI): The AI research lifecycle, bridging the gap between raw datasets, code repositories, machine learning models, and scientific publications.

NFDI-MatWerk Ontology (MWO): Supports the annotation and integration of data related to the materials science lifecycle.

NFDIcore is also expanding through mappings and alignments, including a work-in-progress mapping to DCAT-AP+. The Data Catalog Vocabulary (DCAT) has been extended to DCAT Application Profile (DCAT-AP) for data portals, and DCAT-AP+ adds a provenance layer. This mapping will enable applications of DCAT-AP+ for NFDIcore-compliant ontologies, broaden its use within NFDI projects, and support interdisciplinary data reuse and knowledge transfer with provenance details.

Corresponding Author: Shiva Nahani, shiva.nahani@fiz-karlsruhe.de

ID P42

Multifunctional Data in a Multifunctional Landscape – Development of an Ontology-Based Metadata Description of Near-Surface Geophysical Data

Author: Johannes Rabinger-Völlmer[1]

Co-Authors: Claudia Schütze[2], Till Sonnemann[3], Ulrike Werban[1]

[1] UFZ - Helmholtz Centre for Environmental Research, [2] German Environment Agency,
[3] University of Bonn

A variety of near-surface geophysical methods are used to characterise soil properties at agricultural sites. Such data also provide information on archaeological features and soil dynamics. Metadata standards help making data findable, accessible, interoperable, and reusable according to the FAIR principles.

The aim is therefore to develop a metadata standard for near-surface geophysical prospection. This model, developed with the software Protégé, contains method-independent as well as sensor-specific metadata. The terms were classified hierarchically and restrictions were introduced to ensure all the fundamentals being retrieved, and leaving the possibility to add additional information later. Furthermore, data and object properties were defined in order to specify data formats and also to use defined terms from predefined lists. Existing standards were queried in order to achieve uniform use of selected parameters and thus improve the comparability of the data sets described. Based on the newly developed ontology, and the metadata description derived from it, its outcome will later be made available as a tool for data acquisition to enable interchangeability.

The here presented metadata model represents an initial development stage from FAIR-agro use case 14, which is based on a joint initiative within the NFDI (FAIRagro, NFDI4Objects, NFDI4Earth). It comprises method-independent metadata as well as specific models for electromagnetic induction (EMI) and magnetic prospection. In the future, the model will be expanded and adapted according to the needs of the expert community. In addition, the model will be tested on real measurements.

Corresponding Author: Johannes Rabinger-Völlmer, johannes.rabiger-voellmer@ufz.de

ID P43

Ontology-Based Modelling of Laboratory Processes

Authors: Marta Dembska[1], Martin Held[2], Sirko Schindler[1]

[1] Deutsches Zentrum für Luft- und Raumfahrt (DLR), [2] Helmholtz-Zentrum Hereon

Laboratory processes are fundamental across many scientific and engineering domains, where experimental workflows underpin the generation, interpretation, and reuse of data. While digitalisation of modern laboratories has largely focused on data acquisition and storage, a central challenge remains the formal and shared modelling of laboratory processes and the systematic capture of execution records.

This work presents an ontology-based modelling approach that extends the W3C PROV Ontology (PROV-O) to represent laboratory process plans, executions, and associated provenance across different experimental domains. Building on PROV-O, the ontology provides a formally grounded and interoperable vocabulary for describing both intended workflows and realised executions, while introducing laboratory-specific constructs not covered by generic provenance models. Representations created with this approach are machine-interpretable and reusable across disciplines.

Captured process provenance supports key objectives in experimental sciences, including reproducibility, assessment of error impact, identification of outliers, and the generation of predictions and recommendations. Challenges in applying provenance models to existing and future laboratory workflows - scalability, understandability, and interoperability - are addressed through a deliberately simplified ontology design that captures essential process information while limiting complexity. Storing provenance in knowledge graphs instantiated from the ontology further enhances scalability.

The ontology-based modelling approach can be specialised for domain-specific paradigms, such as the processing-structure-properties-performance framework in materials science, while retaining a common PROV-O-aligned core for cross-domain laboratory process modelling. By providing a formal, reusable structure for process plans, executions, and provenance, this approach facilitates integration of laboratory data across domains and contributes to the FAIRification of experimental data.e of process knowledge and improved FAIRness of laboratory data.

Corresponding Author: Marta Dembska, marta.dembska@dlr.de

ID P44

Ontology-Driven Integration of Preclinical Radiation Oncology Data

Author: Olga Giraldo[1]

Co-Authors: Ina Kurth, Mareike Roscher, Michael Baumann, Rosemarie Lange, Thomas Früchtel, Wahyu Wijaya Hadiwikarta

[1] German Cancer Research Center

Preclinical radiation oncology research produces many different types of data to answer different contexts such as how experiments are designed, how treatments are applied, and how tumors respond to the treatment. This inherent heterogeneity causes challenges in data integration, to ask clear questions across studies, and reusability for new research.

In our research, we approach this problem by demonstrating how this heterogeneuous preclinical radiation oncology data can be organized and consistently connected using an ontology-driven knowledge graph. The demonstration is based on the PTRO ontology [1] as the semantic backbone for data access. Real preclinical datasets are mapped to the ontology and exposed through a SPARQL endpoint.

Overall, this approach shows a real, working example of how complex biomedical data can be better organized, laying the groundwork for further data integration across research domains even beyond cancer or health studies in general.

Reference:

- [1] https://github.com/DKFZ-E220/PTRO

Corresponding Author: Olga Giraldo, olga.giraldo@dkfz-heidelberg.de

ID P45

Semantic Interoperability in NFDI Through Machine-Actionable Ontology Mappings and shared Ontology Development Best Practices

Authors: Ulrik Stervbo[1], Anja Gerber[2], Hendrik Borgelt[3], Kristina Fischer[4], Charles Tapley Hoyt[5], Norbert Kockmann[3], Marta Koscielniak[6], Matthias Löbe[7], Josh Moore[8], Alexander Wolodkin[9], Philip Strömert[10]

[1] NFDI4Immuno, Ruhr University Bochum, [2] NFDI4Objects, Klassik Stiftung Weimar, [3] NFDI4Cat, Technical University of Dortmund, [4] NFDI4Objects, LEIZA-Leibniz-Zentrum für Archäologie, [5] NFDI4Chem, Institute of Inorganic Chemistry, RWTH Aachen University, [6] NFDI4Memory, Bayerische Staatsbibliothek, [7] NFDI4Health, Institut für Medizinische Informatik, Statistik und Epidemiologie (IMISE), Universität Leipzig, [8] NFDI4BIOIMAGE, German BioImaging - Gesellschaft für Mikroskopie und Bildanalyse, [9] NFDI4Earth, Senckenberg -Leibniz Institution for Biodiversity and Earth System Research, [10] NFDI4Chem, Technische Informationsbibliothek

The Nationale Forschungsdateninfrastruktur (NFDI) encompasses 26 consortia across life sciences, humanities, natural sciences, and engineering. Research data infrastructures, including the NFDI, face the challenge of a heterogeneous landscape of terminologies and semantic descriptions. These differences stem from diverse modeling paradigms, ranging from lightweight SKOS vocabularies to expressive OWL ontologies, rather than just disciplinary boundaries. While resources often overlap conceptually, differing logical commitments and formalization levels complicate alignment and reuse.

Despite the shared goal of FAIR data, integration remains a manual and high-effort task. Interdisciplinary research, such as Archaeology, which synthesizes data from molecular biology, building engineering, and material science - requires the reconciliation of diverse data formats. Machine-actionable mappings are therefore essential for practical, scalable interoperability.

To address this, the NFDI Section (Meta)Data Working Group on Ontology Harmonization and Mapping was established in 2023. Its scope includes promoting convergence on shared ontologies, terminologies, or vocabularies whenever possible, adopting technical good practices for their development, documentation, and long-term maintenance, and collaborating with the NFDI Task Force Metadata to standardize metadata. When harmonization is not feasible due to domain-specific frameworks, we focus on the reuse and definition of formal, machine-readable, and commonly shared mappings between such semantic resources.

The research community lacks a mature, reusable solution for the standardized publication and discovery of such mappings. Our working group investigates the technical requirements for a framework to ensure mappings are as FAIR as the data they describe. We advocate for SSSOM as an emerging standard for 1:1 mappings while coordinating with international bodies like the RDA FAIR Mappings WG. By making mappings and guidelines discoverable via shared registries, we enable federated queries across disciplinary knowledge graphs.

We present an overview of the activities of the working group, emphasizing tooling evaluations, international coordination, and the complementary roles of harmonization and mapping in supporting NFDI-wide adoption.

Corresponding Author: Ulrik Stervbo, ulrik.stervbo@elisabethgruppe.de

Wednesday, 29.04.2026

Lecture Hall, 09:00–10:30

Talk Session

Track 2

Software Interoperability for (Meta)data Acquisition

Chair: Martin Held[1]

[1] HMC Projects MetaCook, ELN-DIY-Meta and LabFriend - Helmholtz-Zentrum Hereon

Modern research labs rely on diverse (meta)data acquisition systems — from instrument software to ELNs, LIMs, and workflow managers. To capture the scientific process coherently, these systems must interoperate seamlessly.

This session focuses on practical concepts and existing solutions for bridging across these platforms — from APIs to file-based integration — including the use of semantic or machine-learning tools to enhance interoperability.

ID T07

FDO-Ops Prototype for machine-actionable FDOs

Author: Nicolas Blumenröhr[1]

[1] Karlsruhe Institute of Technology

This talk presents the [FDO-Ops] [1] model as a prototype framework that makes FAIR Digital Objects (FDOs) machine-actionable by discovering, assessing, and executing operations across heterogeneous data resources in an interoperable way.

The prototype builds on a [DOIP/HTTP] [2] client interface where every management function and every Operation FDO is invoked with a uniform request pattern, providing a stable, client-agnostic interaction layer independent of underlying systems. It integrates an identifier and FDO type system supporting base infrastructure, i.e., the [Handle Registry] [3], a [Data Type Registry] [4], and a [Typed PID Maker (TPM)] [5] instance. An extended service component called TPM Adapter is used that (1) supports Elasticsearch-based full-text search over information records; (2) ingests associated FDO-Operation and FDO-FDO relationships into a Neo4j graph for efficient traversal and rule-enforced consistency; (3) uses a mapping component that translates technology-dependent execution protocols of operations into a JSON-based execution map, run by an Executor module (e.g., for Web API calls or script executions).

Conceptually, FDO-Ops advances several interoperability layers defined in different interoperability models, in particular the technical and syntactic layers, by:

- treating operations themselves as reusable Operation FDOs
- separating workflows into phases (discovery, typed metadata assessment, and bit-sequence processing)
- integrating existing standards (e.g., APIs, SKOS/RDF) without requiring changes to established (meta)data systems.

Applicability is exemplified with cross-domain use cases: discovering relevant FDOs, listing associated operations, interpreting SKOS vocabularies via a SPARQL-based endpoint, and executing data-level preprocessing/validation for Numpy and SKOS RDF/XML files.

Overall, this work shows how latest advances in research on machine-actionable FDOs, turning them from passive containers into reusable computational instruments, pave the way towards truly interoperable data ecosystems.

References:

- [1] https://publikationen.bibliothek.kit.edu/1000185135
- [2] https://www.cordra.org/documentation/api/doip-api-for-http-clients.html
- [3] https://www.handle.net/proxy_servlet.html
- [4] https://dtr-pit.pidconsortium.net/#urls/intro.html
- [5] https://github.com/kit-data-manager/pit-service

Corresponding Author: Nicolas Blumenröhr, nicolas.blumenroehr@kit.edu

ID T08

O2A SAMPLES: From Expedition to Digital Database – A Unified Framework for Sample Digitization and Interoperability

Author Maren Rebke[1]

Co-Authors: Stefan Pinkernell[1], Roland Koppe[1], Sonja Hänzelmann[1]

[1] Alfred Wegener Institute for Polar and Marine Research

Comprehensive and standardized sample management is crucial for advancing field research under challenging conditions. We present the "O2A SAMPLES" prototype: a generic Sample Management System designed as a hosted service that unifies diverse sample metadata in one interoperable framework (samples.o2a-data.de). Developed collaboratively by AWI and GEOMAR, O2A SAMPLES not only facilitates automated, digitized workflows for capturing metadata (such as unique sample identification via DataCite, and Nagoya compliance) but also offers seamless integration with well-established Helmholtz systems The main contribution is 1.) a sample management for the Helmholtz community and beyond—enabling institutions, universities, and research organizations to view and loan samples, identify them by QR-code, conduct experiments on actual materials, and benefit from standardized workflows without managing underlying infrastructure and 2.) rigorously documented standardized workflows and standard operating procedures (SOPs) for sample management, ensuring that every aspect of the process is fully described. This integrated approach promises to set a new benchmark for digital sample management, enhancing interdisciplinary collaboration and resource utilization across the research community.

Corresponding Author: Maren Rebke, maren.rebke@awi.de

ID T09

LabID-Prov: Unifying experimental and computational data provenance with LabID

Authors: Jelle Scholtalbers[1], Matthias Monfort[2], Nayeem Reza[2], Laurent Thomas[2], Charles Girardot[2]

[1] EMBL, Heidelberg / LabRise Consulting
[2] EMBL, Heidelberg

LabID [1] is an open-source platform designed to streamline research data management for scientists, research groups, and core facilities of life-science institutes. By integrating sample and dataset management, inventory tracking, and an electronic lab notebook, LabID enables users to organize, annotate, and share experimental data in compliance with FAIR principles. At its core, LabID uses a relational database to document entities—such as instruments, samples, and resulting datasets—and their relationships, constructing a comprehensive knowledge graph that traces data provenance from biological specimens to raw outputs.

We introduce LabID-Prov [2], an extension that enhances this knowledge graph by capturing post-acquisition data processing steps—typically executed via computational workflows. While a variety of platforms and software exist to process raw data with workflows and scripts, there is no central solution to document these processes while preserving data provenance.

With LabID-Prov, the knowledge graph of data provenance has been expanded to incorporate **derived datasets produced by computational workflows**. These datasets are directly linked to their source raw data, while the software, parameters, and processing methods applied are systematically documented. This extension is supported by newly implemented data models for workflow versions and runs, ensuring standardized and comprehensive metadata capture.

To facilitate integration with existing platforms, LabID-Prov supports importing from Git repositories, Galaxy instances, and WorkflowHub, while leveraging Workflow and Workflow Run RO-Crate specifications to simplify sharing and deposition in scientific repositories (WorkflowHub...). Interoperability is further strengthened through an API, a command-line utility, and a Python library, enabling automation and customization. For example, users can import a workflow from Git, enrich its metadata (e.g. license, authors) via LabID's interface, and export it to WorkflowHub, all within a unified ecosystem.

By unifying experimental and computational provenance, LabID-Prov guarantees that derived datasets retain the same level of contextual information as raw data, fostering reproducible computational research.

References:

- [1] https://grp-gbcs.embl-community.io/labid-user-docs
- [2] https://oscars-project.eu/projects/labid-prov-tracking-and-sharing-data-provenance-ro-crate-lab-integrated-data

Corresponding Author: Laurent Thomas, laurent.thomas@embl.de

ID T10

Deploying Kadi4Mat Workflows in Laboratory Environments for Reproducible and Guided Experimental Research

Author: Johannes Steinhülb[1]

Co-Authors: Darya Snihirova[2], Michael Selzer[1], Sven Berger[2]

[1] Karlsruhe Institute of Technology, [2] Helmholtz-Zentrum Hereon

Within the Kadi4Mat ecosystem, Kadi4Mat Workflows support the systematic deployment of laboratory workflows by transitioning execution from previously exclusively local desktop systems (KadiStudio) to remote, containerized infrastructures, such as Docker and Kubernetes. This approach ensures controlled, scalable, and reproducible execution conditions across experimental runs. Interactions via a web interface allows workflows to be controlled and monitored on mobile devices at the point of experimentation. For procedural steps that require human intervention, the workflow provides predefined, structured input fields for capturing user inputs and observations during execution. QR code-based identification of samples and laboratory equipment further improves traceability and reduces manual transcription errors. Experimental data and associated metadata can be persistently stored in the Kadi4Mat repository, enabling comprehensive provenance tracking as well as reproducibility, reuse, and long-term preservation of laboratory research data.

As part of this framework, electrochemical corrosion measurements (e.g., OCP, polarization curves, EIS) are implemented as Kadi4Mat Workflows, enabling guided execution with structured capture of test parameters, sample history and electrolyte composition directly at the point of experimentation. The resulting raw data and metadata are stored persistently in the Kadi4Mat repository, providing end-to-end provenance and ensuring reproducible, comparable corrosion metrics across experimental runs and infrastructures.

Overall, Kadi4Mat Workflows extend electronic lab notebook (ELN) concepts from passive documentation toward guided experiment execution. By integrating structured data capture with workflow-driven experimental guidance, experiments are documented directly during execution rather than retrospectively. This approach embeds ELN functionality into everyday laboratory practice and supports reproducibility by design through the seamless coupling of experimental procedures, data, and metadata.

Corresponding Author: Johannes Steinhülb, johannes.steinhuelb@kit.edu

ID T11

Using Persistent Identifiers (PIDs) as a Vehicle for Achieving Interoperability

Author: Rory Macneil[1]

[1] Research Space

Most research workflows involve use of multiple research tools, services, and IT infrastructure, each addressing one phase of the research life cycle. The lack of interoperability between resources hinders research productivity and prevents streamlined passage of data between tools and sustainable data FAIRification. This presentation discusses implementation of PIDs in RSpace to enhance interoperability between research tools and services used in different phases of the research lifecycle.

RSpace, which has evolved from an ELN into a research orchestration platform, has integrations with 20+ research tools and services, including domain specific research tools, data management planning tools, data repositories, equipment scheduling and colony management tools, computational resources including R, Jupyter Notebooks and Galaxy, protocols.io, and (institutional) data storage solutions. This has resulted in an ecosystem of connected research tools through which data can pass readily and seamlessly.

We will discuss pidification of RSpace and extension of the pids overlay to other tools and services in RSpace's ecosystem. Starting with ORCIDs and RORs, we added support for associating IGSN IDs with physical samples and their metadata, and the ability to pass sample data and associated IGSN IDs from a field data capture notebook (Fieldmark) to RSpace. We then added support for PIDs for projects, RAiDs, and now are incorporating support for instruments, using PIDINST. With IGSNs, RAiD and PIDINST, we describe how product design is driven by consideration of research workflows involving other tools. This includes discussions with developers of other tools and open-source contributors to ensure that information about the PIDs can be shared through RSpace in a streamlined and effective fashion. Finally, we discuss how support for PIDs in a research hub like RSpace enhances an entire ecosystem of tools, services, and research infrastructure using open APIs, SDKs, and MCP tools and making it accessible to both humans and machines.

Corresponding Author: Rory Macneil, rmacneil@researchspace.com

ID T12

TeSSHub - A Federated Training Catalogue Infrastructure

Author: Martin Voigt[1]

Co-Author: Oliver Knodel

[1] Helmholtz-Zentrum Dresden-Rossendorf

The mTeSS-X project ("Multi-space Training e-Support System with eXchange") aims to address one of the central challenges in modern research infrastructures: how to provide coordinated, yet domain-specific training resources across diverse scientific communities. Within the framework of ELIXIR and PaNOSC, the project develops a federated training catalog infrastructure called TeSSHub that connects communities from Photon and Neutron (PaN) Science, the Life Sciences (LS) and beyond, enabling them to share, discover, and reuse training materials and event information across institutional and disciplinary boundaries.

Scientific domains such as LS and PaN share common challenges in training data stewardship, reproducible research, and the application of computational methods. However, their training ecosystems have traditionally evolved independently, often leading to fragmentation and duplication of effort. mTeSS-X directly addresses this by developing a modular, multi-space platform architecture that supports community autonomy while enabling interoperability and content exchange between training catalog instances. The software framework builds upon the ELIXIR Training eSupport System (TeSS) and introduces extensions that facilitate federated content discovery, metadata harmonization, and cross-domain search through standardized APIs and metadata schemas aligned with FAIR principles.

From a technical perspective, mTeSS-X combines robust software engineering with semantic technologies to support training resources that are FAIR (Findable, Accessible, Interoperable, and Reusable) across infrastructures. It introduces an exchange mechanism that allows participating communities to selectively publish, synchronize, and enrich content, while preserving local governance and editorial control. The well-established OAI-PMH 2.0 protocol is supported for import as well as export of content and is used in combination with RDF data utilizing schemas.science, which is based on schema.org properties, and ontologies, such as EDAM, for semantic interoperability.

In this contribution, we will present the conceptual and technical foundations of mTeSS-X. The presentation will also highlight how the exchange feature relates to FAIR training materials.

Corresponding Author: Martin Voigt, m.voigt@hzdr.de

Lecture Hall, 11:00–12:30

Talk Session

Track 3

Ontology-Driven Metadata Harmonization: Closing Semantic Gaps

Chairs: Gerrit Günther[1], Said Fathalla[2]

[1] HMC Hub Matter – Helmholtz-Zentrum Berlin, [2] HMC Hub Information – Forschungszentrum Jülich

This track explores the synergy between domain-specific ontologies and global metadata harmonization. We welcome contributions on the design and engineering of domain and application ontologies that enable the capture of complex data relationships while maintaining alignment with global standards. Submissions may feature case studies on developed ontologies, semantic mapping protocols, and cross-disciplinary solutions for metadata integration.

ID T13

Local Semantics with Global Meaning

Author: K. Gerald van den Boogaart[1]

[1] Helmholtz-Zentrum Dresden-Rossendorf

FAIRification is huge global innovation process. A major challenge is the steep learning curve for every new user especially, if it comes to actionable semantics: Which globally vocabulary or ontology do I use? How can I use it for my data? Essentially every new adopter, who has never collected a single semantic dataset before, is asked to do everything right from the first moment and select the right vocabulary. This however is which strictly speaking would already require an overview.

This talk tries to address this and other problems connected to the introduction of globally harmonized semantics, with a seemingly counterintuitive suggestion, built into the concept of ontology: Split the problem of selecting the vocabulary ontology for collecting and communicating metadata into two: Allow users to collect metadata using fine granular locally defined terms, and to give them a meaning that can be translated into global and harmonized vocabularies later. Both tasks can be performed by creating simple RDFS and OWL statements. Semantic reasoners can do the actual translation.

The short talk will provide several arguments for this proposal: It minimizes hurdles to get started with metadata, because only minimal prior knowledge or decisions are required. It allows adopters to learn about the semantics in a safe local setting, before they need to select the best fitting ontology. It allows to express concepts not yet available in harmonized ontologies and to test new concepts early. It allows to represent the collected metadata in several even contradictory or not yet developed ontologies. It allows to define terms much more specific and concrete than global terms and thus be more concise and better tailored for the actual data needs. It makes the job of data stewards easier, due to a separation of problems and tasks.

Corresponding Author: Martin Voigt, m.voigt@hzdr.de

ID T14

Interoperable Metadata for Describing Health Studies – The NFDI4Health Metadata Schema

Authors: Martin Golebiewski[1], Sophie A. I. Klopfenstein[2], Aliaksandra Shutsko[3], Carina Nina Vorisek[2], Atinkut A. Zeleke[4]

Co-Authors: Elisa Kasbohm[4], Matthias Löbe[5], Guido Prause[6], Carsten Oliver Schmidt[4], Haitham Abaza[1], Wolfgang Müller[1]

[1] Heidelberg Institute for Theoretical Studies, [2] Berlin Institute for Health Research, [3] ZB MED - Information Centre for Life Sciences, [4] University Medicine Greifswald, [5] NFDI4Health, Institute for Medical Informatics, Statistics and Epidemiology (IMISE), University Leipzig, [6] Fraunhofer Institute for Digital Medicine

To structure metadata of health-specific research studies a tailored metadata schema (MDS) has been developed in the context of the German National Research Data Infrastructure for Personal Health Data (NFDI4Health) [1,2]. The MDS supports metadata publication from clinical, epidemiological, and public health studies, while maintaining interoperability with other resources. Designed in a modular fashion, it combines metadata for multiple purposes. Examples are health study design and use case-specific use for nutritional epidemiology, chronic diseases, record linkage, and medical imaging/radiomics. It also comprises bibliographic information and metadata about data sharing and access.

The MDS is based on DataCite for domain-independent metadata, and CT.gov for metadata specific for clinical trials, and other domain-specific schemas. For compatibility with clinical trial registries, the schema was mapped to DRKS and ICTRP. To support metadata exchange with other platforms mappings to the European Clinical Research Infrastructure Network metadata repository, the German Human Genome-Phenome Archive and the Directory of Registries of the European Rare Disease Registry Infrastructure were performed. To further promote semantic and syntactic interoperability of metadata across health research infrastructures, we aligned the MDS with HL7® FHIR® (Fast Healthcare Interoperability Resources) and included standard terminologies in the value sets such as DICOM for the imaging module. The emerging European Health Data Space (EHDS) will play a major role in Europe in the whole medical domain and to prepare for interoperability with the EHDS we have initiated an alignment of the MDS with the HealthDCAT-AP metadata standard underlying the EHDS systems.

Overall, MDS can be applied across a wide range of use scenarios, while maintaining interoperability. NFDI4Health services based on the MDS, such as the German Health Study Hub (https://health-study-hub.de) and the Local Data Hub software (https://www.nfdi4health.de/en/service/local-data-hub.html) can easily interface with external resources.

References:

- [1] https://doi.org/10.2196/63906
- [2] https://www.nfdi4health.de/en/service/metadata-schema.html

Corresponding Author: Martin Golebiewski, martin.golebiewski@h-its.org

ID T15

Integrating Domain Ontologies and Workflow Metadata for Interoperable Computational Experiments

Author: Pavan L. Veluvali[1]

Co-Authors: Jan Heiland[1], Peter Benner[1]

[1] Max Planck Institute for Dynamics of Complex Technical Systems

Computational Science and Engineering relies on complex, multi-step workflows that combine simulations, data processing, and parameter-driven analyses across heterogeneous environments. Ensuring reproducibility in such settings requires not only abstract workflow descriptions but also semantically rich metadata that is interoperable across domains.

In this work, we present MaRDIFlow, a lightweight, metadata-driven workflow framework developed within the MaRDI consortium for research data management in the mathematical sciences. MaRDIFlow executes workflows through explicit input-output relationships between components, enabling structured metadata descriptions at different abstraction levels. Redundant representations of models, code, and data are supported to strengthen reproducibility and reuse.

To address semantic interoperability, MaRDIFlow integrates domain specific ontologies via RESTful APIs and SPARQL endpoints. This allows workflow components and their metadata to be dynamically aligned with standardized vocabularies during both construction and execution. As a concrete example, we integrate Voc4Cat, a domain-specific ontology and SKOS vocabulary from the NFDI4Cat consortium, which serves as a semantic backbone for annotating workflow components and metadata dependencies. Through this integration, knowledge graphs are used to represent and query relationships across workflow layers, supporting automated discovery, validation, and consistent interpretation of data.

The presented use cases demonstrate how combining workflow descriptions with domain ontologies enhances semantic consistency, interoperability, and reproducibility. This work highlights the practical role of domain and application ontologies in building reusable data infrastructures for computational workflows and lays the groundwork for extending MaRDIFlow with additional NFDI ontologies across disciplines.

Corresponding Author: Pavan L. Veluvali, veluvali@mpi-magdeburg.mpg.de

ID T16

MOLSIM: An Interoperable Ontology for Representing Biomolecular Simulation

Authors: Angela Kranz[1], Björn Usadel[1], Fathoni Musyaffa[1], Hannah Dörpholz[1], Holger Gohlke[1], Michele Bonus[2], Rocco Gentile[2]

[1] Forschungszentrum Jülich, [2] Heinrich Heine University Düsseldorf

Molecular dynamics (MD) simulations generate vast amounts of data foundational to structural biology, yet their value is often limited by inconsistent metadata and software-specific formats that create isolated data silos. To address this "semantic gap," we introduce MOLSIM, an interoperable ontology designed to formalize the description of atomistic biomolecular simulations and enhance the implementation of FAIR (Findable, Accessible, Interoperable, Reusable) principles. Developed in adherence to Open Biological and Biomedical Ontologies (OBO) Foundry principles, MOLSIM prevents redundancy by systematically reusing terms from established ontologies such as ChEBI and the Unit Ontology. A core feature of MOLSIM is its software-agnostic labelling, which resolves ambiguities in simulation metadata; for instance, mapping disparate keywords like 'ntt=1' in AMBER and' tcoupl' in GROMACS to a unified Berendsen Thermostat class. The ontology was constructed using a Large Language Model (LLM)-assisted workflow, employing LLM to extract technical terms from software manuals, followed by rigorous expert curation. Currently comprising approximately 2,000 terms, MOLSIM enables simulation data to be structured as Knowledge Graphs. This allows for the seamless integration of MD metadata with external open knowledge bases such as Wikidata, UniProt, and the PDB, providing the necessary semantic granularity to support next-generation community repositories.

Corresponding Author: Pavan L. Veluvali, f.musyaffa@fz-juelich.de

ID T17

Toward FAIR and Reproducible Data Quality Control: A Use Case-Driven Data-Quality Processing Metadata Schema for Time Series Data

Author: Ulrich Loup[1]

Co-Authors: Benjamin Louisot[2], Christof Lorenz[2], David Schäfer[3], Jannis Groh[1] Jürgen Sorg[1], Marc Hanisch[4], Martin Ingenbleek[1], Nicole Büttner[2], Ralf Kunkel[1], Robert Wiesen[3], Romy Fösig

[1] Forschungszentrum Jülich, [2] Karlsruhe Institute of Technology,
[3] UFZ - Helmholtz Centre for Environmental Research, [4] GFZ Helmholtz Centre for Geosciences

High-quality environmental time series data require transparent, reproducible, and well-documented quality control (QC) workflows that integrate automated procedures and expert judgment. While many QC frameworks offer algorithmic methods, the processing information explaining how data quality decisions are made — including parameterization, flag semantics, and manual interventions — is often not formalized enough to be easily reused, reproduced, or exchanged across infrastructures.

In this talk, we present a metadata schema for time series data that enables FAIR and reproducible data quality processing. The schema is designed to describe QC methods, execution contexts, and resulting quality flags in a machine-actionable and interoperable manner. It employs the [OGC SensorThings API] [1] data model enhanced by the [STAMPLATE schema] [2] and the concepts established in the [SaQC] [3] framework. The schema follows the linked-data approach and aligns with standards such as the [W3C Data Quality Vocabulary] [4].

The design of the proposed schema is motivated by concrete use cases for QC of time series data from the TERENO and ACTRIS observation networks. These use cases include detailed analyses of existing automated and manual QC workflows. By comparing and abstracting these practices, we derive common requirements and design patterns for representing QC processing information in a FAIR and reproducible manner. The resulting schema can be used straightforwardly with SensorThings API services and mapped into NetCDF files that align with the [Helmholtz Metadata Guidelines for NetCDF] [5]. It can also be used with [RO-Crates] [6], embedding files in CSV format, for example.

Our metadata schema lays the foundation for a community-driven, FAIR, and reproducible quality control solution. Our goal is to integrate the requirements of other communities and develop a web application that allows users to visually inspect and flag time series data in a manner consistently with our schema.

References:

[1] https://portal.ogc.org/files/?artifact_id=111659&version=2

[2] https://github.com/HMC-STAMPLATE/JSONSchema

[3] https://rdm-software.pages.ufz.de/saqc/index.html

[4] https://www.w3.org/TR/vocab-dqv

[5] https://hmg-netcdf.readthedocs.io/en/latest/index.html

[6] https://www.researchobject.org/ro-crate

Corresponding Author: Ulrich Loup, u.loup@fz-juelich.de

ID T18

Bridging Semantic Gaps: Linked Open Data Modeling of Theresa Hak Kyung Cha's Dictee

Author: Eunji Park[1]

[1] Berlin University of the Arts

Cultural heritage metadata systems face fundamental challenges

representing experimental artworks that resist conventional classification. This research proposes a domain-specific ontology framework to address these challenges, using Theresa Hak Kyung Cha's *Dictee* (1982) – a seminal postcolonial feminist artists' books – as a critical case study.

Dictee is a hybrid work combining untranslated Korean, French, and English text with uncaptioned photographs, organized around nine Greek muses. As a postcolonial feminist intervention, it bypassed patriarchal art institutions through independent publishing while articulating diasporic Korean women's experiences under colonial and postcolonial conditions. Current library metadata (MARC) reduces this work to "fiction," while controlled vocabularies (LCSH, AAT) lack URIs for domain-specific concepts central to the work: "diasporic memory," "translational violence," "dictaphonic structure," and "anti-documentary desire."

The proposed framework develops a domain ontology for artists' books by strategically reusing and aligning existing ontologies:

- Schema.org for basic bibliographic structure
- Linked.art for visual arts context
- CIDOC CRM for cultural heritage semantics
- BIBFRAME for library interoperability

Our semantic alignment strategy implements language tags (@lang="ko", "en", "fr") for multilingual integrity, custom properties for unmappable concepts, and structured relationships connecting nine chapters to historical figures (Yu Gwan-sun, Joan of Arc), literary influences (Gertrude Stein, Marguerite Duras), and historical events (March First Movement, Japanese colonial period).

Preliminary implementation demonstrates how this approach addresses structural gaps in general-purpose vocabularies that systematically marginalize postcolonial feminist epistemologies. The resulting JSON-LD dataset structure enables automatic discovery across disciplines (literature, art history, Asian American studies) through SPARQL queries, while maintaining FAIR compliance and cross-institutional interoperability. This work-in-progress shows how domain ontology design can balance community-specific epistemological needs with broader semantic web standards, providing a replicable model for representing marginalized cultural objects in digital archives.

Corresponding Author: Eunji Park, ejpark8073@gmail.com

Lecture Hall, 13:30 - 14:30

Keynote

Constructing Enterprise RDF Knowledge Graphs: Foundations for Neuro-Symbolic AI

Dr. Jan Portisch[1]

[1] Head of Content Infrastructure at SAP Global Content Group

In his keynote, Jan Portisch will explore the construction and operationalisation of RDF knowledge graphs in complex corporate environments. Drawing on extensive experience in integrating heterogeneous and evolving data sources, he will discuss architectural choices, practical strategies, and common challenges in building enterprise-scale knowledge graphs. The talk will also highlight how such knowledge graphs can support AI applications by improving semantic grounding, robustness, and explainability, and will provide an outlook on neuro-symbolic AI as a bridge between statistical learning and symbolic reasoning.

Foyer + 1 Floor, 17:00 - 19:00

Poster & Demo Session

Track 2

Software Interoperability for (Meta)data Acquisition

Chair: Martin Held[1]

[1] HMC Project - Helmholtz-Zentrum Hereon

Modern research labs rely on diverse (meta)data acquisition systems - from instrument software to ELNs, LIMs, and workflow managers. To capture the scientific process coherently, these systems must interoperate seamlessly.

This session focuses on practical concepts and existing solutions for bridging across these platforms - from APIs to file-based integration – including the use of semantic or machine-learning tools to enhance interoperability.

ID D01

Bridging Repositories, ELNs and Semantic Data Management: A LinkAhead-based Use Case for 3D Additive Manufacturing

Author: Florian Spreckelsen[1]

Co-Authors: Alexander Schlemmer[1], Frank Tristram[2], Henrik tom Wörden[1]

[1] IndiScale GmbH, [2] Karlsruhe Institute of Technology

Managing (meta-)data across interdisciplinary research collaborations often involves using a variety of software tools for storage and sharing. Maintaining data accessibility and synchronization between different sites, working groups, and institutes presents a significant challenge. We developed a solution based on the open source software LinkAhead that combines meta data from different repositories into a single research data management system (RDMS). The meta data import tool was created using the extensible crawler framework provided by LinkAhead. This enables us to import meta data from four different repositories and ELN systems used by the Cluster of Excellence 3D Matter Made to Order (3DMM2O). Where possible, APIs of repositories and ELNs can be used to directly upload meta data entries from LinkAhead (e.g., via export to DataCite XMLs). This allows (semi-)automatic workflows, such as described in the following example:

- A researcher enters meta data in an ELN
- These are imported into LinkAhead by the crawler (possibly enriched by data from other sources)
- They can be sent directly to a repository for publication.

In addition, the RDMS is built in a way that data models and crawler definitions can also be extended and adapted to future requirements by the researchers at any time. Its basic functionality contains a graphical web-interface, as well as an API for automated queries which provides intuitive searching and querying and using metadata from all linked systems.

Corresponding Author: Florian Spreckelsen, f.spreckelsen@indiscale.com

ID D02

Lowering the Barrier to Metadata: Collecting and Ingesting Sample Data for Real-Life Experiments with JSON-Schema Forms

Author: Thomas Gruber[1]

Co-Authors: Eric Hirschmann[1], Subhan Ali[1]

[1] Helmholtz Center Dresden-Rossendorf

Preparing for beamline experiments often requires providing sample information alongside the proposal, especially for external users who may not be present during the experiment. A standardized form based on a JSON schema simplifies this process: it minimizes follow-up questions through built-in validation and constraints, and directly generates a machine-readable JSON object. This object can be seamlessly ingested into the Electronic Lab Notebook (ELN) or sample database, even if the ELN is inaccessible or too complex for untrained users.

The solution is a web-based application designed for both external and internal use. For internal users, a tablet-optimized interface offers guided data entry and direct submission to the ELN. By reusing existing JSON schemas from the ELN, no additional effort is required to maintain the web forms. The presentation will showcase a successful practical implementation of the solution, demonstrated through its application at the ELBE radiation source beamline at HZDR.

Technically, the application leverages json-editor [1], which extends JSON Schema with UI elements, ensuring a user-friendly data entry experience.

Reference:

- [1] https://github.com/json-editor/json-editor

Corresponding Author: Thomas Gruber, t.gruber@hzdr.de

ID D03

The Text+ Registry: Facilitating Cross-Domain Resource Discovery

Authors: Tobias Gradl[1], Daniela Schulz, Kilian Hensen, Leon Fruth, Martin Sievers, Nils Geißler

[1] University Bamberg

The fragmented landscape of scholarly resources presents significant challenges for researchers seeking textual and linguistic data. Resources developed across different projects and institutions often follow varying standards, lack common access points, and suffer from limited interoperability. The Text+ Registry [1] addresses these challenges as a unified cataloguing system within Germany's National Research Data Infrastructure [2] (NFDI), specifically designed to enhance FAIR data practices across the domains of collections, lexical resources, and editions.

Built on three architectural principles—metamodeling, model-driven design, and information layering—the Registry handles multiple data models simultaneously without requiring system-wide redesign when integrating new resource types. Unlike traditional catalogues focused on specific disciplines or formats, our metamodeling approach enables the system to accommodate heterogeneous metadata while maintaining semantic integrity. The model-driven design generates application components from formal models, ensuring technology independence and scalability beyond the current scope.

The Registry's information layering system preserves full provenance while creating enriched resource descriptions. Metadata from multiple sources (including AGATE [3] research information system, catalogues, and databases) are ingested, then stacked as distinct layers with transparent attribution. Manual expert curation supplements automated aggregation without compromising source integrity. This approach produces consolidated descriptions offering more than the sum of their parts.

Integration with authority files (e.g. GND [4]) and an instance of the Research Software Directory [5] creates a rich semantic network connecting textual resources with software tools and institutional actors. The Registry's Search API enables external services like PhilFinder [6] to leverage harmonised metadata through standardised queries, while DataCite [7] mappings facilitate cross-domain discovery.

Our poster + demo showcases the technical implementation addressing harmonisation challenges, practical metadata integration workflows, and the transformation of heterogeneous information into rich resource descriptions. It also places the Registry as a component of an interconnected research infrastructure within NFDI and beyond.

References:

- [1] https://registry.text-plus.org/default
- [2] https://www.nfdi.de
- [3] https://agate.academy/research-projects.html
- [4] https://www.dnb.de/EN/Professionell/Standardisierung/GND/gnd_node.html
- [5] https://research-software.cceh.uni-koeln.de/software
- [6] https://philportal.de/philfinder
- [7] https://datacite.org

Corresponding Author: Tobias Gradl, tobias.gradl@uni-bamberg.de

ID D04

Timely – A Domain Specific Language Extension of HTML to Visualize and Query Time Series Data using W3C Web Components

Author: Andreas Schmidt[1,2]

Co-Authors: Marius Hochmuth[2], Anis Koubaa[2], Fabia Martens[2], Nan Liu[2], Philipp Schmurr[2], Karl-Uwe Stucky[2], Wolfgang Süß[2]

[1] University of Applied Sciences Karlsruhe, [2] Karlsruhe Institute of Technology

We have developed Timely, a domain specific extension of HTML, that enables the visualization, analysis and subsequent export of metadata-enriched time series data within arbitrary websites. The implementation is based on the W3C standard *Web components*, so that the extension can be used in all current browsers without installing additional plugins. Web components are user-defined HTML elements that are implemented as JavaScript classes and executed in the user's browser. They can then be used like normal HTML elements such as *table, h1,* etc. The components can not only have a visual representation but also communicate with backend server processes. A major advantage of this approach is that the web components developed in this way offer website developers a simple yet powerful declarative interface for implementing their functionality, enabling even non-programmers to create and customize simple web applications according to their needs. The developed components allow flexible querying and presentation of time series data in graphical or tabular form. Furthermore, metadata from InfluxDB, such as all available databases, associated measurements (a data structure typically called table in other database systems) and their structures, can be read out and visualized. Like normal HTML elements, web components can be parameterized using attributes. A key feature of our components is the communication between components and their dynamic behaviour in response to changes. When specifying attribute values, property values of other elements can be integrated, which are then monitored for changes. If the value of a property embedded in an attribute value changes, the component is informed and can respond appropriately, for example by executing a new query or adjusting the representation of the visualized data. Similarly, child components are informed of state changes by their ancestor elements. This allows users to easily develop interactive HTML pages for visualising and exporting time series data.

Corresponding Author: Andreas Schmidt, andreas.schmidt@kit.edu

Track 5

Advancing FAIR Metadata with AI: Methods, Challenges, and Synergies

Chair: Santiago Casas[1]

[1] HMC Hub Aeronautic, Space and Transport - Deutsches Zentrum für Luft- und Raumfahrt (DLR)

This session explores the synergy between metadata and AI, where tools such as large language models (LLMs), retrieval-augmented generation (RAG), and machine learning methods enhance metadata quality, completeness, and interoperability. At the same time, FAIR and rich metadata improve AI performance by providing structured, unambiguous context. We will discuss innovative methods, practical challenges, and collaboration opportunities in making metadata workflows scalable, user-centric, and ready for AI-driven research.

ID P04

Metadata Extraction with LLMs for DCAT-AP+ based Research Data Management in Process Engineering and Catalysis

Author: Marc Völkenrath[1]

Co-Authors: Alexander S. Sommer-Behr[1], Hendrik Borgelt[1], Norbert Kockmann[1], Simon Clemens[1]

[1] Technical University of Dortmund

Efficient and FAIR (Findable, Accessible, Interoperable, Reusable) research data management is essential for sustainable data reuse and reproducibility in catalysis research and chemical engineering. Heterogeneous data sources, inconsistent documentation practices, and insufficiently standardized metadata continue to complicate semantic interoperability and long-term accessibility of experimental data.

Within the "Nationale Forschungsdateninfrastruktur" (NFDI) this work presents a workflow that automates the extraction, validation, and semantic enrichment of metadata assisted by Large Language Models (LLMs) from scientific datasets in various file formats.

The workflow applies a customized Ollama-LLM [1] combined with the DCAT-AP+ [2] metadata schemas and the Voc4Cat [3] domain vocabulary. Relevant domain concepts are identified through lexical and semantic matching and assigned to schema-compliant metadata structures. Missing concepts are detected using definition-based reasoning with existing vocabularies and ontologies [4]. If no suitable matches are found, the LLM proposes new, standard-compliant candidate concepts, which are subsequently reviewed and validated by domain experts to ensure semantic correctness and consistency.

The user validated metadata are exported in a standardized, machine-readable representation compatible with existing research data infrastructures. The feedback of domain experts is used to extend the metadata schemas and the controlled vocabularies and ontologies. The resulting enriched metadata and semantic resources enable the construction of interoperable knowledge graphs that can be validated using SHACL and queried via SPARQL to support enhanced literature search, improved research planning and knowledge discovery.

References:

- [1] https://ollama.com/ (Accessed on September 30, 2025).
- [2] P. Strömert, H. Borgelt, M. Doerr, D. Linke, nfdi-de/dcat-ap-plus: Release 0.1.0rc3 (same as rc2), Zenodo 2025.
- [3] D. Linke, N. Moustakas, M. Doerr, J. Schumann, M. Götte, H. Borgelt, M.Nentwich, C. Terboven, F. Flecken, nfdi4cat/voc4cat: Release 2025-10-14, Zenodo 2025.
- [4] A. S. Behr, H. Borgelt, N. Kockmann, Journal of cheminformatics 2024, 16 (1),16. DOI: https://doi.org/10.1186/s13321-024-00807-2

Corresponding Author: Alexander S. Sommer-Behr, alexander.behr@tu-dortmund.de, Marc Völkenrath, marc.voelkenrath@tu-dortmund.de

ID P05

Ontology-driven Data Curation and Knowledge Modeling for Catalyst Layers in Polymer Electrolyte Fuel Cells

Author: Marjan Kohandani[1]

Co-Authors: Kourosh Malek[1], Michael Eikerling[1], Mohammad J. Eslamibidgoli[1]

[1] Forschungszentrum Jülich

Research data management in the hydrogen technology field is challenging because large volumes of heterogeneous data are produced [1]. Electrochemical technologies such as fuel cells and electrolyzers are multicomponent devices, with various manufacturing routes being followed and a wide range of characterization and performance measurements applied. The governing phenomena span multiple length and time scales, creating a complex parameter-property space [2], while data are reported with inconsistent standards and formats. In this work, we build a FAIR and searchable knowledge graph for a concrete use case: catalyst layers in polymer electrolyte fuel cells. Our approach consists of three phases. In the first phase, we define the scope of a domain research question and build an initial ontology. In the second phase, we screen the literature and create a PDF corpus, then extract and curate the data into a structured, machine-readable format guided by the ontology, including terminology alignment and unit harmonization for cross-study comparability. In the third phase, we map the curated dataset into a Neo4j knowledge graph and release it as a FAIR resource. Overall, as will be shown at the conference, this workflow enables standardized, traceable, and AI-ready datasets that can be reused across studies to accelerate data-driven discovery and decision-making.

References:

- [1] Dreger, M., Eslamibidgoli, M. J., Eikerling, M. H., & Malek, K. (2023). Synergizing ontologies and graph databases for highly flexible materials-to-device workflow representations. Journal of Materials Informatics, 3(1), N-A
- [2] Liu, H., Ney, L., Zamel, N., & Li, X. (2022). Effect of catalyst ink and formation process on the multiscale structure of catalyst layers in PEM fuel cells. Applied Sciences, 12(8), 3776.

Corresponding Author: Marjan Kohandani, m.kohandani@fz-juelich.de

ID P46

HVRLocator: A Computationally Efficient Tool for Identifying Hypervariable Regions in Large 16S rRNA Datasets

Authors: Clara Arboleda-Baena[1,2], Felipe Borim Correa[1], Joao Pedro Saraiva[1], Santiago Castillo-Rivadeneira[2], Jonas Coelho Kasmanas[1], Antonis Chatzinotas[1,2,3], Stephanie D. Jurburg[1,2]

[1] UFZ - Helmholtz Centre for Environmental Research, [2] German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, [3] Leipzig University

Metabarcoding of the 16S rRNA gene is widely used to assess microbial diversity due to its cost-effectiveness and efficiency. However, publicly available 16S rRNA metabarcoding datasets often lack standardized metadata, particularly information on the sequenced hypervariable regions or primers used, which are critical to their accurate reuse. To address this, we present HVRLocator, a computational tool that (1) identifies the start and end positions of 16S rRNA amplicons, (2) determines their corresponding hypervariable regions, and (3) detects the presence of primer sequences. This tool was validated on four datasets comprising 41,513 samples generated with different primers and sequencing platforms.

HVRLocator can process archived 16S rRNA sequences from NCBI SRA at an average rate of 6.5 samples per minute. Validation showed it reliably detects amplicon start and end positions across datasets sequenced with different primers and platforms, achieving 100% accuracy within single-platform studies and correctly revealing length heterogeneity across platforms. It also flagged misannotated metadata and problematic sequences, underscoring its value as a sequence data curation tool. Finally, HVRLocator can select comparable sequences to build large 16S rRNA amplicon databases spanning the same hypervariable region, facilitating cross-study comparisons.

In conclusion, this tool overcomes unreliable metadata by accurately identifying 16S rRNA amplicon start and end positions, determining hypervariable regions, and detecting primer sequences, thereby enabling accurate curation and large-scale processing of 16S rRNA data for reliable and reproducible microbial studies, syntheses, and meta-analyses.

Corresponding Author: Clara Arboleda Baena, clara.arboleda@ufz.de

Track 8

From Harmonisation to Action(ability)

Chair: Thomas Jejkal[1]

[1] HMC FAIR Data Commons – Karlsruhe Institute for Technology

Harmonising metadata is the first step in turning data structures into machine actionable items. In this track we explore packaging and exchange formats such as RO-Crates, versatile and uniform query interfaces like SPARQL, and emerging concepts such as FAIR Digital Objects and how they support scientific impact. We particularly encourage submissions presenting practical applications, early showcases, or lessons learned from putting harmonised (meta)data into practice.

ID D13

An Editor for Research Object Crates to Empower FAIR Data Sharing

Author: Christopher Raquet[1]

[1] Karlsruhe Institute of Technology

Research Object Crate (RO-Crate) is a packaging specification for describing and sharing research data. It packages data together with associated metadata in a FAIR manner. For wide community adoption, the format must be tangible for its users. Programmatic APIs are often insufficient for interactive exploration, on-the-fly editing, and introspection by researchers, data stewards, or curators. Although RO-Crate editors have been developed in the past, the most prominent tool (Describo) has recently been discontinued.

To fill this gap in RO-Crate tooling, we present NovaCrate, a web-based editor for viewing, editing, and validating RO-Crates. The Entity Editor shown in the attached screenshot is the core piece of NovaCrate. It allows editing all metadata entities through a tabbed interface, including dynamic property fetching from custom vocabularies and file previews for common formats. Other features of NovaCrate include a File Explorer and a Graph of the metadata relationships. NovaCrate emphasizes an intuitive user interface as well as good usability. This is vital to ensure the accessibility of the editor for a wide range of people with varying knowledge of the RO-Crate specification.

By allowing researchers to directly interact with the metadata of their RO-Crates, NovaCrate makes the RO-Crate concept tangible. NovaCrate is equipped with the groundwork to further extend it for distributed and collaborative workflows, e.g., through the integration of NovaCrate into RO-Crate repositories.

Corresponding Author: Christopher Raquet, Christopher.raquet@kit.edu

ID D14

Download of Time Series Data from SensorThings API Services as a FAIR Digital Object

Author: Martin Ingenbleek[1]

Co-Authors: Christof Lorenz[1], Jürgen Sorg[2], Ralf Kunkel[2], Ulrich Loup[2]

[1] Forschungszentrum Jülich, [2] Karlsruhe Institute of Technology

Time series observations from sensors and monitoring infrastructures are increasingly published via the OGC SensorThings API (STA) [1], providing large infrastructures such as the Earth & Environment DataHub with a standardized and interoperable interface for accessing environmental data. While STA enables uniform access to distributed services, users are still required to possess domain-specific knowledge to locate relevant data endpoints, interfaces and APIs, construct tailored queries, and interpret response structures.

We present a service that allows you to download selected STA time series as a FAIR digital object. In accordance with the WorldFAIR Cross-Domain Interoperability Framework [2] recommendations, we designed the downloaded data package as a self-contained RO-Crate [3]. Each package contains a structured folder layout with one file per time series in a specified format, such as CSV. A top-level metadata file accompanies the package. This file links the dataset to its origin and includes a description using available STA metadata. It also contains contextual information linking the folder contents to the metadata. Additionally, the download service uses the STAMPLATE metadata schema to enrich the RO-Crate metadata with information such as provenance, licensing, and device setup details. To improve citability and interoperability with research data repositories such as Dataverse repositories (e.g., Jülich Data [4]) or PANGAEA [5], a DataCite-compliant description of the dataset is also added.

By offering STA data in widely used, long-established formats such as CSV, scientists can use the data without altering their analysis workflows. The RO-Crate structure provides the necessary metadata for research in a transparent and organized way. Its linked-data format transforms the resulting package into a portable, semantically sound research object that supports automated processing and interpretation. Thus, our STA download service connects established scientific workflows with advanced digital ecosystems, promoting the advancement of environmental research towards FAIRness.

References:

- [1] https://www.ogc.org/de/standards/sensorthings
- [2] https://cross-domain-interoperability-framework.github.io/cdifbook/introduction.html
- [3] https://www.researchobject.org/ro-crate
- [4] https://data.fz-juelich.de
- [5] https://www.pangaea.de

Corresponding Author: Martin Ingenbleek, m.ingenbleek@fz-juelich.de

ID P21

A Flexible, User-Friendly Semantic Uploader for FAIR-Compliant RDF Workflows in Catalysis Research

Author: Yuliia Dikova[1]

Co-Author: Preston Rodrigues[1]

[1] University of Stuttgart

Extended metadata describing experimental context and provenance are becoming increasingly important alongside primary datasets. However, their adoption remains limited due to the technical complexity of storing, querying, and managing semantic data.

Within the NFDI4Cat initiative, several triplestore solutions were evaluated based on a set of defined requirements [1]. Apache Jena Fuseki [2] was selected as a lightweight, standards-compliant backend that integrates well with the NFDI4Cat software ecosystem [3], but its minimalistic interface provides limited support for dataset exploration and metadata inspection, creating a barrier for researchers without expertise in semantic technologies.

To bridge this gap, we developed a lightweight, user-friendly RDF Uploader that enables intuitive management of datasets and graphs, supports the upload of RDF and other Fuseki-compatible serialisation formats (e.g., Turtle, OWL, JSON-LD), and provides full CRUD operations on the triple store without manual endpoint configuration or SPARQL query formulation. In addition, the integration of WebVOWL [4] enables interactive exploration of ontology graph structures, while interoperability with tools such as TRIQ [5] ensures a seamless transition from metadata creation to semantic storage and subsequent use.

To support practical research scenarios, the application provides automated ingestion of RDF serialisation formats into a Fuseki-based triple store with user- and institution-specific storage paths; ontology graph visualisation with WebVOWL for schema-level analysis; template-based SPARQL querying with predefined and user-modifiable templates that allow data exploration without SPARQL knowledge; and flexible search scopes across user-level and institution-level graphs.

The tool was validated on a test triple store populated with real catalysis-related metadata, demonstrating consistent querying, reusable templates, and effective schema-level inspection through visualisation.

The presented approach shows how harmonised semantic metadata can be translated into actionable research workflows using lightweight, user-centric tools. The RDF Uploader lowers the complexity of semantic infrastructures and supports digitalisation and FAIR adoption within the NFDI4Cat community.

References

- 1) Dikova, Y., Kushnarenko, V., Petrenko, T., Rodrigues, P., Huskova, N., and Bönisch, T. (2025). Triplestore software evaluation results. Zenodo. https://doi.org/10.5281/zenodo.152302692)
- 2) Apache Jena Fuseki, 2025. [Online]. Available: https://jena.apache.org/documentation/fuseki2
- 3) S. Schimmler et al., Transforming Catalysis with a Unified Digital Ecosystem. 2nd Conference on Research Data Infrastructure (CoRDI), Aachen, Germany. https://doi.org/10.5281/zenodo.16736060

- 4) WebVOWL: Visualizing ontologies on the Web. [Online]. Available: https://github.com/VisualDataWeb/WebVOWL
- 5) T. Petrenko, Y. Dikova, N. Huskova, V. Kushnarenko, P. Rodrigues, and T. Bönisch, "NFDI4Cat Web Application for User-Friendly Generation and Browsing of Catalysis Metadata within a Semantically Rich RDF Framework," Zenodo, Sep. 2023. Available: https://zenodo.org/records/14389435

Corresponding Author: Yuliia Dikova, yuliia.dikova@hlrs.de

ID P22

Bridging the Gap: Integrating Outdoor Testing Time-Series with Synthesis Metadata for Perovskite Solar Cells

Author: Paolo Graniero[1]

Co-Authors: Carolin Ulbrich, Eva Unger, Marko Remec, Michael Götte, Rutger Schlatmann

[1] Helmholtz-Zentrum Berlin

At PVcomB (Helmholtz-Zentrum Berlin), we are establishing a comprehensive data management framework for the outdoor testing of perovskite solar cells. A major bottleneck in materials science is the disconnection between synthesis data and long-term performance metrics. Addressing this, our initiative supports the increasing drive by research groups at HZB and partner institutes to introduce NOMAD (Novel Materials Discovery) directly into manufacturing laboratories, ensuring standardized metadata capture at the source.

We are deploying a scalable infrastructure to handle high-frequency time-series data from outdoor monitoring. The core innovation lies in the active integration with NOMAD: our system does not merely store testing data but links it to specific synthesis and manufacturing entries within the NOMAD repository. This linkage creates a "digital thread" that connects the initial synthesis conditions of a solar cell directly to its environmental performance and degradation history.

This prototype demonstrates a practical workflow for unifying the sample lifecycle. By bridging the gap between operational time-series databases and the rich metadata capabilities of NOMAD, we enable a feedback loop where synthesis parameters can be correlated with outdoor stability. This work serves as a blueprint for cross-institutional collaboration, moving towards a fully interoperable, FAIR-compliant data ecosystem where manufacturing and characterization processes actively "talk" to each other.

Corresponding Author: Paolo Graniero, paolo.graniero@helmholtz-berlin.de

ID P23

Handle and Share Information on Numerical Earth-System Simulations to Improve Their FAIRness

Authors: Klaus Getzlaff[1], Markus Scheinert[1]

[1] GEOMAR Helmholtz Centre for Ocean Research Kiel

Numerical earth-system simulations are often documented and shared on a personal level while details on simulations are rarely easily findable, even within institutions or departments. This is a hurdle for possible collaboration and re-use, especially for non-experts and people from other research fields. The publicly available documentation of simulations, if any, is mostly part of a scientific article or report, often with a lack on specific details and information. On the other hand, the common approach for a documentation of numerical simulations by throwing everything into a git repository has its weakness too. It completely depends on the creator or curator of the repository to ensure, everything is in place to reproduce or at least repeat the simulation.

In order to easily share numerical earth-system simulations in a reliable way and in compliance with the FAIR principles, we propose a schematic approach, called SimShare. The foundation of SimShare is a standardised, machine readable metadata file which describes the most essential characteristics including references via persistent identifier (PID) for code, modules and input files. The SimShare file can easily be integrated into your institutional metadata repository to improve findability and simple access to the simulation details which in turn could increase the visibility to foster cooperation and re-use.

Here we will describe our institutional workflow to handle and share information on numerical simulations. We will introduce the proposed metadata standard (SimShare) and how we integrate it in the institutional metadata platform.

Corresponding Author: Klaus Getzlaff, kgetzlaff@geomar.de

ID P24

Lightweight Shell Script ‘deploy2zenodo’: Automating FAIR Data Publishing for Zenodo and InvenioRDM APIs

Author: Daniel Mohr[1]

[1] Deutsches Zentrum für Luft- und Raumfahrt (DLR)

The Helmholtz Metadata Collaboration (HMC) advocates for FAIR data stewardship, yet manual publishing workflows -- especially for versioned datasets or software requiring persistent identifiers (PIDs) -- pose persistent barriers for researchers. To address this, we introduce `deploy2zenodo` (https://gitlab.com/projects/51392274; https://doi.org/10.5281/zenodo.10112959), a lightweight, portable shell script designed to automate FAIR-compliant data dissemination via Zenodo’s REST API and, via a configurable flag, the InvenioRDM API (e.g., for DLR’s upcoming local instance). By integrating seamlessly into CI/CD pipelines, the script reduces manual overhead while ensuring alignment with FAIR principles (cf. https://doi.org/10.3289/HMC_publ_01 and https://doi.org/10.3289/HMC_publ_04).

Originally tailored for Zenodo, the script now supports InvenioRDM deployments, bridging current workflows with future Helmholtz-wide infrastructures. This dual compatibility accommodates evolving APIs -- such as Zenodo’s planned adoption of InvenioRDM API enhancing metadata fields (e.g., ROR institution codes, cf. https://github.com/zenodo/zenodo/issues/2544) -- while maintaining backward and forward compatibility.

Environment variables enable flexible configuration, allowing direct publication from Git repositories (e.g. GitLab) via CI/CD pipelines/workflows and linking software versions to DOIs. As the DLR’s InvenioRDM instance advances, `deploy2zenodo` serves as a pragmatic tool to harmonize FAIR adoption across repositories, advancing interoperability and reducing researcher burden.

Corresponding Author: Daniel Mohr, daniel.mohr@dlr.de

ID P25

OSCARS Improving Metadata in Photon and Neutron Science

Authors: A. Paul Millar[1], Melanie Nentwich[1], Patrick Fuhrmann[1]

[1] German Electron Synchrotron (DESY)

The OSCARS project brings together world-class European Research Infrastructures (RIs) in the ESFRI roadmap and beyond to foster the uptake of Open Science in Europe. OSCARS underscores the critical role of persistent identifiers (PIDs) in ensuring the findability, accessibility, interoperable, and reusability (FAIR) of research data. By assigning globally unique PIDs, OSCARS enhances long-term utility, enabling researchers to leverage data for novel applications.

Activities are centred on consolidation, composability, and engagement. Through an in-depth analysis of existing community services, OSCARS has mapped the current landscape and identified typical research workflows within the Photon and Neutron Open Science Cluster (PaNOSC). We aim to seamlessly integrate these services, fostering interoperability while introducing novel functionalities beyond their original scope; particularly in metadata enrichment and PID assignment.

A key initiative expands PID adoption for beamlines and experimental techniques. While instrument PIDs (PIDINST) are already assigned to diverse instruments, beamline PIDs remain an evolving area, particularly regarding the creation of persistent landing pages with standardised metadata. As such, we are exploring Unified Data Sheets: web-based overviews on technical details of the DESY beamline setup, which are aligned with the Ways for Light (WFL) initiative. We seek to separate the representation (webpage) from the underlying data (database) in order to streamline updates for beamline staff and user offices, ensuring consistency across multiple services, including beamline webpages, DOI services, and WFL metadata.

Additionally, OSCARS has enhanced the PaN Experimental Techniques ontology (PaNET). This ontology enables targeted applications, such as the DESY beamline finder, which helps users match beamlines to their experimental needs by querying available techniques. By integrating technique-specific metadata into datasets, PaNET can improve data discoverability while also enabling automated, technique-tailored metadata schemas during measurements.

In summary, OSCARS advances metadata interoperability and PID standardization, fostering a FAIR-aligned infrastructure that supports reproducible, collaborative, and innovative research.

Corresponding Author: Melanie Nentwich, melanie.nentwich@desy.de

ID P26

Semantic x-Lab –Semantic Search on Ontology-based Descriptions of Laboratory Workflows, Resources and Data in Helmholtz

Authors: David Pape[1], Felix Mühlbauer[2], Manja Luzi-Helbing[2], Marc Hanisch[2], Martin Voigt[1], Oliver Knodel[1]

[1] Helmholtz Center Dresden-Rossendorf, [2] GFZ Helmholtz Centre for Geosciences

This contribution presents Semantic x-Lab, a joint Helmholtz project that investigates semantic search and knowledge discovery based on ontology-driven descriptions of laboratory workflows, resources, and research data. The work addresses a common challenge across Helmholtz centers and beyond: while substantial metadata is collected in proposal and publication systems as well as in platforms describing experimental processes and workflows, this information is distributed across heterogeneous and often isolated systems. At the same time, metadata standards have emerged to improve provenance information of research outcomes and to describe methods and experimental steps in a structured manner. However, the lack of semantic integration across systems limits cross-domain exploration and reuse. Semantic x-Lab tackles this challenge by interlinking existing metadata sources and enhancing them in a shared semantic context.

The central outcome of the project is a distributed knowledge graph that enables information to be explored across institutional, system, and disciplinary boundaries. By relating workflows, resources, and data semantically, the approach supports the discovery of connections and knowledge that were previously considered unrelated or confined to specific research contexts. The knowledge graph is developed in a user-centered co-design process involving software developers and use-case partners from laboratories, large-scale facilities, and multiple research domains, ensuring both technical soundness and practical relevance.

Semantic x-Lab is a collaboration between the Helmholtz centers HZDR, GFZ, and GSI and is funded within the HMC Project Cohort 2024. The contribution outlines the project's motivation, conceptual approach, and current status, and illustrates its potential for semantic search and cross-domain knowledge discovery within Helmholtz research.

Corresponding Author: Oliver Knodel, o.knodel@hzdr.de

ID P27

Streamlining Metadata Annotation within FDOs

Author: Angela Kranz[1]

[1] Forschungszentrum Jülich

Comprehensive metadata is essential to ensure data findability and reusability in accordance with FAIR principles. However, keeping pace with the myriad of checklists defined by different research communities and repositories can be an overwhelming task. To address this, DataPLANT, the German NFDI for plants, facilitates the provision of precise metadata through the use of Annotated Research Contexts (ARCs). An ARC serves as a FAIR Digital Object (FDO) that bundles data, metadata, and workflows into a single package, ensuring research remains transparent and reproducible.

Central to this ecosystem is the Swate annotation tool for Excel, which is used within an ARC to create annotation tables based on the Investigation Study Assay (ISA) framework. Swate provides over 40 pre-assembled, curated templates covering a wide range of needs, including minimum information standards like MIAPPE for plant phenotyping and MIMS for metagenome sequences, as well as checklists for major repositories like the European Nucleotide Archive and MetaboLights. These templates utilize integrated ontology terms and prefilled examples to simplify data entry while ensuring consistency.

While researchers can use these templates as provided, the system remains flexible; templates can be modified, extended, or combined to suit specific project requirements. Different research communities can also contribute their own standard operating procedures through documented GitHub workflows. To assist users in navigating these resources, DataPLANT offers a metadata quiz to help identify relevant standards based on the specific organism and data type. Finally, the submission process is streamlined by tools like ARC2Repo, which convert the metadata stored within an ARC into the required formats for repository deposit. By integrating these tools, DataPLANT ensures that plant science data is uniformly annotated and easily shared.

Corresponding Author: Angela Kranz, a.kranz@fz-juelich.de

ID P28

The Behavioral Standard Metadata (BeStMeta): Metadata Standard for Video Tracking Assays (VTAs)

Authors: Steffi Bredemeyer[1], Madhu Nagathihalli Kantharaju[2]

Co-Authors: Deborah Schmidt[2], Ivan Ezquerra-Romano[2], Riccardo Massei[1]

[1] UFZ - Helmholtz Centre for Environmental Research, [2] Max Delbrück Center for Molecular Medicine

Recent advances in imaging and computational methods have led to a rapid expansion of video-based tracking in the biological sciences. The increased adoption of video tracking assays (VTAs) has created new opportunities and challenges for quantitative analysis across multiple scales of biological systems. However, there is a lack of coherent community standards and a metadata reporting schema which are necessary for reproducibility, interoperability and consistency in experimental findings.

The project Behavioral Standard Metadata (BeStMeta) aims to build a generalized reporting schema for VTAs. In our vision, the schema needs to be adaptable to different experimental setups, organisms (e.g., zebrafish, rodents) and analysis pipelines ensuring flexibility for different VTA setups. All metadata needs to be reported in machine-readable formats in order to support interoperability and further processing.

In this work, we present initial results of a comprehensive survey of existing metadata standards on published studies in ecotoxicology and neuroscience using VTAs. Relevant publications were identified through a structured literature search using the open-source PubMed and PMC APIs. We then combined manual extraction with an automated pipeline based on regular expressions and large language models (LLM). These approaches were employed to systematically survey the papers and to create a comprehensive overview of all the metadata that is reported consistently or inconsistently across behavioral studies. Preliminary results on 84 publications indicate major gaps in metadata and data reporting: only 18 studies shared video files in public repositories and less than half reported technical metadata such as camera model, frame rate, or resolution highlighting systemic shortcomings in current practices. In summary, this survey highlights the necessity of establishing a unified metadata reporting schema for VTAs.

Corresponding Authors: Madhu Nagathihalli Kantharaju, madhu.nagathihallikantharaju@mdc-berlin.de, Riccardo Massei, riccardo.massei@ufz.de

Track 9

From Minimum Requirements to FAIR and AI-Ready: Assessing Metadata Quality

Chair: Volker Hofmann[1]

[1] HMC Hub Information & HMC FAIR Data Commons - Forschungszentrum Jülich

Assessing metadata quality can follow many paths — from institutional policies and FAIR principles to domain-specific needs like AI-readiness. Yet, current formalised assessment approaches remain limited in producing objective, actionable results.

This session invites researchers and infrastructure developers to share and discuss methods for evaluating metadata quality, with a focus on strengthening assessment frameworks that can effectively guide implementation towards interoperability, and data reuse.

ID D15

Beyond Static Checklists: A Dynamic, Modular Toolkit for Automated Metadata Assessment

Author: Hamed Hemati[1]

Co-Authors: Alicia Janz[1], Stefan Sandfeld[1]

[1] Forschungszentrum Jülich

As the volume of research output accelerates, the rigorous validation of scientific publications, and the assessment of their claims, objectives, and reproducibility, have become critical bottlenecks. Traditional manual verification is labour-intensive and unscalable, often failing to keep pace with the growing need for structured assessment. To address this, we introduce a domain-agnostic, web-based toolkit designed to automate the assessment of scientific publications and transform the quality control workflow from a static box-ticking exercise into a dynamic metadata verification process.

Moving beyond fixed checklist structures, this toolkit allows users to design custom checklists and review workflows through a visual, no-code designer. This enables domain experts to define specific metadata requirements (e.g., FAIR principles, reproducibility standards, or novel contribution tagging) without programming expertise. For advanced use cases, the toolkit offers a modular architecture that supports custom Python-based extensions, allowing research labs to integrate their own external analysis tools (e.g., for code repository or dataset analysis) directly into the review process definition. The system leverages these workflows alongside flagship Large Language Models (LLMs), such as GPT, Gemini, and open-weight alternatives, to automatically extract and validate information against user-defined schemas.

We validate this toolkit using checklists in the field of Machine Learning as a primary case study for extracting provenance metadata. Furthermore, we illustrate how the tool facilitates collaborative verification by allowing users to assess AI outputs via an integrated human review desk to ensure reliability.

Corresponding Author: Alicia Janz, a.janz@fz-juelich.de

ID P29

Assessing Metadata Quality Across Helmholtz Data Providers: A Practical Approach for Harmonization

Authors: Andreas Pfeil[1], Anis Koubaa[1], Christine Lemster[2], Fabia Martens[1], Hamideh Haghiri[3], Markus Kubin[4], Oonagh Brendike-Mannix[4], Santiago Casas[5], Lucas Lamparter[6]

[1] Karlsruhe Institute of Technology, [2] GEOMAR Helmholtz Centre for Ocean Research Kiel,
[4] German Cancer Research Center, [5] Helmholtz-Zentrum Berlin,
[6] Deutsches Zentrum für Luft- und Raumfahrt (DLR), [6] Forschungszentrum Jülich

As staff of the Helmholtz Metadata Collaboration (HMC), we used metadata harvested by HMC services to conduct a cross-provider analysis of how metadata is currently exposed across Helmholtz and to identify systematic issues that affect metadata quality.

The analysis focused on three priority fields that strongly influence discoverability and integration: identifiers, publication date, and resource type. For each field, we defined simple, reproducible assessment criteria to classify patterns, detect inconsistencies, and highlight common deviations from expected practices. Here we present initial results and the patterns observed across providers.

This work aims to make existing metadata practices visible and to facilitate community discussion around them. The results inform consultations with providers, contribute to understanding how heterogeneous metadata practices impact the Helmholtz Knowledge Graph and related services, and thereby support HMC's efforts to harmonize metadata across Helmholtz.

Corresponding Author: Oonagh Brendike-Mannix, oonagh.mannix@helmholtz-berlin.de

ID P30

Automated FAIR Evaluation with FAIR-Eva and Its Application to the Coscine Research Data Management Platform

Authors: David Schimmel[1], Fernando Aguilar Gõmez[2]

Co-Authors: Ilona Lang[3], Marius Politze[1], Pablo Orviz Fernãndez[2]

[1] RWTH Aachen University, [2] Institute of Physics of Cantabria (CSIC-UC)

The concept of FAIR Digital Objects (FDOs) supports good research data management practices. However, the evaluation of compliance with the FAIR principles suffers from uncertainties: the FDO specification allows different implementations. Furthermore, heterogeneous, domain-specific terminologies and ontologies are used in concrete FDOs. Access methods for data and metadata differ among FDO implementations. The FAIR principles are open to different interpretations.

Consequently, automatic FAIR evaluation tools often focus on well-established terminologies and FDO providers. In contrast, FIDELIS [1] takes a bottom-up approach with FAIR-Eva [2]. FAIR-Eva is a framework for automatically finding and evaluating metadata properties according to the RDA FAIR maturity indicators. It can guide FAIRification efforts by providing useful feedback to creators of metadata records. A key feature of FAIR-Eva is its interface for defining custom plugins, enabling developers to adapt the tool to their specific metadata schemas.

During a first support round, Coscine [3] was chosen as a service for which plugin development was guided. Coscine is a multi-disciplinary data and metadata management platform. In Coscine, data stewards can register metadata schemas tailored to their domain-specific requirements. Thus, a FAIR evaluation of metadata records cannot completely rely on schema standardization.

Furthermore, the structure of Coscine FDOs is not well-supported by existing FAIR evaluation tools. FDOs are identified by handle-based ePIC PIDs. Each PID is associated with a Coscine-specific Kernel Information Profile, linking to the location of the metadata record. This record describes a complex digital object, comprised of a hierarchy of RDF and non-RDF sources, conforming to the Linked Data Platform. They can be navigated by extensions of the object PID. Automatically following the links and evaluating the correct attributes is a challenge.

We propose presenting FAIR-Eva as well as the challenges and insights from the use case of developing a plugin for Coscine.

References:

- [1] https://eosc.eu/horizon-europe-projects/fidelis
- [2] https://github.com/EOSC-synergy/FAIR_eva/blob/main/docs/index.md
- [3] https://about.coscine.de

Corresponding Author: David Schimmel, schimmel@itc.rwth-aachen.de

ID P31

Flexible Metadata Schema Generation for Heterogeneous Physics Experiments Using Object-Oriented Generalization

Author: Saeid Masoumi[1]

Co-Authors: Anke-Susanne Müller[1], Julian Gethmann[1], Marcel Schuh[1], Robert Ruprecht[1], Wolfgang Mexner[1]

[1] Karlsruhe Institute of Technology

Heterogeneous experimental data in modern physics research facilities, such as KARA and FLUTE at KIT, often rely on experiment-specific metadata, making it challenging to define a unified schema for Research Data Management (RDM). We propose an object-oriented approach to address this challenge.

In our framework, each experiment independently generates a dedicated class encapsulating its unique metadata. These classes are then collected and organized into a hierarchical tree structure using the concept of generalization, a core principle in object-oriented design. Shared properties across experiments are promoted to parent classes, while experiment-specific details remain in leaf classes.

This method allows the automated generation of flexible and coherent metadata schemas for a wide variety of experiments, enabling interoperable, extensible, and maintainable RDM solutions. We demonstrate this approach using selected experiments at KIT, showing how class-based generalization produces a scalable and unified metadata representation, bridging the gap between heterogeneous experimental data and FAIR data principles.

Corresponding Author: Saeid Masoumi, saeid.masoumi@kit.edu

ID P32

Interdisciplinary FAIR and Open Data in Helmholtz: New Features in the HMC FAIR Data Dashboard (v3)

Authors: Gabriel Preuß[1], Markus Kubin[1], Mojeeb Rahman Sedeqi[1], Oonagh Brendike-Mannix[1], Pascal Ehlers[2]

[1] Helmholtz-Zentrum Berlin, [2] Deutsches Zentrum für Luft- und Raumfahrt (DLR)

The HMC Dashboard on Open and FAIR Data (also known as the HMC FAIR Data Dashboard) has reached version 3, marking a significant milestone in monitoring and assessing FAIR data practices across disciplines. This release integrates key updates to the HMC Toolbox for Data Mining and incorporates fixes related to recent changes in the Scholexplorer API as well as the FAIRsFAIR metrics and F-UJI API, ensuring continued reliability and consistency of these evaluations. For the first time, the dashboard now includes data publications retrieved directly from library interfaces that are not yet linked to literature publications, broadening its scope and completeness and opening up for future integrations in that direction. Version 3 also enables the identification and visualization of interdisciplinary data publications contributed by multiple Helmholtz centers and research fields, offering a more comprehensive view of data sharing activities within the Helmholtz research community.

Corresponding Author: Gabriel Preuß, gabriel.preuss@helmholtz-berlin.de

Track 10

Harmonisation of Metadata: Closing Semantic Gaps

Chair: Gerrit Günther[1]

[1] HMC Hub Matter - Helmholtz-Zentrum Berlin

The session will focus on the development of metadata standards that balance the specific requirements of a community and interoperability with advancing technologies and existing standards. We invite contributions emphasising on the semantic level, structuring of metadata, technical implementation to local and global infrastructures, up to social dimensions of standardisation, including consensus-building within diverse stakeholder groups. We encourage participants to share their experience, insights, and best practice on this topic.

ID D16

From Hidden Reports to Actionable Data. A Domain-Specific Metadata Standard for Conservation Science

Authors: Kristina Fischer[1], Lasse Mempel-Länger[1], Nathaly Witt[1]

[1] Leibniz Centre for Archaeology (LEIZA)

The conservation of cultural heritage combines humanities and natural science approaches with practical work on objects. Conservation activities provide unique insights into materiality, manufacturing traces, as well as damage mechanisms and modifications of an object. Despite their scientific relevance for object-based research, these data are often hidden in narrative reports, neither machine-readable nor available according to FAIR principles.

This contribution presents a domain-specific metadata standard developed within a Temporary Working Group (TWG) in the interdisciplinary NFDI4Objects consortium. The standard builds on existing documentation traditions while promoting interoperability with established RDM-standards in heritage research. To ensure a shared understanding, the metadata elements are structured as a SKOS-based thesaurus and made interactively accessible through a mock-up page. During the development stage, this practical implementation facilitates comprehension and communication of the rather abstract conceptual work and supports feedback from the conservation community.

The standard establishes a foundation for making conservation data available as structured research data, bridging gaps between object documentation and process documentation. Through contextualisation with data from related disciplines such as archaeology, art history, materials science, and engineering, new research questions and analytical possibilities arise. How do patterns of damage in historical buildings relate to geographical location and environmental conditions? Which historical or contemporary materials become unstable under which conditions? Standardised conservation data enable systematic analyses of object changes, enhance traceability of material interventions, and provide insights into conservation practices and their development over time.

Looking ahead, building on the developed metadata standard, the further development of a CIDOC CRM-based application ontology is planned to digitally represent the complex interrelations and decision-making processes in conservation practice. This approach enhances the long-term value of cultural heritage data, making conservation science a visible and accessible contributor to the wider digital humanities ecosystem and beyond.

Corresponding Author: Kristina Fischer, kristina.fischer@leiza.de

ID D17

The new Circum-Arctic Soil Permafrost Region database (CASPeR)

Author: Friedrich Röseler[1]

Co-Authors: Guido Grosse[1], Gustaf Hugelius[2]

[1] Alfred Wegener Institute for Polar and Marine Research, [2] Stockholm University

Arctic permafrost soil research is significantly limited by the scarcity of reference data. While increasingly more reference data has been published in recent works, it remains scattered among publications, data repositories and institutional archives. Often, the provided data is highly unstructured, important metadata is missing, and previously synthesized databases are becoming outdated and are not detailed enough for machine learning applications. This has led to a significant lack of reliable reference data for permafrost soil research and is particularly relevant in global modelling initiatives, like the Global Carbon Project (GCP).

We developed a new harmonized, standardized, structured, extensive, analysis ready and FAIR soil profile reference database for the northern circumpolar permafrost region, called the Circum-Arctic Soil Permafrost Region database (CASPeR). CASPeR was developed with an emphasis on making it accessible to both the Arctic domain and modelling research community.

CASPeR's extensive metadata gives detailed information on the database attributes, stored using JSON schema and metadata files. This solution offers more flexibility and structure than other commonly used formats (e.g., Excel, text files), but does not require designing complex ontology models. We still make use of such models, reuse existing standards and definitions where possible, and added own definitions if needed. This was and remains a challenging task, as there is not always agreement within the Arctic research community on definitions and semantics, and a lack of structured metadata standards.

Once published, CASPeR will be the largest, most detailed structured soil carbon database for the Arctic permafrost region. We are committed to deliver a high-quality and metadata-rich soil reference database for the Arctic research community. It is planned to distribute CASPeR v1.0 under an Open Source license, together with an interactive dashboard, a Python API for easier access, and a data paper giving an overview on the harmonization methods.

Corresponding Author: Friedrich Röseler, friedrich.roeseler@awi.de

ID P33

A Community-Driven Semantic Strategy for Interoperable Earth and Environmental Data

Author: Dorothee Kottmeier[1]

Co-Authors: Andrea Pörsch[2], Johann Wurz[3], Stanislav Malinovschii[4], Emanuel Söding[4]

[1] Alfred Wegener Institute for Polar and Marine Research, [2] GFZ Helmholtz Centre for Geosciences, [3] UFZ - Helmholtz Centre for Environmental Research, [4] GEOMAR Helmholtz Centre for Ocean Research Kiel

Embedding semantics in research metadata is essential for improving interoperability across datasets and for implementing the FAIR principles. The HMC Hub Earth and Environment is developing a coordinated, community-driven strategy to support the consistent and sustainable use of semantics in Earth and Environment research and to contribute to cross-domain semantic harmonization within the Helmholtz Association.

Central to this effort is the working group AK Metadaten-Semantik, now integrated into the DataHub, which provides a collaborative framework for aligning semantic practices across institutions. The group addresses key semantic dimensions such as observable properties, measurement instruments, and methodologies, and serves as a platform to identify gaps, overlaps, and shared needs across communities and research areas. As a concrete reference implementation, the group is developing a prototype community vocabulary for device types in Earth and Environment research.

The resulting semantic strategy focuses on four core aspects: guidance on navigating and evaluating existing semantic resources; identification of essential metadata elements for semantic annotation; assessment of tools required for practical semantic implementation, including AI-based approaches; and moderation of community processes for agreeing on shared vocabularies and governance structures.

Recommendations and discussion outcomes are published transparently via the HMC Earth and Environment Wiki. Overall, the strategy aims to establish a common semantic foundation, practical guidelines and supportive tools to enable sustainable, interoperable data infrastructures in the field of Earth and the environment and beyond.

Corresponding Author: Dorothee Kottmeier, dorothee.kottmeier@pangaea.de

ID P34

Bridging Biomedical Data Through Harmonised Metadata within NFDI BioMed Interest Group

Authors: Ulrik Stervbo[1], Sebastian Böhm[2], Martin Golebiewski[3], Matthias Löbe[4], Josh Moore[5], Carina Nina Vorisek[6], Christian Busse[7]

[1] NFDI4Imuno, Ruhr University Bochum, [2] NFDI4Immuno, Friedrich-Loeffler-Institut Federal Research Institute for Animal Health, [3] NFDI4Health, Heidelberg Institute for Theoretical Studies, [4] NFDI4Health, Leipzig University, [5] NFDI4BIOIMAGE, German BioImaging - Society for Microscopy and Image Analysis, [6] NFDI4Health, Berlin Institute of Health [7] NFDI4Immuno, German Cancer Research Center

The German National Research Data Infrastructure (NFDI) BioMed Interest Group unites five NFDI consortia - GHGA, NFDI4BIOIMAGE, NFDI4Health, NFDI4Immuno, and NFDI4Microbiota - to address the distinct technical, regulatory and ethical challenges of research data management (RDM) in biomedical science. These consortia represent diverse research communities, each with their own approach to defining knowledge and data organisation. This results in heterogeneity across metadata schemas, terminologies, perspectives, stakeholder groups and even legal frameworks, making comparison of data and interfacing between services difficult. Consequently, systematic mappings between metadata schemas is difficult, but such interoperability is essential to establish an ecosystem of linked resources.

To address this, the group held workshops in 2024 (DKFZ, Heidelberg) and 2025 (ZB Med, Cologne). The first established mutual understanding of current practices through deep-dive sessions on data models, infrastructure, and regulations; the second developed a metadata interoperability roadmap (Figure 1). This roadmap focuses on discovery-level metadata describing study characteristics rather than enforcing deep semantic alignment of all raw variables. A key opportunity lies in the European Health Data Space (EHDS) requirement for the HealthDCAT-AP specification. The group is developing a consortia-overarching metadata layer that aligns with NFDI Task-Force Metadata recommendations while incorporating HealthDCAT-AP and Health Level 7 Fast Healthcare Interoperability Resources (HL7 FHIR). This approach enables both interoperability across data infrastructures through harmonised general metadata and metadata flexibility for the different communities.

This strategy enables interoperability across infrastructures through harmonised discovery metadata while allowing individual communities to maintain the flexibility of their domain-specific schemas. By implementing this overarching layer, it provides a pathway to preserve disciplinary diversity while facilitating cross-repository data discovery and meta-analysis. Furthermore, this framework provides a viable model for a common data catalogue description suitable for the European Open Science Cloud (EOSC) without losing the granular detail required by individual research communities.

Corresponding Author: Ulrik Stervbo, ulrik.stervbo@elisabethgruppe.de

ID P35

Developing FAIR and Interoperable Metadata Standards for 3D Reflection Seismic and Ocean Bottom Seismometer Data

Author: Janine Berndt[1]

Co-Authors: Christian Berndt[1], Daniel Damaske[2], Janine Felden[3], Hela Mehrtens[1], Marianne Rehage[2], Niklas Selke[2], Mechita Schmidt-Aursch[3], Mehrdad Soleimani Monfared[1], Estella Weigelt[3]

[1] GEOMAR Helmholtz Centre for Ocean Research Kiel, [2] MARUM –Center for Marine Environmental Sciences at the University of Bremen, [3] Alfred Wegener Institute for Polar and Marine Research

Because of traditionally large data volumes and the diversity of acquisition and processing workflows, internationally recognized metadata standards for marine seismic data are still lacking. At the same time, there is a growing need within the geophysical community to harmonise metadata in accordance with the FAIR principles, enabling semantic interoperability, long-term reuse, and emerging applications such as machine learning and large-scale data analysis. Within the Helmholtz Metadata Collaboration (HMC) Project MetaSeis, we addressed metadata standardisation for two specific types of marine seismic data: ocean bottom seismometer (OBS) data acquired using artificial seismic sources (active-source experiments) and 3D reflection seismic data. While active-source OBS data have been archived for many years, no commonly agreed metadata standard exists. We therefore reviewed and consolidated existing practices at several international institutions, with a focus on clarifying acquisition, processing, and instrument-related metadata. Based on these findings and aligned with existing efforts for passive-source OBS data, we developed a first draft metadata standard that is currently under discussion in the community and open to further refinement. In parallel, we established an archiving workflow for legacy active-source OBS data acquired by the Alfred Wegener Institute (AWI). Metadata retrieval for these datasets proved challenging due to incomplete or missing metadata, reports, and documentation. These gaps required retrospective reconstruction of metadata in close collaboration with domain experts, highlighting practical challenges in metadata harmonisation. In addition, existing metadata standards for raw 2D reflection seismic data were extended to include 3D reflection seismic data. This step was facilitated by the limited number of institutions in Germany and their prior agreement on the 2D metadata structure.

Corresponding Author: Janine Berndt, jberndt@geomar.de

ID P36

Enhancing Interoperability and Reusability in NFDI4Immuno: Implementation of FAIR Principles

Authors: Ulrik Stervbo[1], Sebastian Böhm[2], Anca Dohoi[2], Katherina Siewert[3], Christian Busse[4]

[1] Ruhr University Bochum, [2] Friedrich-Loeffler-Institut Federal Research Institute for Animal Health, [3] German Federal Institute for Risk Assessment, [4] German Cancer Research Center

NFDI4Immuno develops a federated repository infrastructure for immunological data that implements FAIR principles across diverse data types such as single cell sequencing, including AIRR-seq, cytometry, immunopeptidomics, microscopy, and immune receptor reactivity data. The consortium addresses the challenge of semantic interoperability by building metadata models that balance domain-specific requirements with cross-consortium compatibility.

Our technical approach builds on the core class structure of the Ontology for Biomedical Investigations (OBI). The schema implements a study/subject/sample hierarchy where studies follow defined plans, subjects participate in studies, and samples represent realised time-points from the study plan. We model data type specific metadata from MiAIRR and MiFlowCyt standards within this structure, capturing the sample-to-data transformation process and unifying elements where possible across assay types. As data from human subjects is an essential part of immunology, the metadata model is built with GDPR in mind. The regulation guides the choice of metadata that can be collected and made searchable. The schema is currently in active development and undergoes continuous testing with internal datasets.

At its core, NFDI4Immuno works towards seamless interoperability with other life sciences NFDI consortia, particularly GHGA, NFDI4Health, NFDI4Microbiota and NFDI4BIOIMAGE. Additional efforts focus on mapping to and integrating with existing external resources such as the AIRR Data Commons, creating harmonised data access and standardised data representations. Ongoing work on DCAT-AP and FHIR/HealthDCAT-AP compliance enables broader data discovery - for instance though integration into the European Open Science Cloud (EOSC) and the European Health Data Space (EHDS), respectively. This interoperability strategy fosters cross-domain data discovery and reuse, facilitating novel insights at the intersection of immunology and related disciplines.

This work provides immunology researchers with standardised metadata structures that support data deposition, discovery, and reuse while maintaining interoperability across life sciences domains.

Corresponding Author: Ulrik Stervbo, ulrik.stervbo@elisabethgruppe.de

ID P37

Expressing Conditional Validity of Statements

Authors: Felix Ballani[1], K. Gerald van den Boogaart[1]

[1] Helmholtz Center Dresden-Rossendorf

Domain semantics typically allow to describe things very precisely and with a precise meaning. In many cases, however, not everything that can be said is either true or false, but often true or known only in a limited domain of validity: a certain time span (e.g. the temperature of an object), a spatial domain (e.g. a law), a certain simplification of a theory (e.g. a proportionality coefficient), or even with a certain claim (which still might be unvetted or even wrong). Without general ways of expressing such conditional validity, a multitude of practical challenges arise: Triplets are generated to express facts that are valid only in a specific context and therefore can no longer be used meaningfully together with other triplets from other contexts in automated reasoning, as they would lead to false conclusions. Also, it often requires overcomplications of concepts, like the concept of identity, when a process description needs different identifiers for the same object at different times. While some domain ontologies have tools to express such cases, there seems to be no general solution. Natural languages, however, typically have general grammatical constructs like "According to A. et al. (2024) the Pinatubo was active from April 2, 1991, to June 15, 22:30, ..." to express the conditionality of claims.

The poster exemplifies these challenges and discusses possible general solutions that mimic these natural language capabilities while still enabling complete formal reasoning. The idea is based on the use of identifiers that represent claims, which can then be declared valid under certain conditions. The proposed approach allows to extend deduction rules of the domain ontologies to deduction rules for such conditional claims.

Corresponding Author: Felix Ballani, f.ballani@hzdr.de

ID P38

From Cultural Heritage to Interoperable Data Sets: Metadata Standards as a Basis for Digital Object Biographies

Authors: Kristina Fischer[1], Anja Gerber[2]

[1] LEIZA - Leibniz-Zentrum für Archäologie, [2] Weimar Classicism Foundation

In object-based research within the humanities, such as archaeology, art history, provenance research or conservation science, metadata describe not only digital files but also help to structure information about real-world cultural heritage. The term "metadata" here encompasses systematised object information on object designation, creation history, provenance and usage. Thus, "Creator" refers to the maker of the actual physical object, not the creator of the dataset. Metadata elements such as "object material" describe the relationship between the chemical element "gold" and the documented artefact.

These various pieces of information together trace the object biography [1], i.e. the trajectory of the object from its creation or discovery to the present day. At each stage, from archaeological excavation through collection acquisition, conservation treatment and material-scientific analysis to further research activities, new information emerges. If this is not carefully defined or is lost, gaps in the knowledge about an object arise that are difficult to close.

Furthermore, the heterogeneity of metadata poses a central challenge. Different institutions and disciplines collect different data according to different standards. The consortium NFDI4Objects (N4O) [2] addresses this problem through the development of collaborative metadata standards. The N4O Object Core Metadata Profile [3] establishes a cross-disciplinary minimum metadata set with consensually defined data fields and controlled terminologies to secure data exchange within the wider NFDI. Building on this, for more specialised subject areas more specific standards are being developed, such as the metadata standard for conservation documentation (KuR-MDS) [4]. Through event-based data models like CIDOC CRM [5], these metadata can be linked, thereby enabling the digital representation of object biographies [6] across institutions and research disciplines.

The contribution at HMC 2026 presents experiences from N4O on the development and linking of metadata standards and demonstrates how semantic interoperability sustainably strengthens interdisciplinary research.

References:

- [1] https://zenodo.org/records/18221449
- [2] https://www.nfdi4objects.net
- [3] https://doi.org/10.5281/zenodo.17159183
- [4] https://doi.org/10.5281/zenodo.17367214
- [5] https://cidoc-crm.org
- [6] https://nfdi4objects.wisski.data.fau.de

Corresponding Author: Kristina Fischer, kristina.fischer@leiza.de

ID P39

Let's Improve the Display of Research Results in the INSPIRE Environment!

Author: Andrea Pörsch[1]

Co-Authors: Dorothee Kottmeier[2], Emanuel Söding[3], Johann Wurz[4]

[1] GFZ Helmholtz Centre for Geosciences, [2] Alfred Wegener Institute for Polar and Marine Research, [3] GEOMAR Helmholtz Centre for Ocean Research Kiel, [4] UFZ - Helmholtz Centre for Environmental Research

The improvement of data exchange between research communities and public authorities in Germany is frequently discussed. Both the research data infrastructure in the Earth and environmental sciences and the German Geospatial Data Infrastructure (GDI-DE) are highly developed systems. However, the respective communities tend to operate largely within their own domains. This separation is not only due to established expertise within each community, but also to the use of different metadata standards and interfaces. As a result, the associated data portals emphasize different aspects of information.

Within GDI-DE, spatial (geo-)data are described in great technical detail, whereas scientific communities place particular importance on the proper citation of authors and organizations. To increase mutual acceptance and to foster data exchange between both communities, existing interfaces should be adapted accordingly. In particular, improving the comprehensive display of citation-related information in GDI-DE portals would enhance their acceptance within the research community.

However, the ISO metadata standard currently used by GDI-DE does not provide metadata elements or designated structures in the UML model for persistent identifiers (PIDs) for persons and organizations. To address this gap, a new chapter is currently being introduced into the "Konventionen der GDI-DE", incorporating requirements from a scientific perspective.

This poster presents details and practical examples of the proposed approach and advocates for the implementation of the new convention, enabling the display of persistent identifiers for researchers from ORCID and for organizations from the ROR registry within GDI-DE and INSPIRE portals.

Corresponding Author: Andrea Pörsch, apoersch@gfz.de

ID P40

Project-Oriented Knowledge Graphs for Metadata Alignment Across Repositories

Authors: Stanislav Malinovschii[1], Andrea Pörsch[2], Dorothee Kottmeier[3], Emanuel Söding[1], Johann Wurz[4], Sören Lorenz[1]

[1] GEOMAR Helmholtz Centre for Ocean Research Kiel, [2] GFZ Helmholtz Centre for Geosciences, [3] Alfred Wegener Institute for Polar and Marine Research, [4] UFZ - Helmholtz Centre for Environmental Research

Knowledge graphs enable the structuring and linking of scientific metadata from heterogeneous sources, supporting advanced search and analysis. Their effectiveness, however, depends on the level of metadata consistency and standardization. Graphs linking high-quality topical information can also be very useful for training tailored LLMs and AI models for specific purposes. With this activity, we aim to identify and extract topical datasets from repositories, improve and harmonize discipline-specific information in the metadata, turn them into graphs, and make them available as training data for AI applications.

For this purpose, we explore an approach to building knowledge graphs from project metadata in scientific repositories. We use projects as a linking layer to integrate and align datasets, as well as their associated methods, instruments, keywords, and geographic descriptions. Metadata is extracted and normalized from structured fields and textual descriptions, and then subjected to semantic categorization.

The poster presents the methods applied, encountered challenges, and results, including statistics on harvested metadata, category coverage, and observed differences in metadata representation across repositories. We also discuss the potential of project-oriented knowledge graphs for improving the quality and interoperability of scientific metadata.

Corresponding Author: Stanislav Malinovschii, smalinovschii@geomar.de

Thursday, 30.04.2026

Lecture Hall, 09:00-10:00

Keynote

TBA

Prof. Dr. Oliver Stegle[1]

[1] German Cancer Research Center (DKFZ) and Director GHGA

In his keynote, Prof. Stegle will bring the perspective of scientific infrastructure to the theme of "Metadata in Action," reflecting on how robust infrastructure — from secure data archives to scalable computational ecosystems — underpins effective metadata practices and enables cutting-edge science. Drawing on his experience building and coordinating GHGA and related initiatives that integrate data workflows, governance, and interoperability across institutions, he will discuss how infrastructures can make metadata more actionable, drive reproducible research, and support interdisciplinary collaboration in the era of large-scale genomics and biomedicine.

Lecture Hall, 10:30 - 12:15

Talk Session:

Track 4

Human-Machine Collaboration in (Meta)data Acquisition

Chair: Marta Dembska[1]

[1] HMC Project LabFriend - Deutsches Zentrum für Luft- und Raumfahrt (DLR)

Scientific progress increasingly depends on effective collaboration between humans and machines. Central to this is the digital, standardised acquisition and management of (meta)data in laboratories and fieldwork alike.

This session explores how digital tools (ELNs, LIMS), formalised workflows using ontologies and controlled vocabularies, and automation through robotics or AI agents can enhance (meta)data capture. Emphasis will be placed on metadata standardisation, intuitive data entry, and user support to ensure interoperability and reuse across research domains.

ID T19

A Semantically Integrated Framework for Robotic Data Acquisition in Mechanical Testing

Author: Eric Breitbarth[1]

Co-Authors: David Melching, Ferdinand Dömling, Florian Paysan, Guillermo Requena, Jesco Talies

[1] Deutsches Zentrum für Luft- und Raumfahrt (DLR)

Traditional mechanical testing often relies on manual observation and fragmented data storage, creating bottlenecks in scientific progress. To reduce development times and make mechanical testing more sustainable, we must transition from manual logging to high-throughput, standardized data acquisition. This presentation demonstrates a paradigm shift in human-machine collaboration within fatigue crack growth experiments.

We present an autonomous data acquisition framework for fatigue crack growth experiments in which intelligent robotics, Digital Image Correlation (DIC), and machine learning (ML) operate as closed-loop sensing agents. High-resolution DIC continuously tracks crack tip position and deformation fields, while ML models extract higher-order descriptors, including plastic zone evolution and fracture-relevant damage features, in real time.

Central to the framework is a semantic orchestration layer based on graph databases, domain ontologies, and explicit provenance models. Experimental parameters, sensor states, derived features, and processing steps are represented as first-class entities in a unified knowledge graph. This enables automated metadata capture, cross-modal data alignment, and machine-driven reasoning over experimental context, rather than post-hoc annotation.

By decoupling experimental execution from semantic interpretation, the framework transforms mechanical testing into a self-describing, machine-navigable process. The result is an autonomous experimental pipeline that supports scalable data generation, reproducible analysis, and seamless integration into self-driving laboratory workflows.

Corresponding Author: Eric Breitbarth, eric.breitbarth@dlr.de

ID T20

Automated Metadata Acquisition in Energy Research using BPMN-driven Workflows at the Energy Lab at KIT

Author Jan Martin Reckel[1]

Co-Authors: Malte Holzhäuer[1], Peter Moster[1], Simon Waczowicz[1], Tobias Moser[1], Veit Hagenmeyer[1]

[1] Karlsruhe Institute for Technology

In energy research facilities, such as the Energy Lab at the Karlsruhe Institute of Technology, the systematic acquisition of high-quality metadata remains a significant challenge because manual documentation can be error-prone and time-consuming. To ensure data findability and reproducibility according to the FAIR principles, we present an innovative approach that utilizes Business Process Model and Notation (BPMN) to orchestrate research workflows and automate metadata capture. By using Operaton as a lightweight, open-source BPMN engine, metadata acquisition can be transformed into an inherent component of the experimental lifecycle. Within the Energy Lab infrastructure, experimental sequences, ranging from sensor calibration to data storage, are modelled as executable BPMN diagrams. These models serve a dual purpose: they provide a visual documentation layer for researchers and act as technical instructions for the Operaton engine. By integrating specialized metadata tasks directly into the automated workflow, the engine extracts technical parameters and provenance information in real-time and maps them to standardized schemas without requiring manual intervention. The initial results demonstrate that this orchestration significantly increases metadata completeness and consistency while reducing the administrative burden on researchers. Furthermore, the graphical nature of BPMN facilitates a crucial bridge between domain-specific research and data engineering. This integration provides a scalable framework for "Metadata-by-Design," ensuring that complex datasets generated within the Helmholtz Association are accompanied by high-quality, machine-readable documentation. Ultimately, the use of Operaton for process-driven metadata acquisition represents a robust solution for the long-term usability of energy-research data.

Corresponding Author: Jan Martin Reckel, jan.reckel@kit.edu

ID T21

A Semantic Laboratory Assistant for Metadata Acquisition in Electronic Lab Notebooks

Authors: Tina Boroukhian[1], Marta Dembska[2], Sirko Schindler[2], Martin Held[1], Catriona Eschke[1], Fabian Kirchner[1], Eric Breitbarth[2], Prakul Pandit[2]

[1] Helmholtz-Zentrum Hereon, [2] Deutsches Zentrum für Luft- und Raumfahrt (DLR)

Laboratory data reuse and reproducibility depend on rapid, accurate, and complete capture of experimental (meta)data. In practice, metadata creation in electronic lab notebooks (ELNs) remains a bottleneck because form-based entry interrupts workflows, free-text input is time-consuming and error-prone, and heterogeneous terminology complicates harmonisation across projects and infrastructures. These limitations reduce metadata quality and quantity and impede Findable, Accessible, Interoperable, and Reusable (FAIR) dissemination and integration.

LabFriend is an open, ELN-agnostic laboratory assistant under development to mitigate these issues through semantically structured, context-aware metadata acquisition. Intended functionality includes real-time suggestions, validation of field values against controlled semantics, and optional speech-based capture. Methods combine association-rule mining from historical form instances with ontology- and knowledge-graph-based semantic relatedness, aiming to improve completeness and terminology consistency while keeping interaction lightweight.

A central prerequisite is robust data preparation that converts heterogeneous ELN exports into validation-ready material for semantic methods and evaluation. This contribution focuses on a preparation workflow for transforming records collected in the Chemotion ELN into knowledge-graph-ready representations. The workflow addresses common obstacles in exported records, including a mixture of structured key-value fields and unstructured free-text, missing or implicit units, inconsistent naming, ambiguous identifiers. Preparation steps include structure extraction from Chemotion objects, normalisation of datatypes and units, entity resolution across samples, processes, and instruments, and semantic anchoring to domain vocabularies while preserving provenance of each transformation decision. The resulting material is annotated manually against a closed, schema-driven target model and can be mapped to Resource Description Framework (RDF) statements for knowledge-graph construction and downstream reuse.

Corresponding Author: Tina Boroukhian, tina.boroukhian@hereon.de

ID T22

A Collaborative Approach to Metadata Interoperability: PID4NFDI, TS4NFDI, and RSpace

Authors: Sara El-Gebali[1], Tilo Mathes[2], Rory Macneil[2], Roman Baum[3]

[1] DataCite & PID4NFDI, [2] Research Space, [3] ZB MED - Information Centre for Life Science

This contribution presents a joint collaboration between PID4NFDI, TS4NFDI, and the electronic lab notebook provider ResearchSpace to support interoperable research workflows within the National Research Data Infrastructure (NFDI). It demonstrates how early, structured capture of high-quality metadata and persistent identifiers (PIDs) in ELNs, combined with shared reference schemas and centrally governed terminology services, can reduce redundant effort and improve metadata consistency and data lineage across the research lifecycle [1].

The presentation outlines the complementary roles of the partners. PID4NFDI coordinates PID integration and metadata alignment. TS4NFDI provides centralized terminology services via an API Gateway, ensuring consistent, machine-actionable metadata. RSpace integrates these components into everyday research workflows, enabling structured metadata and PID capture at the point of data creation.

Entity mappings curate DataCite schema alignments with schema.org and DCAT, maintained in Cocoda [2] with versioning and provenance tracking. DataCite properties and vocabularies are available via TIB Terminology Service, providing canonical, machine-actionable terms accessible through TSS widgets [3] and the API Gateway [4]. RSpace integrates these services via embedded widgets, enabling structured metadata capture at data creation and supporting export of NFDI-compliant ELN records.

Overall, the collaboration establishes a reusable, machine-actionable metadata layer based on shared terminology lookup, cross-schema mappings as a single source of truth, and clear service integration patterns. The proof of concept illustrates how PID4NFDI and TS4NFDI can work with ELN and DMP providers to enable interoperable research workflows and inform future NFDI-wide implementations.

References:

- [1] El-Gebali, S. (2024). Concepts for metadata interoperability, harmonization and technical integration of PID infrastructure (1.0). Zenodo. https://doi.org/10.5281/zenodo.14506138
- [2] coli-conc - Cocoda. (2025). Coli-Conc.gbv.de. https://coli-conc.gbv.de/cocoda
- [3] Terminology Service Suite. (2025). Base4nfdi.de. https://terminology.services.base4nfdi.de/tss/comp/latest
- [4] TS4NFDI Api Gateway. (2025). Base4nfdi.de. https://terminology.services.base4nfdi.de/api-gateway/swagger-ui/index.html

Corresponding Author: Tilo Mathes, tilo.mathes@researchspace.com

ID T23

The Agentic Automation Canvas: A Metadata Framework for Human-AI Task Delegation

Author: Sebastian Lobentanzer[1]

[1] Helmholtz Munich

Agentic AI systems—autonomous software driven by large language models (LLMs)—promise significant efficiency gains by performing tasks that traditionally required human judgment. However, their deployment fundamentally involves control inversion: humans must step back and allow the system to take command. The ease of building impressive prototypes with current LLMs creates a dangerous mismatch: stakeholders see quick demos and assume production-ready solutions are within reach, while the bulk of actual work—handling edge cases, ensuring reliability, integrating governance, and validating real-world performance—lies beyond the prototype. Without an explicit contract defining expectations before control inversion occurs, organizations face disillusionment when promised benefits fail to materialize.

We present the Agentic Automation Canvas (AAC), a structured metadata framework that captures the essential agreement between human stakeholders and agentic AI systems. By formalizing this as machine-readable metadata rather than traditional requirements documents, the AAC enables automated validation of stakeholder agreements, cross-system interoperability, and integration with institutional governance workflows. The canvas formalizes user requirements with quantified benefit expectations and balances them with developer feasibility assessments including model baseline capabilities, governance stages with assigned accountability, data access rights, and evaluation metrics for comparing outcomes against expectations. The AAC is implemented as an interactive web application (https://aac.slolab.ai) exporting versioned RO-Crate packages. Where possible, the schema maps to established vocabularies (Schema.org, PROV-O, DCAT, P-Plan, FRAPO, DUO); for agentic-specific concepts such as benefit metrics, baseline capabilities, and control inversion agreements, we introduce new terms under a registered https://w3id.org/aac namespace.

By requiring this contract before control inversion, the AAC bridges the gap between prototype enthusiasm and production reality. The resulting RO-Crate travels alongside the project as a machine-readable artifact designed to support governance, auditable decision-making, and benefit tracking throughout the collaboration lifecycle.

Corresponding Author: Sebastian Lobentanzer, sebastian.lobentanzer@helmholtz-munich.de

ID T24

AIMWORKS: Template-Driven, Agentic Framework for FAIR Knowledge Graph Construction in Hydrogen Technologies

Author: Mohammad J. Eslamibidgoli[1]

Co-Authors: Kourosh Malek[1], Marjan Kohandani[1], Michael Eikerling[1]

[1] Forschungszentrum Jülich

Hydrogen and electrochemical energy research produces rapidly evolving, heterogeneous outputs - protocols, instrument settings, conditioned performance metrics, and multi-scale materials descriptors - that are difficult to curate into FAIR, machine-actionable metadata [1]. Many "LLM-first" knowledge-graph pipelines rely on monolithic prompts and ad-hoc post-processing, which can lead to inconsistent terminology, unreliable unit handling, and missing provenance and dataset details [2]. We present AIMWORKS [3], a template-driven, agentic framework that improves FAIR metadata by design. AIMWORKS uses a stable core vocabulary and a curated library of reusable templates for common experimental patterns (measurements, processes, experimental context, instruments, metrics, and dataset/provenance blocks) [4]. Given a user's natural-language research question, the system selects the most relevant templates, assembles them into a structured knowledge graph, and exports it in standard formats (RDF/JSON-LD). To ensure reliability, each template includes micro-level validation rules (SHACL) and the system applies deterministic checks and repairs to enforce consistent typing, represent conditioned metrics via a DataPoint pattern, normalise quantities and units using QUDT, and generate a complete dataset description (including title, license, and access information). Outputs integrate cleanly into downstream platforms such as Neo4j and institutional knowledge-graph infrastructures. Case studies from hydrogen technologies (polarization curves, impedance spectroscopy, durability protocols, and ionomer–catalyst-layer questions) show that the template-first approach improves metadata completeness and interoperability while reducing manual curation and providing a transparent trace from query to graph.

References:

- [1] Dreger, M., Eslamibidgoli, M. J., Eikerling, M. H., & Malek, K. (2023). Synergizing ontologies and graph databases for highly flexible materials-to-device workflow representations. Journal of Materials Informatics, 3(1), N-A
- [2] Dreger, M., Malek, K., & Eikerling, M. (2025). Large language models for knowledge graph extraction from tables in materials science. Digital Discovery.
- [3] https://meslamib3-aimworks4-streamlit-app7-2kyctx.streamlit.app
- [4] https://github.com/meslamib3/aimworks4/tree/main/templates

Corresponding Author: Mohammad J. Eslamibidgoli, m.eslamibidgoli@fz-juelich.de

www.helmholtz-metadaten.de