Data Stewardship goes Germany 2025

Report of Contributions

Contribution ID: **6** Type: **not specified**

KONDA: An LLM-based Tool for Semantic Annotation and Knowledge Graph Creation Using Ontologies for Research Data

Tuesday, September 30, 2025 2:15 PM (1h 30m)

Abstract

Poster

Author: KIM, Soo-Yon

Co-authors: Mr GÖRZ, Martin (RWTH Aachen University); Prof. GEISLER, Sandra (RWTH Aachen University)

Session Classification: Poster Session

Contribution ID: **7** Type: **not specified**

Facilitating Digitisation and Standardisation in Electrochemical Research through ELN Implementation and Template Development: A Use Case from a Data Steward Perspective

Tuesday, September 30, 2025 4:05 PM (20 minutes)

The digitisation of laboratory workflows brings both opportunities and challenges, particularly in the area of standardised documentation and data management. In the field of electrochemical research, where precision and reproducibility are essential, the lack of standardisation of data documentation significantly affects the efficiency and reliability of research results. This presentation describes the journey of implementing the Chemotion Electronic Laboratory Notebook (ELN) in an electrochemistry research group and highlights the role of the research data management (RDM) support staff in facilitating this transformative process.

Initiation Process and Challenges: The initiation of this transformation came from the research group's strong interest in digitising their laboratory workflows with a view to open research practices, internal standardisation of research data management (RDM) and the development of standard operating procedures (SOPs). A comprehensive needs assessment led to the selection of Chemotion ELN, recognised for its chemistry-specific functionality and its adaptability provided with the LabIMotion generic extension to develop customised documentation templates. The challenge was, firstly, to become familiar with the ELN and its functionalities itself and, secondly, to accurately represent complex laboratory processes within it, given the limitations of available documentation templates for electrochemical research in the ELN.

Solution Development and Collaboration: It was essential for this project to first analyse and structure the laboratory processes to develop a concept for the creation of the templates. At the same time, extensive literature research was carried out on existing documentation recommendations, which were also considered in the templates. Using the LabIMotion extension of the Chemotion ELN, templates for the documentation and data collection of various sub-processes were then created and optimised in terms of structure, content and design using a dynamic and iterative feedback loop. As a result, three comprehensive generic elements and one generic segment were created to reflect the workflow in this electrochemical research area. There was also close collaboration with the ELN developers, where bug reports and feature requests were discussed and subsequently incorporated into new software versions.

Implementation and Community Engagement: The gradual introduction of Chemotion ELN, initiated with a test phase involving a small subgroup of the research team, before extending to the entire group, underscores the importance of a phased approach to technology adoption in research settings. The project aims to go beyond a single research group and contribute to the wider research community. By providing the templates to an expert review process and subsequent provision via the LabIMotion Template Hub, the initiative aims to encourage the reuse and adaptation of this best practice across electrochemical research. In addition, the provision of field names and descriptions, including relevant ontologies where possible, should enable wider applicability across different ELN platforms and documentation workflows.

The implementation of Chemotion ELN in an electrochemistry research group, led by the dedicated support of the research data management staff in all phases, exemplifies a successful model for digital transformation in the laboratory. Through careful planning, collaboration with software developers and community engagement, the initiative is not only streamlining the group's data management practices but also sets a best practice for standardisation and open research in the wider scientific community.

Abstract

Talk

Author: SCHRÖTER, Annett (Friedrich-Schiller-Universität Jena)

Co-authors: BALDUCCI, Andrea (Friedrich Schiller University Jena); HAINTHALER, Andrea (Friedrich Schiller University Jena); KLEIN, Michel (Friedrich Schiller University Jena); JUNG, Nicole (Karlsruhe Institute of Technology)

Presenter: SCHRÖTER, Annett (Friedrich-Schiller-Universität Jena)

Session Classification: Talks

Contribution ID: **8** Type: **not specified**

Implementing FAIR in Practice: A Data Steward's Perspective from the MultiplEYE Project

Tuesday, September 30, 2025 2:15 PM (1h 30m)

This contribution presents a practical use case of embedded data stewardship within the ongoing *MultiplEYE* project—a large-scale, collaborative COST Action that aims to create a standardized, multilingual eye-tracking-while-reading dataset across more than 30 countries and 35 languages. As a Data Steward from the Leibniz Institute for Psychology (ZPID), I am embedded in the project's core team and support the implementation of FAIR principles throughout the entire research data lifecycle.

My role bridges infrastructure and research perspectives: I work closely with the researchers generating the data, while also coordinating with the infrastructure team responsible for building EyeStore, a FAIR-compliant repository hosted by ZPID and tailored to the long-term publication of eye-tracking data.

The contribution visualizes how FAIR principles are put into practice within each phase of the data workflow—from planning and standardized data collection to metadata documentation, quality reporting, and publication. Specific examples include the metadata schema, processing stages, and the role of documentation in supporting data transparency and reuse. It offers a real-world perspective on how an embedded data steward can help align the needs of researchers and infrastructure providers, promote standardization, and facilitate sustainable open data practices in complex, cross-institutional research settings.

Abstract

Poster

Author: MÜLLER, Marie-Luise (Leibniz-Institut für Psychologie (ZPID))

Session Classification: Poster Session

Contribution ID: **9** Type: **not specified**

AI-assisted data annotation for biomedical research consortia

Tuesday, September 30, 2025 2:15 PM (1h 30m)

Accurate and complete metadata is key for making research data findable and thus potentially reusable. However, since data annotation is a tedious and time-consuming process, compliance is often low, leading to missing or incomplete data set descriptions. A possible solution to this problem is the automatic extraction of metadata from data files and associated documents such as research articles and laboratory notebooks. The recent advancement in development of Large Language Models (LLMs) promises to greatly facilitate automatic data annotation, since LLMs excel in extracting specific information from text documents and inferring related concepts from the information provided in the text.

Here we present an approach for LLM-based automatic research data annotation developed by the Research Data Management (RDM) group at the Institute of Medical Biometry and Statistics (IMBI) of the University Medical Center Freiburg. The IMBI RDM group currently supports seven biomedical research consortia by offering training, consulting and software solutions for RDM. Consortium members have access to the open-source electronic laboratory notebook eLabFTW as well as the in-house developed research data platform fredato (Freiburg research data tool). fredato is a platform for collaboration and data sharing built on the open-source components Nextcloud (file storage, sharing and collaborative editing), GitLab (version control and metadata storage) and OpenSearch (search engine). To allow scientists to easily annotate their data with terms most relevant to their use case, the RDM group develops a metadata schema for each individual consortium in close collaboration with the scientists. This metadata schema is implemented as a JSON schema from which a web form for data annotation is built. Metadata are stored as JSON files and can be exported to public data repositories. For interoperability, elements of the metadata schemas are linked to biomedical ontologies and the metadata schemas are published as RDF (Resource Description Framework) knowledge graphs.

While bibliographic metadata of biomedical research articles can be imported from the PubMed database, data sets currently need to be annotated manually by the authors. To minimize the time and effort required for data annotation, we have developed a workflow of metadata prediction from publication PDFs using LLMs. The four-step workflow identifies biomedical entities (e.g. organisms, cell lines, genes, diseases) in the article full text using ChatGPT in combination with the consortium's metadata schema and the PubTator3 database of biomedical entities. We tested this workflow with 23 publications of the Collaborative Research Center 1479 "OncoEscape" and validated the suggested terms in structured interviews with publication authors. Overall, the approach worked very well with a precision of 98% (95% CI: 94%-100%), i.e., the vast majority of predicted biomedical entities were considered correct in the face-to-face interviews. Including or excluding the article's supplementary material did not result in a significant difference in precision of suggested terms.

We are currently working on implementing the LLM-assisted metadata prediction in fredato and testing the performance of several alternative LLMs. Thus, in the near future researchers will only need to review the pre-filled metadata web form and delete incorrect entries. In addition, we are testing the use of LLM-assisted metadata prediction in the schema development process to identify potentially overlooked terms.

Finally, we plan to extend the LLM-assisted metadata prediction to other sources beside article PDFs, including electronic laboratory notebooks such as eLabFTW and files stored on the OMERO

platform for microscopy data. This will allow us to capture metadata directly in the process of data creation, thereby ensuring that research data are immediately findable by interested researchers.

Abstract

Poster

Author: Dr BENADI, Gita (Institute of Medical Biometry and Statistics, Medical Faculty and University Medical Center, University of Freiburg)

Co-authors: Dr GIULIANI, Claudia (Institute of Medical Biometry and Statistics, Medical Faculty and University Medical Center, University of Freiburg); Dr ENGEL, Felix (Institute of Medical Biometry and Statistics, Medical Faculty and University Medical Center, University of Freiburg); Prof. BINDER, Harald (Institute of Medical Biometry and Statistics, Medical Faculty and University Medical Center, University of Freiburg); Dr KAIER, Klaus (Institute of Medical Biometry and Statistics, Medical Faculty and University Medical Center, University of Freiburg)

Session Classification: Poster Session

Contribution ID: **10** Type: **not specified**

Collaborative Infrastructure: Data Stewards and the Base4NFDI Service Ecosystem

Tuesday, September 30, 2025 2:15 PM (1h 30m)

The German National Research Data Infrastructure (NFDI) initiative Base4NFDI addresses interdisciplinary challenges in research data management by supporting the development and integration of cross-domain *basic services*. This poster introduces a selection of Base4NFDI services and highlights the pivotal role data stewards play as stakeholders, contributors, and multipliers in their development and adoption. It also illustrates how basic services support data stewards and other RDM professionals in their role.

Data stewards bridge domain-specific requirements and technical implementation, making them critical stakeholders in basic service development adoption. During basic service development, they can contribute their requirements to tailor services to their target groups' needs. In the *integration phase*, which is the second of three stages of service development focusing on integrating basic services broadly into NFDI consortia, they may facilitate service-related training and advocate for institutional buy-in.

Base4NFDI is developing a growing portfolio of basic services at various stages of maturity. The following selection presents key services that are already in use and particularly relevant for data stewards, as they are community-developed, cross-disciplinary, and directly support the work of RDM professionals and researchers.

Jupyter4NFDI provides a centralized JupyterHub and democratizes access to computational infrastructure via Jupyter notebooks. Thereby it fosters reproducible, FAIR digital research workflows. Data stewards are users and multipliers of this service. They help researchers adopt the platform, show how it can be used for workshops or may provide RDM training and resources via Jupyter notebooks.

IAM4NFDI delivers a state-of-the-art identity and access management system, enabling users to login to NFDI services with their home organizations account. Data stewards can point resource providers to potential integrations with IAM4NFDI solutions, facilitate access to virtual organisations for researchers where appropriate, and communicate user needs to service developers.

TS4NFDI offers widgets for terminologies, supporting harmonization and curation across disciplines. Its Terminology Service Suite (TSS) provides a collection of interactive widgets that ease the integration of terminology service functions into user applications. This makes it easier for data stewards to promote metadata provision, quality and interoperability within their communities.

DMP4NFDI standardizes and streamlines the creation and management of data management plans (DMPs) and software management plans (SMPs) using the open-source RDMO tool. It coordinates template creation, offers guidance, training and support, directly empowering data stewards in their advisory and quality assurance roles.

RDMTraining4NFDI assembles a modular collection of RDM training materials and supports the development of community-specific adaptations. Data stewards are both contributors to and primary users of these train-the-trainer resources, helping to disseminate best practices and foster a culture of data literacy.

By actively engaging with these services - through co-development, feedback, training, and advocacy - data stewards can help shape solutions that are not only technically sound but also aligned with the practical needs of their research community. With their involvement they can directly influence tools and standards that support their daily work, making them more usable, interoperable, and sustainable. Contributing in this way not only enhances their professional expertise and network but also reinforces their role as a key enabler of FAIR data practices.

To engage, data stewards and other RDM professionals can join NFDI sections and consortia or simply contact Base4NFDI's **Service Stewards** who consolidate technical and community requirements. The 2025 Data Stewardship conference in Karlsruhe provides a platform to discuss these dynamics, emphasizing use cases in interoperability, tool integration, and reproducibility.

Abstract

Poster

Authors: Mr GIESLER, André (Forschungszentrum Jülich); Ms MANSKE, Antje (GESIS); Ms TATSCHECK, Jana (ZPID); Dr REINHARDT, Martin (Universität Leipzig)

Session Classification: Poster Session

Contribution ID: **12** Type: **not specified**

Putting FAIR principles into action - Euro-BioImaging's FAIR Data Toolkit

Tuesday, September 30, 2025 2:15 PM (1h 30m)

In the rapidly evolving bioimaging landscape, adopting FAIR (Findable, Accessible, Interoperable and Reusable) principles is critical to maximising the value and impact of research data. Through a network of internationally renowned facilities, the European research infrastructure Euro-BioImaging provides open access to cutting-edge biological and biomedical imaging technologies and data services. As such, Euro-BioImaging is at the forefront of promoting and implementing FAIR practices in the bioimaging community, with the aim of advancing scientific progress and fostering innovation across the life sciences.

Here we present a comprehensive suite of services, resources and training opportunities, designed to promote FAIR practices and support researchers and facilities in making bioimaging data FAIR - in all stages of the data lifecycle. We offer data steward consultatins as well as regular live and on-demand training workshops to support the independent implementation of FAIR practices and the development of FAIR-compliant tools and workflows. The FAIR data toolkit also contains valuable self-learning resources and practical tools such as data management plan templates, repository decision trees, step-by-step data deposition guides, etc. Through close collaboration with major bioimage repositories these ensure seamless data deposition and retrieval.

By providing this toolkit, Euro-BioImaging aims to foster a culture of open science, facilitate data reuse, and accelerate scientific discovery in the bioimaging domain. We invite researchers and facilities to explore our resources and join us in shaping the future of FAIR bioimaging data.

Abstract

Poster

Author: KEMMER, Isabel (Euro-BioImaging ERIC)

Session Classification: Poster Session

Contribution ID: **13** Type: **not specified**

Data formats for UV-Vis spectra and particle/pore size distributions

Tuesday, September 30, 2025 2:15 PM (1h 30m)

Our collaborative research center is specialized in the rigorous design of particulate products. This includes the formation of particle systems through synthesis, the separation, classification, and characterization of these systems, and the simulation and optimization of synthesis and chromatographic separation. In this context, we have identified a lack of standardized formats for interdisciplinary data and metadata exchange that could meet our needs. In particular, ultraviolet-visible (UV-Vis)
spectra (reflectance and extinction) and particle- and pore-size distributions are the information that we generate and exchange frequently. In this contribution, we present two straightforward data
format structures that we have developed to alleviate this situation.

Abstract

Poster

Author: VU, Ngoc (Friedrich-Alexander-Universität Erlangen-Nürnberg, Collaborative Research Centre 1411 –Design of Particulate Products)

Co-authors: GÖTZ, Allison (Friedrich-Alexander-Universität Erlangen-Nürnberg, Collaborative Research Centre 1411 –Design of Particulate Products); MAUCH, Annika (Friedrich-Alexander-Universität Erlangen-Nürnberg, Collaborative Research Centre 1411 –Design of Particulate Products); SPRUCK, Christina (Friedrich-Alexander-Universität Erlangen,-Nürnberg, Collaborative Research Centre 1411 –Design of Particulate Products); PROHASKA, Florian (Friedrich-Alexander-Universität Erlangen-Nürnberg, Collaborative Research Centre 1411 –Design of Particulate Products); NAMDEO, Harsha (Friedrich-Alexander-Universität Erlangen-Nürnberg, Collaborative Research Centre 1411 –Design of Particulate Products); RODESTOCK, Jakob (Friedrich-Alexander-Universität Erlangen-Nürnberg, Collaborative Research Centre 1411 –Design of Particulate Products); SANDNER, Lukas (Friedrich-Alexander-Universität Erlangen-Nürnberg, Collaborative Research Centre 1411 –Design of Particulate Products); BECKER, Maik (Friedrich-Alexander-Universität Erlangen-Nürnberg, Collaborative Research Centre 1411 –Design of Particulate Products); STADELMAIER, Monika (Friedrich-Alexander-Universität Erlangen-Nürnberg, Collaborative Research Centre 1411 –Design of Particulate Products); NEES, Nico (Friedrich-Alexander-Universität Erlangen-Nürnberg, Collaborative Research Centre 1411 –Design of Particulate Products); SALCEDO, Nicolas (Friedrich-Alexander-Universität Erlangen-Nürnberg, Collaborative Research Centre 1411 –Design of Particulate Products); CARDENAS, Paola (Friedrich-Alexander-Universität Erlangen-Nürnberg, Collaborative Research Centre 1411 –Design of Particulate Products)

Session Classification: Poster Session

Contribution ID: **14** Type: **not specified**

Building FAIR from the Ground: Unifying Data Stewardship Across Meteorological and Climate Research Institutes

Tuesday, September 30, 2025 1:10 PM (20 minutes)

The Institutes of Meteorology and Climate Research (IMKs) at KIT operate a diverse range of instruments and models to investigate atmospheric and environmental processes. Historically, each working group has developed independent data management strategies, resulting in fragmented practices across the institutes. To address this challenge and ensure research data sustainability and FAIRness of research data, the IMK Data Management working group is implementing a unified data stewardship framework that integrates modern curation practices.
Our approach is built on three pillars:
1. **Infrastructure development** that supports structured and secure
data storage with robust metadata;
2. **Active data curation** throughout the research lifecycle to ensure data quality, provenance, and documentation;
3. **Training programs** tailored to researchers at all career stages.
This initiative aligns with national efforts such as DataHub and NFDI4Earth while systematically addressing all FAIR principles—enhancing findability through consistent metadata schemas, accessibility through standardized repositories, interoperability through domain-specific standards, and reusability through clear licensing and contextual documentation.
This poster presents our curation-centred methodology, implementation challenges across diverse research contexts, and emerging best practices for establishing scalable data stewardship infrastructure that supports day-to-day research activities while ensuring long-term data value.

Abstract

Talk

Author: BARTHLOTT, Sabine (KIT)

Co-authors: BÜTTNER, Nicole (KIT); Dr FÖSIG, Romy (KIT); HASSLER, Sibylle (KIT); Dr KERZENMACHER, Tobias (KIT); LOEWE, Katharina; LORENZ, Christof (KIT); OBERSTEINER, Florian (KIT); RASTEGAR, Arvin (KIT); REBMANN, Corinna (KIT)

Presenter: BARTHLOTT, Sabine (KIT)

Session Classification: Talks & Poster Pitches

Contribution ID: **15** Type: **not specified**

ReSeeD infrastructure: Advancing Data Stewardship and Sustainable Research Data Management in CRC 1280 "Extinction Learning" at Ruhr University Bochum

Tuesday, September 30, 2025 2:15 PM (1h 30m)

Abstract

Poster

Author: Ms ALASHLOO, Maryam (Ruhr University Bochum (RUB), Germany)

Co-authors: Ms PACHARRA, Marlene (Ruhr University Bochum (RUB), Germany); Ms WINTER, Nina (Ruhr University Bochum (RUB), Germany); Mr OTTO, Tobias (Ruhr University Bochum (RUB), Germany)

Session Classification: Poster Session

Contribution ID: **16** Type: **not specified**

GEOMAR's data steward concept and a data steward's view on day-to-day research data mangement activities

Tuesday, September 30, 2025 1:30 PM (20 minutes)

Research data management comes more and more into the focus of daily scientific workflows. This is driven by the increasing demand and focus on data-driven science and open science policies. Here, data steward or data experts or data champions or data managers, which ever role title is used, they play a crucial role in the research data management processes to achieve the envisioned goals of FAIR and open science.
Within this presentation we would like to introduce the institutional approach of GEOMAR's established data steward concept. We shortly highlight the different roles involved in research data management, such as data manager, data steward and data curator. And we will specifically describe the role of our domain-specific data stewards and their embedding into research projects but also the active partricipitation and exchange within the individual communities locally or across topics and domains such as within the Helmholtz Association or the Nationale Forschungsdateninfrastruktur (NFDI) and the federal state initiative FDM-SH.

Abstract

Talk

Author: GETZLAFF, Klaus (GEOMAR)

Co-author: Dr MEHRTENS, Hela (GEOMAR)

Presenters: Dr MEHRTENS, Hela (GEOMAR); GETZLAFF, Klaus (GEOMAR)

Session Classification: Talks & Poster Pitches

Contribution ID: **17** Type: **not specified**

Data Stewardship in Everyday Research: Use Cases and Best Practices at HZDR

Tuesday, September 30, 2025 2:15 PM (1h 30m)

Research data management (RDM) requirements have intensified over the past several years, particularly at large-scale Research Institutes (RI). At the Helmholtz-Centre Dresden-Rossendorf (HZDR), the interaction between internal processes and external developments illustrates this shift: growing demands for Open Science and FAIR data, funding policies of the DFG and the EU, the formation of national (NFDI) and European (EOSC) open science infrastructures, and strategic procedures within the Helmholtz Association—such as the Helmholtz Open Science Policy and Helmholtz Indicators for Research Data Management and Open Science—are shaping today's RDM environment. Simultaneously, day-to-day research at HZDR is marked by interdisciplinary collaboration, diverse scientific requirements, and operational complexities of large-scale facilities with international user communities. This environment demands data stewards and libraries—in collaboration with central scientific service departments—to practice both flexibility and strategic planning.

Since 2019, HZDR has appointed a Data Librarian to support RDM activities at the intersection of library services, IT infrastructure, and scientific institutes. With this role, several challenges typical for most research institutions but particularly pronounced at HZDR have been brought to light:

• The key challenge is to strike a balance in the institutional data repository RODARE: to establish standards that are detailed enough to meet the specific requirements of individual disciplines, while remaining flexible enough to describe the diverse outputs of interdisciplinary research.
• Data generation at large-scale research facilities or experiments, such as ELBE, HLD, and the Ion Beam Center—operated by shifts—still makes metadata curation and data sovereignty particularly challenging on a daily basis.
• Although the FAIR principles have widespread acceptance, their implementation is still perceived by many researchers as an extra burden, requiring not only technical services but rather a genuine cultural change.
• Entrenched practices ("We've always done it this way") are a barrier to innovation, despite the fact that funders increasingly request Open Access and RDM compliance.

At HZDR, support for RDM varies from repositories like ROBIS and RODARE, DOI registration, and data management plan consulting to workshops for postdocs and doctoral researchers. Among the success factors is the close collaboration of the library and the Scientific Data Management Group within the central IT department, which develops solutions for heterogeneous needs in a collaborative effort. Projects like HELIPORT and HERMES attempt to enable low-threshold access to RDM services with improved integration of metadata processes. The integration of facility-driven data streams into generic RDM frameworks remains a challenge—compounded by a project-based, temporary workforce.

Recognizing that progress in RDM cannot be achieved through isolated initiatives, HZDR initiated a center-wide process in 2024 to develop a sustainable, comprehensive approach that combines technical, organizational, and cultural components. Guided by our Data Policy and the HZDR Data Lifecycle Model, we are developing a Data Management Strategy that is purposely designed as a dynamic, living document. Its purpose is to create a catalog of demands that aligns existing resources (repositories, PID services, DMPs, lab notebooks) with the differing needs of our scientific communities—while maintaining congruence with national (DAPHNE4NFDI, PUNCH4NFDI) and international (PaNOSC/EOSC) RDM initiatives.

Importantly, this strategy is not a top-down requirement but a co-creative endeavor that engages researchers directly and gives them the structures and resources necessary to meet rising expec-

tations. Initial workshops have already been held to take in input and detail demands, with further targeted sessions planned. Rather than imposing requirements, our strategy is to enable researchers, while establishing a basis to integrate RDM more profoundly into scientific practice and academic reward—through novel incentive mechanisms such as counting RDM outputs, relevance for Helmholtz's Program-Oriented Funding (POF), and alignment with Helmholtz quality metrics.

This talk will provide insight into how HZDR is navigating the challenging balancing act between generic infrastructure and discipline-specific customization, between service provision and the cultural change that is needed, and between Open Access and technology transfer. It becomes clear that data stewardship in everyday research is far more than a technical task —it is also a process of organizational development that requires lengthy coordination processes, overcoming resistance, and building bridges between diverse stakeholders and communities. The ongoing development of our Data Strategy should therefore not be seen as an endpoint, but as an expression of this continuous negotiation process —including how we identify requirements and address the need for additional resources. In my talk, I will illustrate, using concrete examples from HZDR practice, how we are shaping this process, what challenges are emerging, and which approaches we are taking to gradually embed research data management as a natural part of the everyday practice of research of a large, international, and interdisciplinary research center.

Abstract

Poster

Author: FIEDLER, Maik (HZDR)

Co-author: Dr KNODEL, Oliver (HZDR)

Session Classification: Poster Session

Contribution ID: **18** Type: **not specified**

Data Stewards as a Service - NFDI-MatWerk

Wednesday, October 1, 2025 9:00 AM (20 minutes)

In NFDI-MatWerk, there have been extensive efforts of integrating data stewards into the overall concept of the consortium in the past. This led to the development of a dedicated data stewardship concept [1] to capture the needs and assist to tackle them within the consortium. In the first steps of the consortium, this was necessary and has fulfilled its function.

Now we want to roll out the solutions, tools, and general knowledge of the consortium to the MSE community. However, MatWerk has only very limited capacities with a total of 1.5 Data Stewards. Therefore, we will offer what is called Data Stewardship as a Service (DSAAS) as part of our so-called Community Membership in the future. MSE researchers can register at a low level as MatWerk Community Members and thus take advantage of the DSAAS. In this way, we aim to support researchers, increase their engagement with NDFI-MatWerk, and strengthen their trust in us.

Specific services that DSAAS will offer include: dedicated regular support hours, onboarding support, introduction to tools, workflow guidance, requirements discussions, trainings⋯

We –the data stewards of NFDI-MatWerk - would like to take this opportunity at DSGG to present the DSAAS, exchange ideas, and further develop our concept with qualified feedback from our peers.

[1] Daei Rezaei Moghaddam, A. (2024). Data Steward's Workflow in NFDI-MatWerk. Zenodo, URL: https://doi.org/10.5281/zenodo.14012860

Abstract

Talk

Authors: Mr MOGHADDAM, Amirreza (RWTH Aachen University); Dr WINKENS, Georg (FZ Jülich); Dr MOHRBACHER, Julia (Albert-Ludwigs-Universität Freiburg); Mrs GRÜNWALD, Katharina (RWTH Aachen University)

Presenters: Mr MOGHADDAM, Amirreza (RWTH Aachen University); Dr WINKENS, Georg (FZ Jülich); Dr MOHRBACHER, Julia (Albert-Ludwigs-Universität Freiburg); Mrs GRÜNWALD, Katharina (RWTH Aachen University)

Session Classification: Talks

Contribution ID: **19** Type: **not specified**

NFDI & Data Stewards: Working together for a sustainable research data management

Tuesday, September 30, 2025 2:15 PM (1h 30m)

The German National Research Data Infrastructure (NFDI) is a science-led network that aims to establish a sustainable, decentralised research data ecosystem in Germany. The objective is to provide reliable services, establish cross-disciplinary standards, secure access to high-quality research data and enable their reuse in accordance with the FAIR principles. 26 consortia from all research areas and the Base4NFDI initiative are jointly developing concepts, services and infrastructures to simplify the handling of research data for the future. They are working together to realise the vision of the NFDI Association: data as a common good for excellent research, organised by the scientific community in Germany.

Data stewards and NFDI play central roles in research data management (RDM). As bridge builders between research, IT and administration, data stewards design local implementations of RDM standards. They can put NFDI impulses into practice at their institutions. We present the structure and strategic goals of NFDI as well as possible participation opportunities for data stewards.

The poster is intended to open a dialogue: How can cooperation between NFDI and data stewards be strengthened in order to ensure sustainable research data management?

Abstract

Poster

Authors: RATHGEB, Nathalie (Nationale Forschungsdateninfrastruktur (NFDI) e.V.); WILJES, Cord (NFDI); FUCHSLOCH, Stefanie (Nationale Forschungsdateninfrastruktur (NFDI) e.V.)

Session Classification: Poster Session

Contribution ID: **20** Type: **not specified**

Curate Me If You Can: Repository Metadata in re3data

Tuesday, September 30, 2025 2:15 PM (1h 30m)

The Registry of Research Data Repositories re3data is a major directory of research data repositories that makes metadata on repositories for the permanent storage and access of data sets openly available according to re3data's Metadata Schema Version 4.0 (https://www.doi.org/10.48440/re3.014). The international index currently lists over 3350 digital repositories across all scientific disciplines –critical infrastructures to enable the global exchange of research data. It is widely used by researchers to identify repositories to find and deposit data. Funders and publishers recommend re3data to guide investigators and authors to repositories to meet their requirements for sharing the data that support grant-funded research and publications. Librarians and lecturers use re3data to promote data literacy. Additions of new repositories from users like data stewards, repository operators or researchers in re3data are managed by an international Editorial Board. The editorial process includes a multi-stage review, adapting best practices in science. A team of research data professionals thoroughly analyzes the repositories and ensures the metadata completeness and quality of the records.

Therefore, re3data promotes open science practices and the visibility of science-driven open infrastructures for research data. The service makes its data open, available under a CC0 license. Its web interface and API enable end users and third-party providers to utilize the largest index of research data repositories in the world.

The poster will focus on the use and benefits which re3data provides to the research data community. It will demonstrate the simplicity of the suggest and the update process of a repository record in the registry. We want to facilitate updates through the community via a guided process of submitting change requests for repositories. Curating the repository information together with the research data community is crucial and contributes to high standards in data management, follows the FAIR principles and promotes a culture of open science. A use case for this is currently being implemented with members of the NFDI. The detailed process will be illustrated on a poster and walked through by the presenters.

Abstract

Poster

Authors: Mrs NEIDIGER, Charlotte (re3data / Karlsruhe Institute of Technology); ULRICH, Robert (re3data / Karlsruhe Institute of Technology)

Session Classification: Poster Session

Contribution ID: **21** Type: **not specified**

Learning through Experience: Bootstrapping an Electronic Lab Notebook as a central IT-Service for Thuringian universities –Challenges and Lessons Learned (*)

Tuesday, September 30, 2025 2:15 PM (1h 30m)

Electronic lab notebooks (ELNs) have emerged as vital tools in modern research, significantly impacting the way data and metadata are documented, managed, and shared throughout the research data lifecycle. By enabling structured, standardized, and consistent metadata collection from the start of data generation, ELNs notably improve metadata quality. Their digital nature not only facilitates backups or collaboration, but also enhances the discoverability of new connections between datasets, thereby promoting data reuse. As the demand from researchers and funding bodies for robust data management practices grows, the need for institution-wide ELN solutions becomes increasingly urgent.

Despite these well-known benefits of ELNs, action at the universities in Thuringia had not yet gone beyond initial efforts. Until early 2025, none of the universities offered their researchers a centrally hosted and administrated, ready-to-use generic ELN service. Instead, individual research groups were left alone to navigate the complexities of selecting, implementing, and maintaining their own ELN solutions. This fragmented approach led to inefficiencies, duplication of effort, and an uneven landscape where smaller or less technically resourced groups are at a disadvantage.

Recognizing this critical gap in provision and equipped with initial experiences in inter-institutional collaboration with the federal network of university computing centers from a previous project, the Thuringian Research Data Management (RDM) state initiative TKFDM took up the challenge to overcoming the institutional level and initiated a joint effort to establish an ELN as a shared IT-service for all universities in the state.
Now that the goal has been set and in sight, should we not expect a straightforward and direct path to it and also expect to be able to focus soon on introductory courses and on supporting end users in research teams using the ELN?

This poster reflects on the technical and organizational challenges encountered in the ongoing efforts. Key issues include the need for software development to ensure compatibility with the DFN Authentication and Authorization Infrastructure (AAI), the development of a sustainable concept for shared support and onboarding, and the complexities of cross-institutional coordination –both internally (between project coordination and operational teams) and externally (with stakeholders such as staff councils and data protection officers). It also discusses the importance of clear governance structures, effective communication channels and the need to adapt to obstacles and evolving requirements.

By sharing our experiences and lessons learned, we aim to provide data stewards and research IT professionals with practical insights into the complex bottom-up process of establishing a federated IT service, and to foster discussion on strategies for successful cross-institutional research data infrastructure initiatives.

*) „[…] so far.“H. J. Simpson in The Simpsons Movie, 2007, USA

Abstract

Poster

Authors: Mr LINDT, Kevin (Thüringer Kompetenznetzwerk Forschungsdatenmanagement/ Thuringian Competence Network for Research Data Management (TKFDM)); KIRSCH, Stefan (Thüringer Kompetenznetzwerk Forschungsdatenmanagement/ Thuringian Competence Network for Research Data Management (TKFDM))

Session Classification: Poster Session

Contribution ID: **22** Type: **not specified**

Building Bridges through Trust –Data Trusteeship as a Key Enabler for Science-Industry Data Sharing

Tuesday, September 30, 2025 2:15 PM (1h 30m)

Collaborations between academia and industry are essential to address complex societal challenges —but they often fail due to a lack of trust in data sharing. While intra-academic data sharing is increasingly supported by established infrastructures and norms, researchers remain reluctant to share data with commercial actors. Concerns about data misuse, loss of control, or legal liability create real barriers.

The proposed poster introduces the concept of data trusteeship as a trust-based solution to this dilemma. Envisioned in the EU's Data Governance Act, data trustees act as neutral intermediaries who mediate data access or analysis without pursuing their own interest in the data. Their key role is to ensure that the conditions and terms of data use respect the intentions of both sides—thus making sensitive or strategically valuable data more shareable, especially in contexts where science and industry collaborate across institutional boundaries.

Rather than presenting a ready-made use case, this poster advocates for an integration of the trusteeship idea into ongoing aspirations in Research Data Management. It aims to stimulate reflection within the RDM community: Rooted in trust, data trusteeship may be the institutional innovation we need to bridge the gap between science and industry—enabling data sharing beyond science while upholding autonomy, compliance, and research integrity.

Abstract

Poster

Authors: BRENNEIS, Andreas; EGGERICHS, Lisa

Session Classification: Poster Session

Contribution ID: **23** Type: **not specified**

RDM Compas –strengthening professional competencies in data curation

Tuesday, September 30, 2025 2:15 PM (1h 30m)

As Open Science and the FAIR data principles continue to reshape the research landscape, the roles of data stewards and data curators are gaining increasing importance. These professionals are essential for implementing sustainable data practices, ensuring long-term accessibility, and improving the quality of research data throughout their lifecycle. Within this context, Research Data Centers (RDCs) play a central role by curating, safeguarding, and providing access to high-quality data for secondary use, particularly in social sciences. However, in contrast to researchers who have various training opportunities, the education of data curators and other Research Data Management (RDM) professionals remains underdeveloped and scattered across institutions and disciplines. To address this gap, RDM Compas (Research Data Management Competence Base) was developed as an open accessible online platform aimed at strengthening the competencies of professionals working in RDM and research data curation, particularly in social, behavioural, educational, and economic sciences. RDM Compas is a project of KonsortSWD - a consortium of the German National Research Data Infrastructure (NFDI).

RDM Compas provides structured and flexible learning paths tailored to the specific needs of data stewards, data curators, and other RDM staff, aligning these with the Data Curation Lifecycle (Higgins, 2008). While the more widely known research data lifecycle focuses on researcher's perspective and various stages data undergoes during a research project, the Data Curation Lifecycle (Higgins, 2008) emphasizes the process-oriented activities required to manage, curate, and preserve data for future reuse. This perspective underpins the platform's approach to offered training and information materials.

RDM Compas offers a comprehensive knowledge base –RDM Knowledge Base - designed to support day-to-day work of RDC staff. It covers a wide range of RDM topics including metadata standards, persistent identifiers, legal and ethical aspects, data documentation etc. The articles are useful for both emerging and experienced professionals, and they are aimed at facilitating knowledge exchange between RDM professionals and researchers. To complement the Knowledge Base, RDM Compas includes an interactive, self-paced Trainingcenter. It offers internally developed courses and links to external training opportunities, emphasizing practical workflows in areas such as data quality assurance, anonymization, legal compliance etc. The training materials adhere to Open Educational Resources (OER) principles, ensuring that content is freely available and adaptable to individual learning needs and schedules.

In addition to the generic content, the platform also addresses specific challenges associated with different types of research data generated in social, behavioural, educational, and economic sciences. For instance, curating qualitative data often requires particular attention to ethical concerns, anonymization, and detailed documentation to ensure contextual integrity. This is especially important when dealing with sensitive topics or vulnerable populations. In contrast to qualitative data, corporate data, which is frequently collected for non-research purposes, presents unique challenges related to contractual agreements, data restructuring, and legal compliance.

RDM Compas is a community driven platform that aims to foster collaborations, share best practices and develop a foundation for professional standards in RDM. It encourages contributions from RDCs and other institutions to expand its informational and training resources. In doing so, it serves as a training tool and as a community for RDM professionals and thus represents a signifi-

cant step towards the standardisation of research data curation in social, behavioural, educational, and economic sciences.
By presenting specific challenges encountered in some phases of the Data Curation Lifecycle, we will illustrate the most relevant and demanding competences required in practice. Drawing on insights from selected RDCs, the poster highlights how these competencies are applied in everyday data curation scenarios.

Abstract

Poster

Author: CHICHERINA, Ekaterina (GESIS –Leibniz Institute for the Social Sciences)

Session Classification: Poster Session

Contribution ID: **24** Type: **not specified**

Building a Comprehensive Federated Network of RDM Helpdesks

Wednesday, October 1, 2025 11:00 AM (20 minutes)

RDM helpdesks are the primary point of contact for researchers with questions about research data management (RDM). In RDM helpdesks, data stewards provide support in various formats, ranging from a first-level-support hotline to long-term complex consultancy. Data stewards are working at a wide range of institutions: universities (including clusters of excellence and collaborative research centres), non-university research institutions, data competence centres, RDM federal state initiatives, and national and international infrastructure initiatives (National Research Data Infrastructure (NFDI), European Open Science Cloud (EOSC)).

RDM helpdesks are often specialised non-specific disciplines, topics, data types and / or target groups. This results in a diverse landscape of RDM helpdesks across Germany, causing significant challenges to researchers and data stewards. Researchers struggle to find the right helpdesk for their needs, which leads to missed opportunities in research and funding as well as data that is less fit for re-use. Data stewards face huge workloads because they constantly have to acquire new knowledge beyond their key subjects on a wide range of topics and don't have resources to compensate for a higher workload during peak periods.

To meet these challenges, the RDM Helpdesk Network was initiated in 2023 in a two-day Community Workshop [1] by 50 representatives of RDM helpdesks. The vision of the RDM Helpdesk Network is to connect all RDM helpdesks to a comprehensive federated network. This will allow researchers to easily identify the helpdesk that is best suited to answer their questions, and enable RDM helpdesks to efficiently cooperate and dynamically allocate resources. To bring about this vision, the RDM helpdesk Network is working to
- connect RDM helpdesks so they can learn and benefit from each other's experience;
- create a central signpost service that refers researchers to the helpdesk best suited for their needs;
- connect the existing ticketing systems of RDM helpdesks to enable referring joined consultations;
- establish a ticketing service that can be used by helpdesks without an own ticketing infrastructure;
- create a comprehensive catalogue of RDM helpdesk, allowing users to search and find the helpdesk that best meets their requirements.

The RDM Helpdesk Network is a joint initiative of the German National Research Data Infrastructure (NFDI), RDM Federal State Initiatives, and institutional RDM support helpdesks. It was initiated in November 2023 in a community workshop constituted as a working group of the NFDI Section Training & Education [2]. We published a charter [3] that outlines the motivation and objectives of our working group. This document serves as a guide for our collaborative efforts. Additional workshops online and at conferences (BiblioCon 2024; 1st Base4NFDI User Conference; e-Science-Tage 2025) were conducted, and we created a mailing list [4] that has already attracted 175 subscribers. In order to implement the strategy for a comprehensive federated Helpdesk Network in a concrete and sustainable manner, we submitted the proposal Support4RDM to the NFDI basic service initiative Base4NFDI [5] in April 2025.

In our talk, we will present the RDM Helpdesk Network, give an overview of its activities and plans, and invite participants to join the working group. We are particularly looking forward to a dynamic and engaging dialogue with the data stewardship community, and we strongly encourage data stewards to actively contribute to our working group's discussions and activities.

[1] https://www.forschungsdaten.org/index.php/2023-11-20_Workshop_Vernetzung_der_FDM-Helpdesks
[2] https://www.nfdi.de/section-edutrain
[3] https://doi.org/10.5281/zenodo.14035822

[4] https://www.listserv.dfn.de/sympa/info/fdm-helpdesks
[5] https://base4nfdi.de/

Abstract

Talk

Authors: WILJES, Cord (Nationale Forschungsdateninfrastruktur (NFDI) e.V.); ENGEL, Judith (Universität Bremen); WEIMER, Lukas (Universität Göttingen); SCHMIDT, Marcus (ZALF e.V.); LEMAIRE, Marina (Universität Trier); STELLMACHER, Martha (SLUB Dresden); MÜLLER-PFEFFERKORN, Ralph (TU Dresden)

Presenters: WILJES, Cord (Nationale Forschungsdateninfrastruktur (NFDI) e.V.); ENGEL, Judith (Universität Bremen); WEIMER, Lukas (Universität Göttingen); SCHMIDT, Marcus (ZALF e.V.); LEMAIRE, Marina (Universität Trier); STELLMACHER, Martha (SLUB Dresden); MÜLLER-PFEFFERKORN, Ralph (TU Dresden)

Session Classification: Talks (joint session with FDM@Campus)

Contribution ID: **25** Type: **not specified**

Metadata by Design –Building Profiles for FAIR and Reusable Data

Tuesday, September 30, 2025 2:15 PM (1h 30m)

Creating well-structured metadata profiles is essential for enabling long-term access, understanding, and reuse of research data. However, many researchers struggle with deciding what information to include, how to structure it, and where to find standardized terminology. This poster offers practical guidance tailored to everyday researchers who aim to store their data in platforms like Coscine.

The process begins with identifying relevant metadata elements from available data sources. From there, researchers select appropriate metadata standards—such as Dublin Core, schema.org, or DCAT—which are widely used and serve general-purpose needs. To ensure semantic consistency and clarity, controlled vocabularies and ontologies are incorporated into the profiles.

To meet the specific needs of various disciplines, researchers often rely on schema extensions or domain-specific vocabularies. These help enrich the metadata and make it more meaningful within particular scientific contexts.

A key component of this approach is the use of the Resource Description Framework (RDF) to formally represent metadata, making it both machine- and human-readable. Tools like the AIMS Metadata Generator support this process by assisting in the creation of RDF-based metadata.

By building metadata profiles thoughtfully and using established standards and tools, researchers can ensure that their data remains understandable and reusable—even years after its initial creation.

The poster presents a step-by-step strategy to guide researchers through designing effective metadata profiles. Applying common standards not only improves internal data quality but also enables metadata exchange across repositories, ultimately fostering interoperability.

Abstract

Poster

Author: DUKKART, Kseniia (IT Center RWTH Aachen)

Session Classification: Poster Session

Contribution ID: **26** Type: **not specified**

Empowering Educators and RDM Professionals: The GesundFDM Trainer Tandem Approach to Teaching Subject-specific Data Literacy

Tuesday, September 30, 2025 4:25 PM (20 minutes)

Abstract

Talk

Authors: KOCH, Katharina (Hochschule Bochum); Ms WERNER, Stephanie (Hochschule Darmstadt); Prof. QUILLING, Eike (Hochschule Bochum)

Presenters: KOCH, Katharina (Hochschule Bochum); Ms WERNER, Stephanie (Hochschule Darmstadt); Prof. QUILLING, Eike (Hochschule Bochum)

Session Classification: Talks

Contribution ID: **27** Type: **not specified**

Open Science in Action: Hands-On Spatial Data Training with Jupyter Notebooks

Tuesday, September 30, 2025 3:45 PM (20 minutes)

The Jupyter ecosystem for data-driven research

The Jupyter ecosystem offers a powerful and intuitive approach to modern data-driven research. It seamlessly integrates explanatory narratives, executable code, and visualisations within a single, web-based interactive environment, lowering the barrier to engaging with complex data workflows (Kluyver et al., 2016). A centralised infrastructure, such as the Jupyter4NFDI platform (used by initiatives like NFDI4Biodiversity), further enhances accessibility. These platforms provide scalable, cloud-based computing environments. Our training materials are compatible with the Jupyter4NFDI hub and can be accessed and run directly through a web browser. This seamless integration allows users to interact with the training materials in their browsers without needing complex local software installations or managing software dependencies. This makes advanced computational tools more accessible, supports reproducibility across different teams and institutions, and aligns with open science principles by promoting shared, standardised environments.

Furthermore, the training materials offer a valuable blueprint for creating similar open educational resources. The entire process from authoring interactive Jupyter Notebooks to publishing a polished, publicly accessible static website is streamlined using standard Git version control and managed through platforms like GitLab. This automated CI/CD (Continuous Integration/Continuous Deployment) pipeline allows the repository to be easily cloned and adapted, enabling others to efficiently build upon our framework for their teaching needs or develop training in different spatial science domains. This significantly lowers the barrier to creating and sharing high-quality, interactive learning experiences.

Showcasing reproducible spatial data science

This contribution presents Jupyter notebook training materials (see https://training.fdz.ioer.info) developed as a hands-on, open educational resource. They offer interactive tutorials and reproducible workflows for students, researchers, and practitioners conducting fully transparent, reproducible, narrative-embedded data-driven research focusing on spatial data (Dworczyk et al., 2025). The training materials provide users explicit guidance through the entire research output lifecycle (Higgins, 2008). Chapters cover topics ranging from understanding the FAIR (Findable, Accessible, Interoperable, Reusable) principles (Wilkinson et al., 2016) and good scientific practices for handling data to accessing, processing, analysing, and visualising spatial data and, ultimately, publishing findings. Each chapter of the training material features practical examples, including Python code, visualisations, and described processes. For instance, we demonstrate how to make code more broadly applicable and reusable by adhering to conventions like notebook parameterisation. This technique allows code to be easily adapted for different contexts, such as applying the same spatial data processing to a new region, thereby enhancing the reusability and impact of the shared work.

We use illustrative data examples, such as accessing biodiversity observations via the Global Biodiversity Information Facility (GBIF) API, contextualised with data relevant to the Lebendige Atlas der Natur Deutschlands (LAND), and retrieving raster-based land use and environmental data from the Monitor der Siedlungs- und Freiraumentwicklung (IÖR-Monitor). These datasets demonstrate reproducible research practices from programmatic data ingestion (e.g., API querying, handling various file formats) and pre-processing (e.g., data cleaning, geospatial operations like reprojection, clipping, and overlays) to advanced spatial analysis and visualisation (e.g., creating static and interactive maps).

Crucially, the material emphasises the creation of a “Replication Package”, a versioned archive

containing notebooks, scripts, data subsets, generated outputs, and a detailed README. This ensures the work can be fully understood, reproduced, and cited (e.g., via deposition in repositories like ioerDATA or Zenodo). While accessible via Jupyter4NFDI, we also detail using the versioned Carto-Lab Docker container. A containerised setup can encapsulate the precise software environment, including specific versions of Python and numerous cartographic and geospatial libraries, to guarantee full computational reproducibility and long-term preservation.

The collaborative online framework proved invaluable for efficient teamwork and knowledge sharing. We leveraged Git for robust version control and issue tracking to manage development and ensure incremental improvements. Real-time collaborative editing within Jupyter sessions enabled direct interaction and mentoring between junior and senior team members. This entire ecosystem, supported by a centrally maintained IT infrastructure, streamlined the research and publication process by offloading technical burdens from individual researchers and simplifying the integration of new collaborators.

Our training materials showcase practical tools and workflows that enhance data stewardship and foster robust reproducibility in spatial science. By providing open-source code, example data, detailed methodological explanations within an executable format, and explicit guidance on creating citable "Replication Packages," these resources actively support the creation and dissemination of FAIR research outputs. The material is designed for open dissemination and re-use, promoting learning, critical engagement with data-driven methods, and the widespread adoption of open science practices across the research community.

Acknowledgement

The development of these training materials was made possible through support from NFDI4Biodiversity (phase II).

Keywords

Jupyter Notebook, Jupyter Book, Open Science, Reproducibility, FAIR Data, Spatial Data, Geodata, Python, NFDI4Biodiversity, Biodiversity Informatics, Environmental Data Science, Training, Open Educational Resources, Carto-Lab Docker, Replication Package.

References

Dworczyk, C., Dunkel, A., Rafiei, F., Syrbe, R.-U., 2025. Exploring Spatial and Biodiversity Data with Python and JupyterLab. https://training.fdz.ioer.info

Higgins, S., 2008. The DCC Curation Lifecycle Model. International Journal of Digital Curation 3, 134–140. https://doi.org/10.2218/ijdc.v3i1.48

Kluyver, T., Ragan-Kelley, B., Pérez, F., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., Willing, C., Jupyter Development Team, 2016. Jupyter Notebooks –a publishing format for reproducible computational workflows, in: Positioning and Power in Academic Publishing: Players, Agents and Agendas. IOS Press, pp. 87–90. https://doi.org/10.3233/978-1-61499-649-1-87

Wilkinson, M.D., et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018. https://doi.org/10.1038/sdata.2016.18

Abstract

Talk

Authors: DELLA CHIESA, Stefano (Leibniz Institute of Ecological Urban and Regional Development); DUNKEL, Alexander (Leibniz Institute of Ecological Urban and Regional Development); DWOR-

CZYK, Claudia (Leibniz Institute of Ecological Urban and Regional Development); SYRBE, Ralf-Uwe (Leibniz Institute of Ecological Urban and Regional Development)

Presenters: DELLA CHIESA, Stefano (Leibniz Institute of Ecological Urban and Regional Development); DUNKEL, Alexander (Leibniz Institute of Ecological Urban and Regional Development); DWORCZYK, Claudia (Leibniz Institute of Ecological Urban and Regional Development); SYRBE, Ralf-Uwe (Leibniz Institute of Ecological Urban and Regional Development)

Session Classification: Talks

Contribution ID: **28** Type: **not specified**

Beyond the ELN: RSpace as an Integrator for Data and Metadata in the Research Data Lifecycle

Tuesday, September 30, 2025 2:15 PM (1h 30m)

Research data management increasingly requires seamless integration of multiple components across the entire research lifecycle. While Electronic Laboratory Notebooks (ELNs) provide crucial functionality for experimental documentation, today's data-intensive research demands more comprehensive solutions that connect primary data generation, analysis, and preservation workflows.

This presentation explores how RSpace functions not merely as an ELN but as a central integrator that bridges the gap between everyday research activities and institutional RDM requirements. We demonstrate two practical implementation cases that showcase vertical interoperability illustrating how researchers can maintain focus on their scientific workflows while simultaneously enhancing data interoperability through machine-readable formats and standardized metadata.

Our first use case examines field collection data management, demonstrating how RSpace improves the FAIRness of physical samples and sample metadata by using globally resolvable identifiers throughout the entire research lifecycle. The second showcases RSpace's unique Data Management Plan (DMP) integration capabilities, allowing researchers to fulfill reporting requirements directly within their working environment rather than treating them as separate administrative exercises.

Both examples highlight RSpace's role in creating an infrastructure where persistent identifiers, standardized metadata, and repository connections coexist within researchers' daily workflows. This approach transforms data stewardship from an administrative burden into an integrated component of the research process, ultimately ensuring the reproducibility of research data and results without disrupting scientific productivity.

The presentation will also address collaboration with the NFDI, specifically highlighting RSpace as a reference example in the ELN focus group of PID4NFDI, creating synergies that provide practical tools for researchers across Germany. By sharing these concrete implementations, we aim to contribute to the broader conversation about practical data stewardship solutions that serve both individual researchers and institutional RDM objectives.

Abstract

Poster

Author: Mr MACNEIL, Rory (ResearchSpace)

Co-author: MATHES, Tilo (ResearchSpace)

Session Classification: Poster Session

Contribution ID: **29** Type: **not specified**

LLM-Assisted Variable Annotation using the I-ADOPT Framework

Tuesday, September 30, 2025 2:15 PM (1h 30m)

Effective data stewardship in research hinges upon the consistent and FAIR (Findable, Accessible, Interoperable, Reusable) representation of scientific variables across diverse environmental disciplines. Within the Helmholtz Earth and Environment DataHub initiative, we are therefore developing an innovative approach utilizing Large Language Models (LLMs) to support data producers by automating the semantic annotation of research data. Our service employs the community-driven I-ADOPT framework, which decomposes variable definitions from natural language descriptions into essential atomic components, ensuring naming consistency and interoperability.

In this poster, we present our approach to developing an LLM-based annotation service, highlighting key challenges and solutions as well as integration into higher-level infrastructures of the Helmholtz DataHub and beyond. The proposed annotation framework significantly streamlines the integration and harmonization of environmental data descriptions across domains such as climate, biodiversity, and atmospheric sciences, aligning closely with the objectives of the NFDI and the European Open Science Cloud (EOSC).

This contribution demonstrates how advanced semantic annotation tools can effectively support data stewardship in practical research contexts, enhancing reproducibility, interoperability, and collaboration within the scientific community.

Abstract

Poster

Authors: RASTEGAR, Arvin (KIT IMK-IFU); Dr LORENZ, Christof (KIT IMK-IFU)

Session Classification: Poster Session

Contribution ID: **30** Type: **not specified**

MARIE's Integrated Research Data Management Framework

Tuesday, September 30, 2025 2:15 PM (1h 30m)

Abstract

Poster

Author: MARSON, Marcel (IT.SERVICES, Ruhr University Bochum, Bochum, Germany)

Co-author: Dr REHWALD, Stephanie (Research Data Services (RDS), University Library, University of Duisburg-Essen, Duisburg, Germany)

Session Classification: Poster Session

Contribution ID: **31** Type: **not specified**

Solar Cell Research Data Digitization

Tuesday, September 30, 2025 2:15 PM (1h 30m)

While FAIR data is essential for the long-term sustainability of science, researchers need practical incentives to engage with data stewardship. At HZB, the data steward team has collaborated closely with solar cell research groups to implement efficient and comprehensive data management workflows. This initiative is built on a strong partnership between data stewards and scientists, aiming to transform research outputs into FAIR-compliant datasets while providing field-specific analytical tools to support the exploration, visualization, and analysis of both individual and group-level data.
A key element of this effort is the collaboration with FAIRmat, a consortium within the German National Research Data Infrastructure (NFDI), which develops and maintains the NOMAD platform. NOMAD Oasis enables the creation of a customized NOMAD instance, allowing research teams to benefit from NOMAD's features while tailoring data management to their specific needs and institutional policies. Leveraging this scalable and interoperable framework, we have developed domain-specific schemas, ontologies and plugins for solar cell research. These enable the structured capture of data and metadata from laboratory processes and measurements.

The HZB NOMAD plugin has shown encouraging results: it is being adopted by an increasing number of research groups, including a large EU consortium project- PERSEUS, which focuses on printed perovskite solar cells for large-area applications. This work serves as a practical example of how data stewardship can be integrated into daily research activities to foster reproducibility, interoperability, and long-term value.

Abstract

Poster

Author: GKOGKOU, Dimitra (Helmholtz-Zentrum Berlin (HZB))

Co-authors: Dr NANDAYAPA, Edgar (Helmholtz-Zentrum Berlin (HZB)); Dr GÖTTE, MIchael (Helmholtz-Zentrum Berlin (HZB)); GRANIERO, Paolo (Helmholtz-Zentrum Berlin (HZB)); Prof. UNGER, Eva (Helmholtz-Zentrum Berlin (HZB)); Dr RAZEGHI, Yousef (Helmholtz-Zentrum Berlin (HZB))

Session Classification: Poster Session

Contribution ID: **32** Type: **not specified**

A Global Network for Data Stewards at the MWS – Chances and Challenges

Tuesday, September 30, 2025 2:15 PM (1h 30m)

The Max Weber Foundation –German Humanities Institutes Abroad (Max Weber Stiftung –Deutsche Geisteswissenschaftliche Institute im Ausland (MWS)) supports German research in the humanities abroad and acts as a link between the host countries and Germany. The institutes are distinguished –among other things –by the diversity of research focus, scope, methods and funding providers.

In November 2023, the MWS adopted Research Data Management (RDM) guidelines 1 in order to establish a foundation-wide RDM concept. Subsequently, data stewards were introduced and onboarded at the individual research institutes and offices abroad. Following the train-the-trainer-principle, a working group was set up for the data stewards. The staff was recruited internally at the institutes and offices abroad. Consequently, the background and expertise in the group are diverse, coming from different fields of library, research, IT and others.

One of the primary challenges was creating a shared understanding and knowledge regarding best practices for data management plans, data publication, teaching, and consulting. To address this, a communication infrastructure and workflows were established, functioning exclusively in a virtual environment. After more than one year of working together, the group never had a chance of in-person-meetings because of the distribution of institutes and offices worldwide. Even in a virtual work environment and all modern infrastructure in place, time zone differences between Europe, Asia and North America make it impossible to meet during a normal business day.

However, the data stewards meet regularly with the central RDM staff and meetings are recorded for those who cannot attend. We started off by building up an internal knowledge base and discussing discussing challenges and requirements arising from specific on-site conditions. The "MWS Best Practice Handbook on Research Data Management" 2 serves as the basis for training, information, and communication, providing further details on individual research data infrastructure offerings of the MWS for the public. This living handbook is intended for MWS researchers, data stewards, and other RDM professionals.

Additionally, we discovered that competences and knowledge regarding RDM vary significantly among researchers, depending on their background in Digital Humanities, Social Sciences, Arts, or History. This led to the planning a "quickstart package", consisting of a central RDM website and introductory video tutorials. The prototye of the website and video tutorials were prepared by the DHI Rome in cooperation with the central office, offering other individual institutes and offices abroad the possibility to adapt and reuse the work. The launch of the website is scheduled for Autumn 2025.

Before establishing the working group of data stewards, another crucial step was the introduction of the central Research Data Management Organizer (RDMO) platform for MWS researchers. DMP software, such as RDMO, supports researchers in creating data management plans required for third-party funding and planning the handling of research data at an early stage. A specially adapted questionnaire was developed to meet the requirements of MWS researchers, querying both superordinate project information and metadata on individual datasets. This information is also required, at least in part, in other research infrastructures. To avoid multiple entries, a connection to standardised interfaces is necessary.

Project-specific information in the questionnaire (RDMO), relevant for monitoring and reporting, as well as feeding into the publicly searchable MWS project database, only needs to be entered once by the researchers. Both databases are synchronized via an interface, and updates can be made within RDMO by completing or editing the questionnaire. This workflow gives all the power back

to the researchers who control the information of their research project from the beginning to the publication. They plan their projects in RDMO and keep editing the information in only one tool. The MWS project database harvests the relevant metadata fields automatically. A new project entry also triggers an email notification to the person responsible for the project database and the central RDM team. The latter initiates central RDM support, if needed, and often leads to the early clarification of open questions.

Looking back after more than a year, both opportunities and challenges have become apparent.

It has become clear that institutional RDM cannot be implemented by a person or a small team at the central office, but must be accompanied by trained staff on-site who can provide deeper insider knowledge and consider specific needs. Regular exchange, discussions, and training sessions have led to a wider acceptance of the RDM concept.

However, the central RDM staff and their data stewards still face challenges. The scalability of consulting services in a distributed infrastructre is a challenge that is handled in cooperation with several NFDI consortia (Text+, NFDI4Culture) and the RDM-helpdesk-network in the NFDI section CommonInfra. Task Forces have been established among the data stewards in the MWS working group to create training materials, to plan specific workshops and to advance the RDMO questionnaire to discipline-specific needs of researchers. The task to define which metadata should be collected in RDMO for specific topics,is also connected to the Interest Group in the Research Data Alliance (RDA) “Collections as Data”. In this particular context, we try to connect the idea of “Datasheets for Cultural Heritage data” to DMP-tools like RDMO.

There is still a long way to go. However, with the active group, creating a good ratio of consultant vs. researcher, an established workflow for information, communication and training, plus the technical infrastructure (digital publishing services, RDMO, collaboration), we are kicking off in a good spirit.

References

1 Max Weber Stiftung - Deutsche Geisteswissenschaftliche Institute im Ausland. (2023). Guidelines for the Handling of Research Data within the Max Weber Foundation. Zenodo. https://doi.org/10.5281/zenodo.10225705

2 Max Weber Stiftung - Deutsche Geisteswissenschaftliche Institute im Ausland. MWS Best Practice Handbook on Research Data Management. https://projects.academiccloud.de/projects/fdm-praxis-in-der-mws/wiki/fdm-praxishandbuch-mws (21.5.25)

Links:

https://www.maxweberstiftung.de/en/about-us.html

https://rdmo.maxweberstiftung.de

https://www.maxweberstiftung.de/en/research/projects/project-database.html

https://qed.perspectivia.net

https://www.rd-alliance.org/groups/collections-as-data-ig/plenary-participation/?application_id=188038

Abstract

Poster

Authors: RISSLER-PIPKA, Nanette (Max Weber Stiftung); GERSTNER, Eva-Maria (Max Weber Stiftung); GRÜNEWÄLDER, Jan-Peter (DHI Rom)

Session Classification: Poster Session

Contribution ID: **33** Type: **not specified**

Click, Upload, Confuse? Usability Testing of Dataverse and FAIRDOM-SEEK”

Tuesday, September 30, 2025 2:15 PM (1h 30m)

As research across disciplines becomes increasingly data-intensive, effective and user-friendly data management platforms are essential. Among the most widely adopted tools are Dataverse and SEEK, both designed to support FAIR (Findable, Accessible, Interoperable, Reusable) data practices. While each platform offers robust capabilities for data curation, sharing, and metadata management, there is a growing need to understand how they perform from a usability perspective, especially for researchers, data stewards, and clinicians who may not have specialized technical backgrounds.

This poster outlines the design and goals of an ongoing usability study comparing Dataverse and SEEK. Despite their widespread adoption in academic and clinical settings, little comparative research exists on how actual users interact with these platforms in everyday data workflows. Our aim is to fill this gap by evaluating how intuitive, efficient, and user-friendly these systems are for diverse user groups.

We are conducting a mixed-methods usability evaluation involving task-based testing, user surveys, and structured interviews. Participants will include researchers, data stewards, and clinicians with varying levels of experience in research data management. The study focuses on key user interactions such as dataset deposition, metadata annotation, access management, stress testing, and integration with external tools like ORCID and Git repositories.

The evaluation is guided by the following research questions:

How do users perceive the usability and learning curve of Dataverse versus SEEK?

What specific tasks or workflows present the greatest friction for different user roles?

How well do the platforms support users with limited technical backgrounds, such as clinicians or early-career researchers?

What usability improvements could enhance the adoption and effectiveness of each platform?

While results are still forthcoming, this poster will present our preliminary observations from early testing. By sharing our process, we aim to encourage similar usability-driven assessments in the broader research infrastructure community.

Abstract

Poster

Author: JABEEN, Hajira (Biomedical Informatics, University of Cologne)

Co-authors: Mr ZOUBIA, Oussama (University of Cologne); Dr AVILA-CALERO, Sergio (University of Cologne); Dr VORHAGEN, Susanne (Institue of Biomedical Informatics, University Hospital Cologne); Ms FU, Yu-Ting (Institue of Biomedical Informatics, University Hospital Cologne)

Session Classification: Poster Session

Contribution ID: **34** Type: **not specified**

Assessing Data Management and Compliance in Large Research Collaborations for Consent, Data Sharing, and Knowledge Transfer

Tuesday, September 30, 2025 2:15 PM (1h 30m)

This paper addresses the complex challenges of data management, compliance with legal and ethical standards, and knowledge transfer in large-scale collaborative research centers (CRCs), particularly in the context of linguistic data. The management of sensitive information, such as personal and health data, and the need for proper anonymization of audio and video recordings, pose challenges for researchers and data management practices. These legal issues, together with ethical requirements, complicate data sharing and reuse, which are essential components of adherence to FAIR (findable, accessible, interoperable, reusable) data principles. However, legal and ethical requirements must be complied to when it comes to study planning, data management plans, and further good research practices, even before the data have been collected (Jorschick et al. 2024). Structured research data management (RDM) can improve project documentation and collaboration efficiency (Mittal et al. 2023), but RDM strategies need to be tailored to the needs of the individual projects in order to allow for effective collaboration and reduced time spent on data management (Kanza & Knight 2022, Pascquetto et al. 2017).

In order to meet the issues outlined above, we introduce the development of an integrated platform to support structured RDM within a large Collaborative Research Centre (CRC) in linguistics. This platform includes (1) a comprehensive formal *ontology*, a conceptual schema, composed of definitions of abstract classes and their properties as well as the relation between them, which serves as the foundation to data and metadata representation according to FAIR principles, (2) a consent form *wizard* that supports researchers in setting up studies in accordance with RDM, e.g., by automatically generating legally compliant consent forms tailored to the specific requirements of each experiment, and (3) a *knowledge base* that records the collected information using the classes and relations of the ontology as well as guidelines for good research practices. In later phases, the combination of knowledge base and ontology will ground the platform's (meta)data, in that every dataset will be automatically linked to its participant-level consent, allowing researchers to query and monitor status for data sharing and (re-)use (Jorschick et al., 2024).

As the initial phase of platform development, we designed a semi-standardized interview protocol to assess each CRC project's data management and data protection practices. The guided questions trace the entire data life-cycle –from collection through storage, stewardship, legal-ethical safeguarding, and sharing –while also mapping the team's technical skills, workflows, and perceived training gaps. By confronting researchers with the interview questions, the method guided not only to evaluate current practice but also to raise awareness of legal and ethical obligations and identifies where training and infrastructure support were still needed. The collected data were then used to start refining the platforms knowledge base and the ontology, building on existing GDPR-compliant ontologies like GConsent (Pandit et al. 2018). We present knowledge graphs as a visualization of the information gained, representing the interrelationships between the projects in the CRC, such as the type and use of collected data, or research goals. These graphs allowed for the identification of potential synergies and connections between projects, promoting inter-project collaboration, and facilitates the clearer identification and continued pursuit of future research directions and management.

By integrating technical, legal, and ethical considerations into the research infrastructure, both the interview process and the ontology development for the platform aim to improve the overall sustainability and compliance of collaborative research. This presentation describes the development of this user-oriented solution. It will outline the structure of the platform, the role of the ontology, and the key benefits of implementing such a solution for large-scale interdisciplinary

research collaborations.

References

Jorschick, A., Schrader, P., & Buschmeier, A. (2024), What can I do with this data point? Towards modeling legal and ethical aspects of linguistic data collection and (re-)use. *Proceedings of the Workshop on Legal and Ethical Issues in Human Language Technologies at LREC-COLING 2024.* ELRA and ICCL, 47–51. https://aclanthology.org/2024.legal-1.8

Kanza, S. & Knight, N. J. (2022), Behind every great research project is great data management. In: *BMC Research Notes*, 15(20). https://doi.org/10.1186/s13104-022-05908-5

Mittal, D., Mease, R., Kuner, T., Flor, H., Kuner, R., & Andoh, J. (2023), Data management strategy for a Collaborative Research Center. In: *GigaScience*, 12, https://doi.org/10.1093/gigascience/giad049

Pandit, H. J., Debruyne, C., O'Sullivan, D., & Lewis, D. (2018). *GConsent –A Consent Ontology based on the GDPR*. Retrieved from https://openscience.adaptcentre.ie/ontologies/gconsent/main.html.

Pasquetto, I. V., Randles, B. M., & Borgman, C. L. (2017), On the reuse of scientific data. *Data Science Journal*, 16(8), 1–9. https://doi.org/10.5334/dsj-2017-008

Abstract

Poster

Authors: JORSCHICK, Annett (Universität Bielefeld); MOHAMMADI, Maryam (Universität Bielefeld); POLITT, Katja (Universität Bielefeld); BUSCHMEIER, Hendrik (Universität Bielefeld)

Session Classification: Poster Session

Contribution ID: **35** Type: **not specified**

Best practices for using tiered storage in research

Tuesday, September 30, 2025 2:15 PM (1h 30m)

Research today generates vast amounts of data which has to be analyzed, stored and, eventually, archived. This often necessitates the use of large storage systems, which consume considerable amounts of energy and natural resources. Therefore, it is paramount to use these storage systems in an efficient manner. To achieve this goal, it is necessary to use a tiered combination of different storage media, from fast SSDs for currently used data to long-living magnetic tapes for inactive data, and to set criteria on when data should be kept on a specific media and when it should be moved. In this talk I will present best practices for the efficient access to and use of different storage media in your day-to-day research and illustrate how you can set storage tier criteria that fit your research workflow.

Abstract

Poster

Author: MANGER, Natascha (KIT)

Session Classification: Poster Session

Contribution ID: **36** Type: **not specified**

Strengthening Research Data Management in Sport Science: A Community-Based Approach to Guidelines, Practices, and Challenges

Tuesday, September 30, 2025 2:15 PM (1h 30m)

In response to the growing importance of research data management (RDM) across scientific disciplines, the German Society of Sport Science (dvs) established an ad hoc committee on Research Data Management in October 2022. The committee's primary objective is to develop community-specific RDM guidelines that support sustainable, transparent, and open research practices within the diverse field of sport science. This initiative aligns with broader national and international efforts to foster responsible data stewardship and to address the practical challenges researchers face in their day-to-day work.

While RDM has become a central topic in research policy and funding, its implementation in everyday research activities remains uneven. Studies from various disciplines have highlighted a range of barriers, including limited awareness, lack of institutional support, concerns about data sensitivity, and uncertainty about legal and ethical frameworks. These challenges are particularly pronounced in interdisciplinary fields like sport science, where research often involves sensitive personal data, complex methodological designs, and diverse data types ranging from physiological measurements to video recordings and qualitative interviews.

To better understand the current landscape of RDM in sport science and to inform the development of practical, community-driven guidelines, the dvs ad hoc committee conducted a pilot survey between August and September 2023. The online survey was designed to capture a snapshot of RDM practices, knowledge, and needs among sport scientists in Germany. It consisted of three main sections: (1) demographic and professional characteristics of the participants, (2) current RDM practices across the research lifecycle, and (3) general knowledge and attitudes toward RDM.

A total of 122 researchers from all sections of the dvs participated in the survey, providing a broad and representative overview of the community. The results revealed a heterogeneous picture: while some researchers reported well-established practices in data documentation, storage, and sharing, others indicated significant gaps in knowledge and infrastructure. For example, many participants expressed uncertainty about how to handle sensitive data in compliance with data protection regulations, and only a minority reported using standardized metadata or data repositories. At the same time, there was a strong interest in improving RDM practices, with many respondents welcoming the idea of community-specific guidelines and training opportunities.

The survey findings underscore the need for targeted support mechanisms that address the specific challenges of sport science research. Based on the results, the dvs ad hoc committee has identified three core areas for future action:

1. Development of RDM Guidelines: The committee will draft practical, discipline-specific guidelines that reflect the realities of sport science research. These guidelines will cover key aspects such as data planning, documentation, storage, sharing, and long-term preservation, with a particular focus on ethical and legal considerations.
2. Handling of Sensitive Data: Given the prevalence of personal and health-related data in sport science, special attention will be given to the management of sensitive data. This includes guidance on informed consent, anonymization techniques, and secure data storage, as well as clarification of responsibilities under the General Data Protection Regulation (GDPR).
3. Promotion of Best Practices: The committee aims to collect and disseminate examples of suc-

cessful RDM strategies from within the community. These use cases will illustrate how researchers have integrated RDM into their workflows, navigated institutional and legal frameworks, and benefited from open data sharing.

By focusing on these three pillars—guidelines, sensitive data, and best practices—the committee seeks to foster a culture of responsible data stewardship in sport science. Importantly, the approach is grounded in the everyday realities of researchers and informed by empirical evidence from the community itself. This bottom-up strategy not only enhances the relevance and usability of the guidelines but also encourages broader engagement and ownership among researchers.

The pilot survey and the ongoing work of the committee offer valuable insights for other disciplines and professional societies facing similar challenges. They demonstrate how community-driven initiatives can bridge the gap between abstract RDM principles and practical implementation. Moreover, they highlight the importance of understanding disciplinary cultures and workflows when designing support structures for data stewardship.

In conclusion, this initiative represents a promising model for integrating RDM into the daily routines of researchers. By combining empirical research, community engagement, and practical guidance, it contributes to the broader goal of making research more transparent, reproducible, and sustainable. As the project evolves, it will not only support sport scientists in managing their data more effectively but also provide a foundation for future research on the facilitators and barriers to open and sustainable RDM practices.

Abstract

Poster

Author: Dr KELLER, Katja (KIT)

Co-authors: ROSE, Christian (KIT); KRÜGER, Melanie (Universität Hannover); Dr NIESSNER, Claudia (KIT)

Session Classification: Poster Session

Contribution ID: **37** Type: **not specified**

Lessons Learned: RDM Tools in CRC 1430

Tuesday, September 30, 2025 2:15 PM (1h 30m)

Research data management (RDM) in the scope of collaborative research centers (CRC) faces a very specific set of challenges. CRCs usually comprise researchers from different disciplines leading to very heterogeneous requirements with regard to RDM tools, storage capacity and data types involved. This is especially true for consortia in the field of biomedical research which include a plethora of different specialties, such as bio-informatics, medicine, structural biology and cell biology. Each of these fields has its own workflows, data formats and even scientific traditions. Furthermore, data is often stored in disparate locations. From this, the necessity arises to integrate and consolidate all information in a centralized platform and to establish a meaningful data structure. Learning from experiences with other CRCs, carefully selected tools have been introduced to address this issue and to ensure data within the CRC 1430 follows FAIR principles \[1\].

A modular framework of RDM tools with focus on free and open-source software has been established based on previous experiences \[2\] and comprises of the **electronic lab notebook** (ELN) *eLabFTW* \[3, 4\], the **collaborative cloud storage** solution *Nextcloud* and an **internal repository** based on *Dataverse* \[5\]. This framework already allows for an efficient data management workflow including primary documentation of experimental results and storage and sharing of raw data within the project. Due to its biomedical scope with focus on cell-state transitions, microscopic images play a significant role in the overall data of the CRC 1430. However due to their large storage requirements and very diverse file format situation a central platform for storage, file format conversion, annotation and sharing of microscopic data was urgently needed. Hence, the **imaging management platform** *OMERO* was introduced \[6, 7\].

Finally, the internal repository *Dataverse* was envisioned to integrate information from all of the above-mentioned sources in a single platform to allow easy and intuitive access to research data. To this end, a **custom metadata schema** specifically tailored to the data produced within the scope of CRC 1430 was developed \[8\]. This allows researchers to directly and easily reference related data in *OMERO*, *eLabFTW* and *Nextcloud* from within the internal repository as well as to search the CRC 1430 data based on relevant keywords, experimental methods or project collaborations. The metadata schema is continuously refined based on the involved researcher's feedback. Somewhat unexpectedly, the repository roll-out triggered an intense and constructive discourse within the CRC 1430 about publication practices, especially with regard to data publication. We therefor conclude that *Dataverse* as an internal repository not only serves as staging area for later data publication but actively catalyzes discussion and awareness for this topic.

To encourage upload of research data predating the launch of the internal repository, full-day **data upload events** were held. During these events, RDM staff was available on-site to assist with any issues and questions that arose in the process and to engage in discussions about possible enhancements to the repository and the underlying metadata schema. As a socializing and networking component, food and drinks were provided throughout the event which encouraged participants to stay and discuss even after their data uploads were finished. These events prove to be very successful and will be a corner stone in upcoming on-boarding processes of research groups and individual researchers.

Another successful initiative is the introduction of **Data Champion Awards** \[9\]. These are awarded to early-career researchers within the CRC for outstanding RDM practices, including innovative use of the above-mentioned tools. The overall goal is to create more incentives that will contribute to the establishment of good RDM as a routine for a new generation of researchers.

References

1. Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). doi: 10.1038/sdata.2016.18
2. Rehwald, S. Research Data Services for INF Projects. in *Posters des SFB-INF Workshops: Let's share! Gute Praktiken und Fallstricke im Datenmanagement in SFBs und Forschungsgruppen* (2025). doi: doi:10.5281/ZENODO.14904441.
3. Carpi, N., Minges, A. & Piel, M. eLabFTW: An open source laboratory notebook for research labs. JOSS 2, 146 (2017). doi: 10.21105/joss.00146
4. Timm, H., Wittkamp, A. & Beyer, S. Automatisiertes Deployment von Elektronischen Laborbüchern mit Ansible. in *E-Science-Tage 2021* (heiBOOKS, 2022). doi: 10.11588/heibooks.979.c13744
5. King, G. An Introduction to the Dataverse Network as an Infrastructure for Data Sharing. *Sociological Methods & Research* 36, 173–199 (2007). doi: 10.1177/0049124107306660
6. Burel, J.-M. et al. Publishing and sharing multi-dimensional image data with OMERO. *Mamm Genome* 26, 441–447 (2015). doi: 10.1007/s00335-015-9587-6
7. Lama, S., Koch, J., Schulze, N. & Rehwald, S. Sharing, storing and analysing cell image data in a Collaborative Research Centre using OMERO. (2023) doi: 10.5281/ZENODO.10069503
8. Minges, A. & Rehwald, S. CRC 1430 Metadata Block Configuration for Dataverse. *DuEPublico: Duisburg-Essen Publications online, University of Duisburg-Essen, Germany* (2025). doi: 10.17185/DUEPUBLICO/83389
9. Minges, A. & Rehwald, S. OMERO and Data Champions in CRC 1430. *Informationsmaterialien der Research Data Services an der Universitätsbibliothek Duisburg-Essen* (2025) doi: 10.17185/DUEPUBLICO/83183.

Abstract

Poster

Author: Dr MINGES, Alexander (Universität Duisburg-Essen)

Co-author: REHWALD, Stephanie (Research Data Services (RDS), University Library, University of Duisburg-Essen, Duisburg, Germany)

Session Classification: Poster Session

Contribution ID: **38** Type: **not specified**

Data Accessibility and Reusability Engine (DARE)

Tuesday, September 30, 2025 2:15 PM (1h 30m)

The widespread adoption of Electronic Laboratory Notebooks (ELNs), repositories, and other Research Data Management (RDM) and Research Software Engineering (RSE) tools has created a wealth of digital research artifacts. However, the lack of standardized formats and metadata hampers the efficient extraction, sharing, and reuse of these valuable resources. By developing a system to extract scientific data from diverse file formats, standardize it according to common standards, and publish the original data with rich metadata, we can unlock the full potential of these digital assets. The benefits of this approach are multifaceted. By achieving data interoperability, researchers can combine data from disparate sources, fostering novel insights and accelerating the discovery process. The automatic generation of machine-readable formats and metadata enables the integration of data into various systems, streamlining the research workflow. Moreover, the rich automated metadata generated through this process facilitates the reproducibility of research results, allowing scientists to verify and build upon existing findings with confidence. We show our first project results that aim to create a foundation for a more open, collaborative, and efficient scientific ecosystem. Harnessing the power of AI to create interoperable research data could be a major use case for generative AI for data stewards.

Abstract

Poster

Author: TRISTRAM, Frank (KIT)

Session Classification: Poster Session

Contribution ID: **39** Type: **not specified**

Self-Hosting Research Data Infrastructure with Kadi4Mat: A Practical Use Case for Managing Physics Data at IBPT, KIT

Tuesday, September 30, 2025 2:15 PM (1h 30m)

Abstract

Poster

Author: MASOUMI, Saeid (KIT)

Co-authors: MÜLLER, Anke-Susanne (KIT); GETHMANN, Julian (KIT: IBPT/LAS); Dr SCHUH, Marcel (KIT - IBPT); MEXNER, Wolfgang (IBPT)

Session Classification: Poster Session

Contribution ID: **44** Type: **not specified**

From Workshops to Community –Advancing Data Stewardship in Germany

Wednesday, October 1, 2025 9:20 AM (20 minutes)

Abstract

Talk

Authors: TRAUTWEIN-BRUNS, Ute (RWTH Aachen); DELLA CHIESA, Stefano (Leibniz Institute of Ecological Urban and Regional Development); DIERKES, Jens (Universität zu Köln); HAUSEN, Daniela (Heinrich-Heine-Universität Düsseldorf); KIRSCH, Stefan (Thuringian Competence Network for Research Data Management (TKFDM)/ Ernst-Abbe-Hochschule UAS Jena); MÄRZ, Hans (EVH Bochum)

Presenters: TRAUTWEIN-BRUNS, Ute (RWTH Aachen); DELLA CHIESA, Stefano (Leibniz Institute of Ecological Urban and Regional Development); DIERKES, Jens (Universität zu Köln); HAUSEN, Daniela (Heinrich-Heine-Universität Düsseldorf); KIRSCH, Stefan (Thuringian Competence Network for Research Data Management (TKFDM)/ Ernst-Abbe-Hochschule UAS Jena); MÄRZ, Hans (EVH Bochum)

Session Classification: Talks

Contribution ID: **45** Type: **not specified**

Data Organization Made Easy: Comprehensive Folder Structure Template for Early Career Life/Natural Science Researchers

Tuesday, September 30, 2025 2:15 PM (1h 30m)

Creating findable, accessible, interoperable, and reusable (FAIR) data and metadata is essential for researchers. Effective research data management (RDM) is crucial for achieving this, as recognized by funding organizations. Both data FAIRness and RDM rely on well-structured and documented data, including organized storage in a clear folder structure. Establishing a suitable folder structure early on is beneficial, yet junior scientists often struggle due to uncertainty about their needs. Ready-made templates can help establish an initial structure and foster good RDM habits. While templates exist for small-scale projects, a comprehensive folder structure tailored to doctoral research is lacking. This paper presents a folder structure template designed for PhD candidates in life and natural sciences, encompassing locations for code, results, figures, manuscripts, background information, and administrative paperwork. The template offers a practical solution for digital data organization, along with best practice RDM recommendations and metadata recording prototypes. By implementing a thoughtful structure early on and maintaining it consistently, researchers can manage their data more efficiently, leading to improved FAIR data outcomes and faster publication.

Abstract

Poster

Author: DEMERDASH, Yasmin (Institute of Molecular Biology (Mainz))

Co-authors: WILBRANDT, Jeanne (Leibniz Institute on Aging –Fritz Lipmann Institute); DOCKHORN, Ron (TU Dresden)

Session Classification: Poster Session

Contribution ID: **48** Type: **not specified**

Poster Pitches

Tuesday, September 30, 2025 1:50 PM (25 minutes)

Session Classification: Talks & Poster Pitches

Contribution ID: **49** Type: **not specified**

Poster Session

Contribution ID: **50** Type: **not specified**

RADAR - Using AI for better, FAIR research data management

Tuesday, September 30, 2025 4:45 PM (1 hour)

RADAR, the research data repository developed and operated by FIZ Karlsruhe, supports the secure archiving, publication and long-term availability of research data across disciplinary boundaries. Since going live in 2017, RADAR has been continuously developed to meet the increasing demands of Open Science.

The system offers comprehensive metadata support, persistent identifiers, semantic enrichment (e.g. Schema.org, FAIR Signposting), subject-specific terminologies via TS4NFDI and integrations with platforms such as GitHub, GitLab and WebDAV. Different operating variants (RADAR Cloud, RADAR Local) and subject-specific publication offerings (e.g. RADAR4Chem, RADAR4Memory) ensure a high degree of flexibility and broad connectivity to the needs of the scientific community.

As part of its continuous innovation work, we are currently testing AI-supported functions to further improve FAIR data practices. These include:

- **AI-powered metadata enrichment**, through automatic extraction of relevant keywords from existing metadata and documents linked to the dataset;
- **AI-based FAIRness checks** that provide targeted feedback and recommendations to optimize the FAIRness of datasets.

These functions are intended to support researchers and curators in meeting the growing demands on metadata quality and data responsibility - and at the same time reduce manual effort.

In the session, we will present our previous approaches and work and discuss them together with the participants. We are particularly looking forward to the exchange with participants who would like to contribute their own experiences in the use of AI in research data management.

Abstract

Presenters: BACH, Felix (FIZ Karlsruhe); BONATTO MINELLA, Christian (FIZ Karlsruhe)

Session Classification: Working Groups

Contribution ID: **51** Type: **not specified**

How Many Data Stewards Do We Need? Let's Talk Roles, Capacity, and Reality

Tuesday, September 30, 2025 4:45 PM (1 hour)

What does effective data stewardship look like in practice and how many stewards does it take? This interactive workshop invites participants to explore feasible role definitions, priorities, and team structures for data stewardship in research institutions of different sizes and capacities.

Presenter: DEMERDASH, Yasmin (Institute of Molecular Biology gGmbH (IMB))

Session Classification: Working Groups

Contribution ID: **52** Type: **not specified**

Data Stewards, let's work together: Shaping a living Community

Wednesday, October 1, 2025 9:40 AM (50 minutes)

Building on the outcomes of the Data Stewardship goes Germany (DSgG) workshops and the ongoing working group activities toward a sustainable and inclusive Data Stewardship Community in Germany, this workshop provides a space for reflection, exchange, and collaborative vision-building.

The Workshop will explore how a nationally recognised and globally connected data stewardship ecosystem can evolve from currently fragmented efforts. In this highly interactive workshop, participants are invited to share their perspectives on the future role and structure of the community, reflect on its potential added value and discuss how engagement across different initiatives can be fostered. The aim is to initiate concrete next steps for strengthening cooperation, avoiding parallel developments, and building a sustainable framework for exchange and joint action.

All data stewards and RDM professionals - whether already involved or newly interested - are invited to help shaping the future of Data Stewards in Germany. Together, we will explore how to turn shared goals into coordinated action and how the Data Stewardship Community can grow as a strong and visible force in the German research data landscape.

Abstract

Authors: TRAUTWEIN-BRUNS, Ute (RWTH Aachen); DELLA CHIESA, Stefano (Leibniz Institute of Ecological Urban and Regional Development); DIERKES, Jens (Universität zu Köln); HAUSEN, Daniela (Heinrich-Heine-Universität Düsseldorf); KIRSCH, Stefan (Thuringian Competence Network for Research Data Management (TKFDM)/ Ernst-Abbe-Hochschule UAS Jena); MÄRZ, Hans (EVH Bochum)

Presenters: TRAUTWEIN-BRUNS, Ute (RWTH Aachen); DELLA CHIESA, Stefano (Leibniz Institute of Ecological Urban and Regional Development); DIERKES, Jens (Universität zu Köln); HAUSEN, Daniela (Heinrich-Heine-Universität Düsseldorf); KIRSCH, Stefan (Thuringian Competence Network for Research Data Management (TKFDM)/ Ernst-Abbe-Hochschule UAS Jena); MÄRZ, Hans (EVH Bochum)

Session Classification: Working Groups

Contribution ID: **53** Type: **not specified**

Basisdienste in der NFDI: Nutzen für Forschende und Langzeitperspektive (streamed from FDM@Campus in Göttingen)

Wednesday, October 1, 2025 11:20 AM (20 minutes)

In this session, we will introduce the initiative Base4NFDI which aims to build so-called basic services for the German scientific landscape (and beyond). You will get insights on why such an approach is useful and how sustainable solutions are discussed.

Abstract

Author: JANDER, Melina

Presenter: JANDER, Melina

Session Classification: Talks (joint session with FDM@Campus)