SLUB
Wir führen Wissen.

TU9
GERMAN UNIVERSITIES OF TECHNOLOGY

Data Stewardship goes Germany

Workshop Proceedings

DSgG 2023, 25-26 September 2023, SLUB Dresden, Germany

Edited by
Marlies Krause, Arne Rümmler, Katharina Zinke, Janna Little,
Kay-Michael Würzner

Editorial

This volume contains the abstracts of the lectures and posters from the second 'Data Stewardship goes Germany' (DSgG) workshop, an annual workshop series organised by the Research Data Management Working Group of TU9 (Alliance of leading Technical Universities in Germany). The workshop took place on 25 and 26 September 2023 at the Open Science Lab at the Saxon State Library – State and University Library Dresden. The central theme of the workshop was the collective exploration of the increasingly established concept of data stewardship - an institutionally organised support system for all aspects of sustainable research data management. It is increasingly recognised that such support is neither trivial nor can be considered a basic competency of scientific staff in research institutions. Consequently, more and more scientific institutions and projects are relying on data stewards to establish sustainable work processes and raise awareness among researchers about the importance of good research data management.

The DSgG workshop series provides a framework for direct exchange within the research data management (RDM) community. It invites data stewards, research data (handling) officers, research software engineers, data experts, data curators, and data managers to discuss their professional realities and current development perspectives in Germany and beyond. The workshops also serve as a forum for presenting innovative solutions, addressing critical issues, and providing comprehensive assessments in the field of data stewardship.

Similar to its premiere in Braunschweig in 2022, this year's edition also garnered significant interest. Over 60 participants from across Germany and Switzerland came together to present and discuss ideas and solutions, as well as to conceptually exchange views on data stewardship in Germany. The well received keynote address[1] by Jens Dierkes (University and City Library Cologne) set the stage for the discussion, highlighting essential results from the DataStew project[2] and initiating further exploration of the status quo and perspectives for data stewardship in the German-speaking region. Barcamp discussions on topics such as job titles, tasks and roles, institutional placement, training, and the establishment of a data steward network, further served to explore these topics in more detail.

Thanks to numerous submissions, there were six exciting lectures and 21 poster presentations that focused on various aspects. The topics ranged from concrete tools for practical work to concepts for institutional embedding. A common theme of all contributions is acknowledging the gap between vision and reality in research data management. They propose a variety of strategies and tools to bridge this gap, emphasising the importance of research data management specialists, i.e., data stewards, throughout the research data lifecycle.

[1] Presentation Jens Dierkes: https://doi.org/10.5281/zenodo.10069006
[2] Final report of the DataStew project [German]: https://repository.publisso.de/resource/frl:6441397

Contents

Between institutional autonomy and structural requirements. Definition and tasks of data stewards at the University of Basel.

DOI: https://doi.org/10.5281/zenodo.10125156

Christina Besmer[1], Markus Bossert[1], Iris Lindenmann[1]

[1]University Library Basel, University of Basel

In Switzerland, the majority of universities started to establish research data management (RDM) support services for researchers at the latest in 2017, when the Swiss National Science Foundation (SNSF) made data management plans mandatory when submitting project applications. The topic got even more attention when swissuniversities, the Rectors' Conference of the Swiss Universities, published a national strategy on Open Research Data (ORD) in 2021 and an accompanying Action Plan, which specifies the measures of the ORD Strategy. The ORD Strategy promotes for Switzerland the establishment of structures for RDM and data stewardship in a collaborative way, while preserving the autonomy of institutions. In the strategy data stewardship is mentioned as a key element to achieve this aim.

At the University of Basel, various bodies offer RDM support to researchers. These include the University Library and various IT services, legal and ethical advice and other research support services at different institutional levels. Close collaboration between these bodies is crucial to ensure the best support for researchers across disciplines and for all aspects of the research data lifecycle. The Rectorate of the University commissioned an RDM Network to create structures for coordinated RDM support and to improve collaboration. The successful collaboration of this network, co-coordinated by sciCORE, the University's high-performance computing facility, and the University Library, began as a project in September 2017 and was consolidated by a Rectorate decision in September 2020. This decision also authorized a two years' pilot project for data stewardship, which aims to implement data stewards in each faculty and department, to bridge the gap between the subject-specific needs and the central services. The role of data stewards is assigned to people who already have a research-related support function at the departments. A train-the-trainer programme is part of the project in order to enable the data stewards, to fulfil their role as data stewards. Courses are provided on all aspects of RDM.

The pilot project started in autumn 2021. In 2022, as part of the ORD Action Plan, swissuniversities launched a call for the establishment of data stewardship at all Swiss universities. With the additional funding of swissuniversities, the project can now be continued, expanded and consolidated until the end of 2024.

The University of Basel has chosen a path that gives the faculties and departments as much freedom as possible in the institutional embedding of the data stewards. For example, the Faculty of Science has integrated data stewardship primarily into its IT services, while the Faculty of Theology appointed its coordinator of the doctoral program for this purpose.

In terms of the definition and the tasks of the Data Stewards, the University only sets minimal requirements and thus allows great flexibility in the implementation. The University defines data stewards as individuals who provide subject-specific support for RDM. They are the first point of contact for researchers in their respective subject area. In addition to providing direct advice and support, they also act as a link to the members and bodies being part of the RDM Network.

Overall, this leads to a high degree of diversity in the implementation of the data stewardship and in the range of skills of the data stewards at the University of Basel. Our experience so far shows that there are different needs and demands for RDM, not only between faculties, but even within faculties and their departments. The needs for RDM are by no means uniform within a discipline. Data stewardship roles have been filled by people working in their unit in areas as diverse as research, service or administration. In addition, the number of data stewards per faculty varies widely.

In our presentation, we will present the definition of data stewardship and the tasks of the data stewards at the University of Basel, as well as the structural framework within which data stewardship at the University of Basel operates. We would like to give an insight into the status of our project and will illustrate the implementation and embedding of data stewardship in the faculties and departments with practical examples. Our focus is on the Faculty of Science, but we will also address related general and interdisciplinary challenges and issues.

In the follow-up discussion, we would like to address the following questions:

- What are the advantages and disadvantages of an open approach to the definition and tasks of data stewardship?
- What are fundamental requirements for the role of data stewards? What are their essential tasks? What should be the limits of their responsibility?
- What are the challenges of collaborating on RDM within a very broad discipline such as the natural sciences?

The presentation is aimed at people who are involved in the establishment and implementation of data stewardship in their institutions, who are concerned with the definition of data stewardship and the role of data stewards, who themselves have experience as data stewards, or who simply have a basic practical or theoretical interest in these topics.

The Transformative Power of a Data Steward

DOI: https://doi.org/10.5281/zenodo.8378528

Stefano Della Chiesa[1]

[1]Leibniz Institute of Ecological Urban and Regional Development

In the context of a research institution, a Data Steward assumes a position of paramount importance, as it directly contributes to the efficient management and maximisation of the institution's valuable research data assets. The presence of a skilled and transformative Data Steward becomes instrumental in overcoming the silos that tend to emerge within institutions, allowing for the seamless integration and coordination of data-related efforts across various departments and disciplines. By bridging these silos, the Data Steward plays a crucial role in promoting the principles of open science and fostering a collaborative research environment. However, many institutions need help comprehending the precise nature and scope of the Data Steward role. This lack of clarity often leads to ambiguity in identifying skills and expertise. Additionally, integrating the Data Steward position within the organisational structure is also conducted heterogeneously across institutions due to the need for a well-defined framework for the Data Steward's responsibilities and reporting lines, which may hinder their ability to realise their transformative potential. Therefore, some aspects are crucial and need further discussion in this context, such as 1) The need to embed Data Stewardship into a strategic management position within the institution organigram, 2) The importance of the cross-cutting multidisciplinary competencies, 3) The essential requirement for a certain degree of research experience and 4) The project management skills.

1) Strategically embedding the Data Steward within the institution's organisational structure is crucial to establishing their authority, visibility, and influence. This ensures that research data management initiatives are integrated seamlessly with the institution's overall mission, governance, and research program and aligned with other strategic plans (Transfer, Publication, and Funding strategy, to mention a few). Research data should be considered a strategic asset; consequently, a Data Steward holds a strategic profile for a research institution. In this way, the institution acknowledges the importance of data governance and management as a strategic priority. A Data Steward fosters collaboration and ensures that data management considerations are integrated into the decision-making processes at all levels. Serves as a central point of contact for all matters related to data management ensuring compliance with relevant regulations and standards. Strategically embedding the Data Steward in the organigram also enables effective communication and coordination with different teams and individuals across the institution, bridging the gap between technical experts, such as IT professionals and data analysts, and non-technical stakeholders, including researchers, administrators, and policymakers. Additionally, they play a pivotal role in fostering a culture of data literacy and awareness among staff members across organisational units.

2) It is crucial to recognise that Data Stewards hold cross-cutting multidisciplinary profiles as their tasks span multiple departments and organisational units. Data Steward helps bridge the gap between technical and non-technical stakeholders. They act as a liaison between researchers and other actors such as IT, Knowledge transfer, DPO officers, Librarians, research management, and administration, facilitating effective communication and understanding of data-related requirements and fostering effective research data governance. This bridging function enables a shared understanding of data-related challenges, requirements, and opportunities, promoting a more holistic and collaborative approach to data management. This ensures that data infrastructure, tools, and systems are aligned with the institution's objectives and enable efficient data workflows across different actors.

3) Data Stewards with research experience bring several advantages to the role, as it allows a deep comprehension of the research data life cycle, starting from the project proposal to the data collection and methods conceptual design, operative data handling, analyses and output publication. They can fully understand the researcher's challenges, efforts, concerns, and specific needs. They can empathise with researchers, speak their language, and develop tailored solutions that meet their requirements. This understanding helps build trust and credibility among researchers, facilitating collaboration and data sharing.

4) Project management experience holds significant value for Data Stewards, providing them with the

essential skills to develop and proficiently develop and implement Research Data Management (RDM) strategies, governance, and stewardship practices. In the context of RDM, the utilisation of the Work Breakdown Structure (WBS) approach in designing Data Management Plans (DMPs) enables a detailed identification of research outputs, the delineation of work packages, deliverables, dependencies, and the determination of requirements throughout the entire data life cycle. This methodological approach allows for effective planning and coordination of efforts, resulting in the systematic implementation of the FAIR (Findable, Accessible, Interoperable, and Reusable) principles across the entirety of the data's life cycle. Additionally, monitoring FAIR or open science progress requires the development of Key Performance Indicators (KPIs). These indicators serve as measurable benchmarks that help evaluate the effectiveness and progress of Data Stewardship activities. Moreover, implementing RDM practices often involves significant changes in processes, policies, and workflows. Change management is another tool that supports Data Stewards in engaging with the RDM actor understanding their perspectives, managing expectations, and involving them in the change process. Hence, by leveraging project management principles and tools, Data Stewards can navigate the complexities of RDM and address all the challenges. In conclusion, a skilled and well-embedded Data Steward is vital in advancing data-driven decision-making, facilitating collaboration, and cultivating an environment of responsible and transparent research data management within academic institutions. By acknowledging the importance of these fundamental aspects, which arise from the deconstruction and elaboration of the three terms "Research," "Data," and "Management", institutions can fully harness the Transformative Power of Data Stewards, enabling effective data asset management, fostering open science practices, and optimising the transfer and impact of research outcomes, ultimately yielding broader societal benefits.

References

[1] Constanze Curdt, Jens Dierkes, et al. Data Stewardship im Forschungsdatenmanagement - Rollen, Aufgabenprofile, Einsatzgebiete: Überblick: 11. DINI/nestor Workshop, 16. und 17.11.2020. *Bausteine Forschungsdatenmanagement*, 3:70–81, 2021. https://doi.org/10.17192/bfdm.2021.3.8347.

[2] European Commission. *Strategic Research and Innovation Agenda (SRIA) of the European Open Science Cloud (EOSC)*. Publications Office of the European Union, 2022. https://doi.org/10.2777/935288.

[3] Konrad U. Förstner, Eva Seidlmayer, et al. Ergebnisse des Projektes DataStewForschung unterstützen - Empfehlungen für Data Stewardship an akademischen Forschungsinstitutionen, 2023. https://doi.org/10.4126/FRL01-006441397.

[4] Niklas K. Hartmann, Boris Jacob, and Nadin Weiß. RISE-DE – Referenzmodell für Strategieprozesse im institutionellen Forschungsdatenmanagement. Zenodo, 2019. https://doi.org/10.5281/zenodo.3585556.

[5] Patrick Helling, Felix Rau, et al. The Importance of Demand and Environment for Defining and Establishing the Role of Data Stewards. Zenodo, 2022. https://doi.org/10.5281/zenodo.6511185.

Data stewardship in CRC 1280 "Extinction Learning": From policy to dedicated workflows in an institutional data management system

DOI: https://doi.org/10.5281/zenodo.8388876

Marlene Pacharra [1], Johannes Frenzel [2], Andreas Schramm [2], Tobias Otto [3]

[1]CRC 1280 "Extinction Learning", Biopsychology, Ruhr University Bochum
[2]IT.SERVICES, Ruhr University Bochum
[3]Cognitive Psychology, Ruhr University Bochum

Data stewards of Collaborative Research Centers (CRCs) funded by the German Research Foundation (DFG) support the establishment and operation of sustainable infrastructures for the management of research data and metadata in their CRC (Schwandt 2019 [1]). There exist clear technical challenges such as the resource-saving adaptation of existing infrastructure systems to discipline-specific requirements instead of the new development of systems from scratch. However, also organizational factors such as lack of access to researchers' needs, unclear roles and governance, and lack of acceptance in the CRC were identified as major problems for the success of information management (INF) projects in which data stewards are often embedded (Brand & Dierkes 2020 [2]; Engelhardt & Kusch 2021 [3]). To address these critical issues, many useful steps can be taken to engage researchers and thereby increase their commitment to research data management (RDM) as well as the usability of (further) developed and implemented infrastructure systems. Based on the experiences of the CRC 1280 "Extinction Learning" successful measures and processes are presented that can ideally be extended to other CRCs or even the broader research institution.

Early-stage data stewardship activities can increase CRC researchers' investment in RDM and motivate them to use the infrastructures developed. Therefore, CRC 1280 initiated the still ongoing iterative process of needs assessment related to RDM even before the start of its first funding period in 2017. Since then, researchers from the CRC's groups as well as support staff have been meeting at irregular intervals to discuss RDM in the CRC. In this process, the support of experts from the infrastructure facilities (library, IT) of the Ruhr University Bochum (RUB) was invaluable e.g., in advising on metadata standards and technical options. To increase reusability of internally shared research data, researchers from all CRC groups agreed on a common folder structure to store data and a metadata schema with 16 fields that is partly mapped to Dublin Core terms and DataCite properties (Zomorodpoosh et al. 2023a [4]). Since no institutional data management infrastructure was available in 2017, a central network drive of the central IT was set up as an interim solution to share data and metadata, as well as Java open-source software applications developed that assist in (1) creating standardized metadata (Zomorodpoosh et al. 2023b [5]) and (2) searching existing metadata records (Diers et al. 2023 [6]).

RDM coordination processes in the CRC were formalized in the second funding period: A recognized Research Data Management Board was established to include researchers from the various disciplines, all status groups (early career researchers, principal investigators, support staff) and all CRC institutions. As a first step, the board developed a draft RDM policy on which all CRC members could comment. After several feedback loops, a revised policy was adopted unanimously by the CRC's principal investigators. The policy covers ethical and legal constraints, open science, data documentation and organization, quality assurance, collaboration and data sharing, and archiving (Pacharra et al. 2022 [7]). It defines the roles of all CRC groups, including their tasks and responsibilities, which increases transparency also regarding INF governance. While providing a framework for the entire CRC an important point for the researchers was that the adopted RDM policy refers to a dynamic internal knowledge base that can continue to adapt to changing RDM requirements. The adoption of the policy and the formal establishment of the board

This work was funded by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG), SFB 1280 Extinction Learning (316803389 – Project INF).

were important steps to truly anchor RDM in the CRC.

Adherence to the CRC's policy means adapting site-specific workflows and changing the local RDM culture - a process that can be time-consuming for researchers and often lacks clear, short-term incentives. To avoid frustration, it has paid off to develop and implement low-threshold formats to broadcast and ingrain RDM knowledge such as the CRC's knowledge base and Data Cleaning Days. Such personal and general data management measures flanked by custom developed tools, provide key incentives for researchers to incorporate and implement RDM into their workflows.

While using a shared network drive in combination with the developed apps has been a practical and intuitive interim solution for the CRC, only dedicated digital infrastructure can serve the requirements of researchers e.g., regarding data publication and 10-year preservation. To this end, needs assessments in the CRC have repeatedly revealed the desire for a central platform providing a uniform front end for all relevant RDM activities that enables role and rights management in accordance with the CRC's policy and provides simple authentication of CRC scientists from outside RUB via ORCID iD. To create sustainable solutions in this regard, the CRC has jointly worked with the central IT infrastructure and library services at RUB. In a first step requirements were compiled into a specification that was based on the CRC's consensus on a common folder structure and metadata, the experiences of the tools developed, as well as the CRC's policy. In a second step these specifications were used to design features that are currently being implemented into a new open-source, institutional data management system based on Hyrax 3[1]). As a result, the system will meet the CRC's policy requirements and reflect the CRC's organizational structure by a three-step roles and rights system. The associated three-step workflow ensures that highest quality data is being preserved or/and published according to the FAIR principles (Wilkinson et al. 2016 [8]) and the DFG's guidelines (DFG, 2022 [9]). With extended beta testing phases the system is continuously adapted to researchers' needs. We hope that by participation of researchers to create a system that fits into the CRC's RDM journey and is accepted by researchers, while helping them to meet policy requirements.

Overall, close collaboration between researchers and infrastructure is required to ensure both sustainability and discipline specificity of the solutions developed. In addition to this work on infrastructure, however, organizational anchoring of RDM in a CRC appears necessary: Through measures such as common policies, committee work, low-threshold training and consulting formats, and hands-on technical assistance RDM best-practices become part of researchers' daily routine. In addition, the necessary motivation and knowledge is built up to use the new infrastructure in a goal-oriented manner.

References

[1] Silke Schwandt. Forschungsdatenmanagement in Sonderforschungsbereichen: Best Practices. *Bausteine Forschungsdatenmanagement*, 2:22–30, 2019. https://doi.org/10.17192/bfdm.2019.2.8084.

[2] Ortrun Brand and Jens Dierkes. Failures and major issues. *Bausteine Forschungsdatenmanagement*, 2:89–96, 2020. https://doi.org/10.17192/bfdm.2020.2.8102.

[3] Claudia Engelhardt and Harald Kusch. 5.3 Kollaboratives Arbeiten mit Daten. In Markus Putnings, Heike Neuroth, and Janna Neumann, editors, *Praxishandbuch Forschungsdatenmanagement*, pages 451–476. De Gruyter Saur, 2021. https://doi.org/10.1515/9783110657807-025.

[4] Setareh Zomorodpoosh, Erik Diers, et al. Wiki of the SFB MetaDataApp project [Documentation], 2023. https://gitlab.ruhr-uni-bochum.de/sfb1280/metaapp_2/-/wikis/home.

[5] Setareh Zomorodpoosh, Erik Diers, et al. Metadataapp (v2.0). Zenodo, 2023. https://doi.org/10.5281/zenodo.8040229.

[6] Erik Diers, Setareh Zomorodpoosh, et al. DatabaseApp (v2.0). Zenodo, 2023. https://doi.org/10.5281/zenodo.8039273.

[7] Marlene Pacharra, Nina O. C. Winter, et al. Research Data Management Policy of the collaborative research centre SFB 1280 "Extinction Learning" (v1.0). Zenodo, 2023. https://doi.org/10.5281/zenodo.8004432.

[8] Mark D. Wilkinson, Michel Dumontier, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1):160018, 2016. https://doi.org/10.1038/sdata.2016.18.

[9] Deutsche Forschungsgemeinschaft. Guidelines for Safeguarding Good Research Practice. Code of Conduct. Zenodo, 2022. https://doi.org/10.5281/zenodo.6472827.

[1] Samvera community (2021). Hyrax (v3.0) [Source code]. Accessed 2023-07-13. https://github.com/samvera/hyrax

Standardized Data Interchange between ELNs using ELN File Format – One step towards “I” in FAIR principles

DOI: https://doi.org/10.5281/zenodo.10069050

Manideep Jayavarapu[1], Michael Selzer[1], Britta Nestler[1,2,3]

[1] Institute of Nanotechnology (INT-MSS), Karlsruhe Institute of Technology (KIT), Karlsruhe

[2] Institute of Applied Materials (IAM-MMS), Karlsruhe Institute of Technology (KIT), Karlsruhe

[3] Institute of Digital Materials Science (IDM), Karlsruhe University of Applied Sciences, Karlsruhe

As part of many research activities, Electronic Lab Notebooks (ELNs) are widely used in documenting the results of experiments and investigations, simulations, etc. With the existing ELN solutions, one can manage, store and publish the results of the research activities. But the data sharing between different ELNs was not possible. In this work, we introduce The ELN File Format, which uses RO-Crate [1] inside a ZIP archive to address the issue of data interoperability i.e., the exchange of (meta)data between the ELNs. The concept of ELN File format is being developed as part of an ELN Consortium[1]. The ELN File format is a structured archive format for the exchange of results and data, which can be read and produced by Electronic Lab Notebooks (ELNs).

1. Background The shift to open science has raised the demand for and the significance of publishing re-search results [2]. It is suggested that the publication of these workflows and data should ad-here to the FAIR principles [3]; namely, Findable, Accessible, Interoperable, and Reusable. To meet the challenges of creating FAIR data, many research groups and institutions are concerned with the design and implementation of tools for gathering, finding, and sharing data. The ability for other researchers to use this data with different ELNs is the component that is still lacking.

The deployment of Findable Accessible Interoperable Reproducible (FAIR) objects made it possible by integrating RO-Crate (or ELN File Format) implementations in the systems like ELNs.

2. Methodology The main goal of this ELN File Format is to standardize the exchange format and to enable the exchange of information or data between different ELNs. This file format is developed based on RO-Crate.

A generic packaging format called RO-Crate is an open and lightweight approach for bundling (or packaging) research data that contains datasets and associated descriptions(i.e., metadata) following FAIR Linked Data standards. The format is based on annotations in JSON-LD from schema.org [4], which enables comprehensive metadata representation. This JSON-LD format allows to link to external information which makes it flexible and machine readable. The metadata of the data entities as well as contextual entities such as people, software etc., as a series of linked JSON-LD objects using common vocabularies, mainly schema.org.

3. Implementation in Kadi4Mat Kadi4Mat (Karlsruhe Data Infrastructure for Materials Science)[2] is an open-source research data infrastructure being developed at IAM-MMS of the KIT. It has been kept as generic as possible, though the software has been developed with materials science in mind. Kadi4Mat is best described as a virtual research environment that combines the features of ELNs and repositories. The main types of resources available in Kadi4Mat are Records, Collections, and Templates, out of which records are the central component of Kadi4Mat. A record can represent any type of digital or digitalized object, e.g. arbitrary research data, samples, experiments, or even individual processing steps. In addition to the basic metadata of a record, each record can contain an arbitrary amount of generic metadata.

[1]The ELN Consortium: https://github.com/TheELNConsortium

[2]Kadi4Mat landing page: https://kadi.iam.kit.edu/

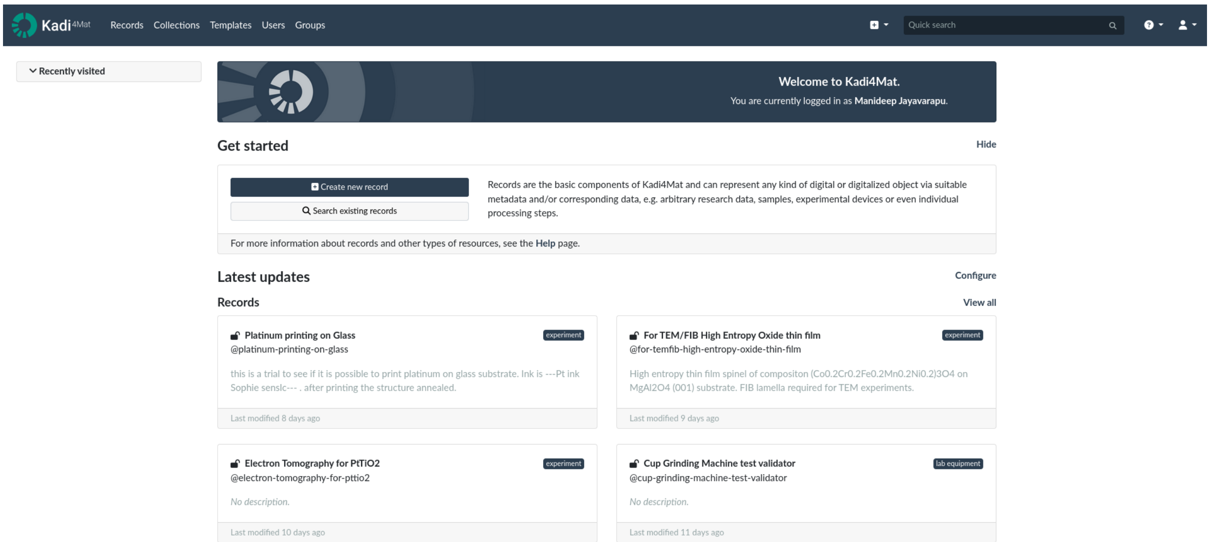

Figure 1 Screenshot of Kadi4Mat's web interface

The RO-Crate export (or ELN file format) of a record or collection in Kadi4Mat [5] is developed as part of the ELN Consortium[3]. The goal of this consortium is to define some common specifications and exchange file formats to allow the interchange of data and metadata between different ELNs.

Record Example:

Name	Size	Compressed
demo_record	1 Folder, 1 File	
demo_record	1 Folder, 2 Files	
files	4 Files	
demo_record.json	8,7 KiB	8,7 KiB
demo_record.ttl	4,4 KiB	4,4 KiB
ro-crate-metadata.json	4,8 KiB	4,8 KiB

Collection Example:

Name	Size	Compressed
dummy-collection	2 Folders, 1 File	
dummy-title	1 Folder, 2 Files	
files	9 Files	
dummy-title.json	18,2 KiB	18,2 KiB
dummy-title.ttl	7,0 KiB	7,0 KiB
instrument-data	1 Folder, 2 Files	
files	3 Files	
instrument-data.json	14,3 KiB	14,3 KiB
instrument-data.ttl	4,1 KiB	4,1 KiB
ro-crate-metadata.json	10,2 KiB	10,2 KiB

Figure 2 The Folder structure of the ELN File Format in Kadi4Mat

The above folder structure follows the archive structure[4] in the ELN File Format specification. All the data of a research activity is bundled in a ZIP archive along with a JSON-LD file named ro-crate-metadata.json, which describes data entities (i.e, payload files/directories) and contextual entities (i.e, person, software, etc.) in a structured form of JSON.

Along with the data entities (i.e., payload files/ directories) of a resource in Kadi4Mat, the JSON export and RDF (Turtle format) export of the resources are also included in the archive, which holds most of the metadata of each resource.

4. Specification for the ELN File Format The ELN File format is a ZIP archive that follows all the standard ZIP specifications and can be readable by zip utilities. This archive format is a zipped RO-Crate, with a '.eln' extension, which was decided in ELN Consortium. The core of this ELN File format is a JSON-LD file named 'ro-crate-metadata.json' which follows RO-Crate specification [1] and is machine readable.

The rest of the archive is composed of zero or more folders that each describe one resource (i.e., a record in Kadi4Mat) or a set of resources (i.e., a collection in Kadi4Mat). Thus, the archive can accommodate one or several sets of data.

[3] The ELN Consortium: https://github.com/TheELNConsortium

[4] Structure of the contents in ELN File Format: https://github.com/TheELNConsortium/TheELNFileFormat/blob/master/SPECIFICATION.md#structure-of-the-archive

5. Conclusion The presented work provides a data exchange format between different ELNs. By using the ELN File Format, one can bundle the data along with the metadata which can easily be shared between different ELNs.

References

[1] Peter Sefton, Eoghan Ó Carragáin, et al. RO-Crate Metadata Specification 1.1.2, Jan 2022.

[2] Eoghan Ó Carragáin, Carole Goble, et al. A lightweight approach to research object data packaging. bioinformatics open source conference (bosc), ismb/eccb 2019, basel, switzerland. Zenodo, 2019. https://doi.org/10.5281/zenodo.3250687.

[3] Peter Sefton. FAIR Data Management; It's a lifestyle not a lifecycle., 2021. http://ptsefton.com/2021/04/07/rdmpic/.

[4] R. V. Guha, Dan Brickley, and Steve Macbeth. Schema.org: Evolution of structured data on the web. *Commun. ACM*, 59(2):44–51, 2016. https://doi.org/10.1145/2844544.

[5] Nico Brandt, Lars Griem, et al. Kadi4mat: A research data infrastructure for materials science. *Data Science Journal*, Feb 2021. https://doi.org/10.5334/dsj-2021-008.

Automatic Generation of Interoperable Scientific Data – A Use Case with Electrochemistry Data

DOI: https://doi.org/10.5281/zenodo.10069113

Albert K. Engstfeld[1], Johannes Hermann[1], Nicolas Hörmann[2], Julian Rüth[3]

[1] Institute of Electrochemistry, Ulm University, Ulm
[2] Theory Department, Fritz-Haber-Institute, Berlin
[3] Julian Rüth GmbH, Güntersleben

Storing scientific data according to the FAIR principles[1] requires annotating such data with additional information or metadata. Research data consists, for example, of single values (the temperature at a specific time), time series data (observation of temperature with time), or images, all providing information on the state of a system at a particular time. The additional information on the system, in turn, usually must be stored separately, such as in a laboratory notebook. While text-based descriptions are valuable resources, data-interchange formats, such as JSON, YAML, or XML, are desirable for further data processing by a machine. Annotating research data is usually challenging, especially when many parameters should be stored, such as information on the users, equipment, software, involved materials, or type of measurement. In that sense, automation is highly desirable.

This is the topic of this work, where we[2] present solutions to annotate time series data, the automated generation of related metadata, and show use-case examples of such annotated files in the research area of electrochemistry.

To annotate research data, we developed "autotag-metadata"[3], which observes a local folder for file changes and writes a file with metadata provided in a template YAML file in the same folder. These YAML files can contain any structured information, preferably based on an existing metadata schema of the respective research area. We developed a simplified metadata schema containing general objects for any research area and objects specific to the electrochemical community[4]. Since the data acquired with particular instruments are often in very different and sometimes proprietary file formats, we convert the data into CSV files and store the information on the CSV and the YAML file in a Frictionless datapackage[5] (JSON file). Besides the metadata from the YAML, the datapackage also contains information on the columns in the CSV, such as the units. Since scientific results are often compared with and validated against data in scientific publications, but such data is usually only available as a plot in a scientific publication, we developed "svgdigitizer", which extracts the data from carefully prepared scalable vector graphics and generates a similar datapackage as for the experimental data [1].

Both the experimental as well as extracted datapackages can now be used with software solutions provided by Frictionless to do data validation or create static websites. To interact with our datapackages, including units associated with the data in the CSV and additional metadata, we augmented the Frictionless Python API with the "unitpackage" module [2]. It provides simple access to the metadata associated with a CSV resource and allows unit conversion or visualizing of the underlying data. On the other hand, it can also be used to generate a collection of unitpackages, which can be pictured appealingly on a website[2].

Finally, possible further applications can be discussed, such as the automated generation of electronic lab notebook entries. In total, storing scientific data with associated metadata as datapackages in the local file system or a remote repository provides a highly interoperable approach to working with scientific data in general.

References

[1] Albert K. Engstfeld, Johannes M. Hermann, et al. svgdigitizer. Zenodo, 2022. https:

[1] https://www.go-fair.org/fair-principles/
[2] https://www.echemdb.org, https://github.com/echemdb
[3] https://echemdb.github.io/autotag-metadata/
[4] https://github.com/echemdb/metadata-schema
[5] https://framework.frictionlessdata.io/, https://frictionlessdata.io/

//doi.org/10.5281/zenodo.5874747, https://echemdb.github.io/svgdigitizer/.

[2] Albert K. Engstfeld, Johannes M. Hermann, et al. echemdb/unitpackage: 0.7.1. Zenodo, 2023. https://doi.org/10.5281/zenodo.7942320, https://echemdb.github.io/unitpackage/.

From Planning to Practical Implementation and Support: Fusion of Multiple Open Source Applications for Data Management Workflows in Psychology and Neuroscience

DOI: https://doi.org/10.5281/zenodo.10069259

Julia-Katharina Pfarr[1,2], Stefan Lenze[2,3],Christian Berger[2], Ortrun Brand[4], Dominik Endres[5]

[1] Philipps-University Marburg, Department of Psychology, General and Biological Psychology
[2] Philipps-University Marburg, University Computer Center, Research-related Applications
[3] Philipps-University Marburg, Department of Physics, Neurophysics
[4] Philipps-University Marburg, eResearch-Service Center
[5] Philipps-University Marburg, Department of Psychology, Theoretical Cognitive Science

Data Management has become an incredibly important term and concept which is invariably present in every workspace. When entering the "Data Management World" professionally, one soon realizes that this world is vast with a lot of potential ways to navigate through. Specifically in research data management, the challenge of data reproducibility, sharing, and the training of researchers in data management adds even more dimensions. Moreover, in the field of psychology and neuroscience potential sensitive data has to be considered as well. Finding an intersection between researchers' needs, their skills in data management, data access restrictions, and proper funding is thus a complex but at the same time determined challenge.

At the University of Marburg this challenge is taken on by the team of the so called "Data Hub"[1]. The team consists of multiple people with different responsibilities, backgrounds, and affiliations (project management staff, scientific staff, data stewards, data scientists, technical administrative staff, located in Marburg and Gießen). The Data Hub is funded by The Adaptive Mind (TAM)[2] and supported by the infrastructure project of the SFB135[3] (NOWA[4]), which are consortia in the field of psychology and neuroscience, with over 50 involved PIs, based in several locations in the federal state of Hesse, Germany. To date, two Data Stewards are responsible for training and support of the researchers in TAM and the SFB135. Along the way of development of the Data Hub, the team faced obstacles and found solutions regarding different requirements.

Requirements for the data management tools: Even though the fact that the kind of research data to manage in the two consortia is restricted to the field of psychology and neuroscience, a major challenge (besides the ones already mentioned above) is the need of harmonizing heterogeneous data. The scope reaches from behavioral data (e.g., questionnaires, interviews) over experimental and analysis code to output data from a variety of hard- and software (e.g., eye trackers, magnetic resonance scanners, EEG etc.). The data management workflow needs to be applicable to heterogeneous in- and output data, project size, and number of researchers involved. Furthermore, tools need to be able to integrate those heterogeneous data by utilizing a harmonizing standard in the field (i.e., BIDS [1]). To fulfill the need of reproducibility of research findings, as heavily discussed in the concerned fields of science, an integration of version control and provenance tracking should be available.

Requirements for the researchers: The profile of a researcher includes many skills: creative but critical thinking, advanced statistics, programming, writing up and presenting scientific work, project management, and, of course, vast knowledge of theories and previous work in their respective field of science. Undeniably, data management represents a necessary skill as well. However, education in data management during studies is highly unusual and thus researchers are left on their own to gain those skills. This leads to a mentality of "I don't need data management; I know my work.".

[1] https://www.uni-marburg.de/de/forschung/kontakt/eresearch/unsere-angebote/services-thematisch/data-hub, https://www.theadaptivemind.de/research/data-hub.html
[2] https://www.theadaptivemind.de/
[3] https://www.sfb-perception.de/index-en/
[4] https://www.sfb-perception.de/projects/inf.html, https://sfbs.pages.uni-marburg.de/sfb135/nowa/nowa.site/

While this is plausible, it is not applicable to the growing world of digital data, collaboration and replication. Thus, researchers need to have a basic understanding and training in research data management. This involves multiple training sessions to not only know how to use the tools but also to know why the tools behave this way and how to troubleshoot.

Requirements for the Data Stewards: The responsibilities of data stewards are diverse. On the one hand, data stewards need to have an understanding of topics such as software development and technical hardware concepts. On the other hand, data stewards need to have an understanding of how research is done. What is more, even though workflows in research share common phases, the ways of collecting, processing, and storing data is more or less individual to the specific research area. This requires not only a basic understanding of research work but also research area specific knowledge (i.e., what does the kind of data actually mean? Which research questions can be answered with this kind of data and how?). Only then, data stewards can serve the intersection of software and research world. What is more, serving this intersection means conceptualizing trainings for the researchers in data management. For this, again, data stewards must have an understanding at which point to pick up the researchers: How much background knowledge about software do they have and how much do they really need? Which functions of the software are necessary and which ones can be skipped because they'll never apply to the researchers work? Do they need a lot of hands-on practice or is the concept enough?

In our presentation we will first briefly introduce the Data Hub of the University of Marburg and its technical architecture. We will then present the data management tools utilized in the Data Hub (i.e., GIN [2], DataLad [3], Git [4], GitLab [5], JupyterHub [6]). We will introduce their main functions and specifically focus on reasons for using these tools compared to other popular research data management tools. Further, we will show how these tools are interconnected, i.e., the research data management workflow of the Data Hub. That followed, we will outline the challenges for both the researchers as well as the Data Stewards regarding training and support and maintenance of the services. Furthermore, we address the question of how different (or even no or at least no academic or standardized) education in "How to be a Data Steward" affects the daily work of a data steward.

Acknowledgements and Funding

The adaptive Mind

NeurOscientific Workflow Assistance NOWA

Funded by Deutsche Forschungsgemeinschaft German Research Foundation

References

[1] Krzysztof J. Gorgolewski, Tibor Auer, et al. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, 3(1):160044, 2016. https://doi.org/10.1038/sdata.2016.44.

[2] Achilleas Koutsou, Michael Sonntag, et al. GIN: Open platform and services for efficient research data management, collaboration and data publication. INCF Neuroinformatics Assembly, 2021. https://neuroinformatics.incf.org/node/216.

[3] Aref Kalantari, Michał Szczepanik, et al. How to establish and maintain a multimodal animal research dataset using datalad. *Scientific Data*, 10(1):357, 2023. https://doi.org/10.1038/s41597-023-02242-8.

[4] John D. Blischak, Emily R. Davenport, and Greg Wilson. A quick introduction to version control with git and github. *PLOS Computational Biology*, 12(1):1–18, 2016. https://doi.org/10.1371/journal.pcbi.1004668.

[5] Prithwiraj Choudhury, Kevin Crowston, et al. Gitlab: work where you want, when you want. *Journal of Organization Design*, 9(1):23, 2020. https://doi.org/10.1186/s41469-020-00087-8.

[6] Michael Milligan. Interactive hpc gateways with jupyter and jupyterhub. In *Proceedings of the Practice and Experience in Advanced Research Computing 2017 on Sustainability, Success and Impact*, PEARC '17. Association for Computing Machinery, 2017. https://doi.org/10.1145/3093338.3104159.

A hands-on workshop to propagate knowledge and positive experiences with RDM in natural sciences

DOI: https://doi.org/10.5281/zenodo.10134464

Johannes Vosskuhl[1,2], Stefan Harfst[1,2]

[1] Scientific Computing – Research Data Management, School of Mathematics and Natural Sciences, Carl von Ossietzky University, Oldenburg

[2] SFB 1372 – Magnetoreception and navigation in vertebrates

Awareness and knowledge about research data management (RDM) as well as the application of the techniques is highly heterogeneous among researchers in natural sciences and mathematics. Also, the required techniques – and relevance of concepts – are highly diverse. To generally and individually improve the handling of research data, we have developed a modular workshop and hands-on session using realistic data. A secondary aim is to increase individual motivation to professionally manage and share own data by overcoming negative prepossessions against RDM and data sharing. During the workshop, general concepts of RDM are interactively developed together with the participants, then these concepts are practiced and deepened in a hands-on part and third the learned skills are transferred into the daily routines of the participants in a guided discussion format.

For the hands-on part, participants are split into two groups, each of which works on a different set of realistic practice data with the same task: First, to find as many details as possible about the data and use it as metadata. Second, to structure the data in a meaningful way, for example by renaming files, sorting files into folders, etc. Third, to upload the data and describe it concisely in the institutional data repository. After the successful submission, both teams are instructed to look at the other team's upload, try to identify similarities between the datasets and discuss options for additional data analyses with the combined datasets.

A special emphasis was put on the design of the practice data. It comprises folders of raw data, processed data, figures, unused files, analysis-scripts and dissemination products from two neuropsychological experiments. Together the material contains all relevant metadata in file names, read-me files, and the folder structure to fully describe the dataset in rich detail. It is however arranged in a sub-optimal but realistic way at the end of a research project. The datasets consist of a mixture of files in different formats (.csv, .mat, .vhdr, etc.), with varying degrees of FAIRness (e.g. proprietary formats, specialized code, readme-files) and realistic filenames (oldWorkflow.m, allSamples.csv, filtered_01-80_Sept2019.mat). To adapt task difficulty, files can be pre-sorted and re-named to highlight relevant information. All these problems and imperfections of the dataset shall be identified and are addressed at the end of the hands-on session.

The hands-on part is wrapped up in a guided discussion in which the participants shall evaluate the quality of the datasets with respect to the FAIR principles. In a second step, the learned concepts and experiences are transferred to the datasets usually handled by the participants.

After the workshop, participants understand that RDM generates benefits and basic techniques are easy to integrate in day-to-day work routines. They have experienced a feeling of success and thus associate positive emotions with data sharing. The realistic quality of the practice data illustrates that data must not be in perfect shape to be shared and thus increase the motivation to actually share own data.

Diversity of data stewardship within a federated research infrastructure

DOI: https://doi.org/10.5281/zenodo.8383283

Steffi Genderjahn [1], Meistring, Marcel [1], Roland Bertelmann [1]

[1] Helmholtz Association, Helmholtz Open Science Office

As the requirements for research data and its handling are constantly changing, the range of tasks and topics in scientific institutions is also becoming more diverse and larger. This poster will outline the diverse developments, challenges and difficulties surrounding the topic of research data in a federated, decentralized research infrastructure such as the Helmholtz Association from a cross-institutional and cross-disciplinary perspective. As interest in managing research data increases in many parts of the research ecosystem, policies and infrastructures are being implemented in more and more contexts.

Data management teams can consist of data stewards, librarians, software developers, and scientists from various research areas. They create conditions to ensure digital reproducibility and best practices for research data. Data stewards collaborate with researchers in a research ecosystem on different levels. In a large scientific organization, they can be deployed for specific research disciplines, priority initiatives, major projects or in infrastructure-related services. Data stewardship includes responsibility for data and coordinating the necessary activities for data exchange processes, including the collection, storage, backup and use of data [1] also with regard to licensing agreements and ethical standards. Therefore, data stewards assist researchers in practical applications, enabling scientists to benefit from research processes.

It is a challenge to make different data sets from the scientific process of various disciplines findable and reusable. A central tool is the use of metadata to enrich research data with standardized and machine-readable information. To coordinate this process, central infrastructures such as the Helmholtz Metadata Collaboration (HMC) and the Helmholtz Federated IT-Services (HIFIS), bundle scientific expertise from various research areas on the subject of metadata or build and maintain an IT infrastructure for networking research areas and centers. For instance, they offer consulting services, provide information about suitable technical solutions, and ensure uniform scientific standards. These structures create synergies and increase the efficiency in the use of resources and knowledge throughout the research association. With these supporting infrastructures, data stewardship can benefit from a variety of services and developments, related to culture change in research data provision and reuse. At the same time, it is an overarching goal to provide research data with metadata and make it FAIR (Findable, Accessible, Interoperable, Reusable) [2]. Here, the data stewards can support the digital workflow from data generation to documentation to analysis and archiving by using appropriate interfaces or linking via persistent identifiers (PIDs) on an infrastructural level.

An additional challenge in the field of data stewardship is the wide range of developments at national and international level with regard to a standardized approach to data. Data stewards also act as an interface between scientists and scientific support units, such as infrastructures internal to the institution (IT department or libraries) or even to national infrastructures such as the National Research Data Infrastructure (NFDI).

Guidelines and policies were drawn up to coordinate the handling of research data and software in the Helmholtz Association. The requirements for the research data cycle depend on the discipline and also differ in the individual research areas. With the ongoing digitization of research and teaching, the number of research software solutions for scientific knowledge acquisition is constantly increasing. In many cases, the provision of appropriate software is of great importance for the reproducibility of data analyzes and for the subsequent use of the respective research data. For this reason, suitable referencing and provision of the software used or developed is essential. The requirements for a data steward for research software also include licensing issues. In this case, cross-research consulting offers can support the work of a data steward.

In addition, facets of open science will be considered as an essential component of FAIR data stewardship. In order to organize open science according to the principles of openness, transparency, quality assurance, networking and sustainability, researchers will be supported in the implementation by providing infrastructures, services, consulting and training. In perspective, central open science infrastructures such as publication platforms, in-house publishers, repositories and consulting services are to be established and expanded. The requirements for a cultural change with regard to open science are set out in the Helmholtz Open Science policy[1]. A data steward plays a central role in communicating with scientists in particular and can accompany cultural changes.

In closing, current developments on research evaluation and incentives play a central role in the transformation of the scientific enterprise. For an established, sustainable research data management (RDM), the recognition of the research performance of scientists is another important component. In this context, the evaluation of new indicators for data and software products is necessary. Hereby, pure metrics are not a solution, quality issues should be central. In general, data stewardship can support research assessment in several ways. Data stewards can provide insights into the researcher's data management practices and help them identify areas where additional support is needed[2]. At the same time, they can help establish best practices for data management and develop services to support them. Data stewards can assist to evaluate the quality of data management plans (DMPs) and provide feedback to researchers on how to improve them[3]. In addition, they can support identifying areas where policies and guidelines need to be developed or updated. In their work, data stewards gain insights into the use of data infrastructures and can determine the need for services or tools that researchers require. They can work in both discipline-specific and general infrastructure areas, and embedding them in a data management team is one way to encourage transfer and foster changes in assessment practices.

The requirements for data stewardship should be considered in their diversity based on the federated, decentralized research association. Science supportive units offer a broad portfolio to support sustainable and FAIR RDM as well as export and publishing routines. The tasks of data stewards depend on the embedding in the scientific ecosystem. Data stewardship is intended to support cultural change in research data and software, as well as Open Science at discipline-specific or general infrastructure level.

References

[1] W. van Donge, N. Bharosa, and M.F.W.H.A. Janssen. Data-driven government: Cross-case comparison of data stewardship in data ecosystems. *Government Information Quarterly*, 39(2):101642, 2022. https://doi.org/10.1016/j.giq.2021.101642.

[2] Martin Boeckhout, Gerhard A. Zielhuis, and Annelien L. Bredenoord. The FAIR Guiding Principles for data stewardship: fair enough? *European Journal of Human Genetics*, 26(7):931–936, 2018. https://doi.org/10.1038/s41431-018-0160-0.

[1] Helmholtz-Gemeinschaft, 'Helmholtz Open Science Policy. Version 1.0. Approved in the 119th General Assembly of the Helmholtz Association on 20-21 September 2022', 2022, https://doi.org/10.48440/os.helmholtz.056

[2] RDMkit, https://rdmkit.elixir-europe.org/data_steward_research.html, accessed 22 July 2023.

[3] Data Stewards: Overview - The Turing Way, https://the-turing-way.netlify.app/collaboration/research-infrastructure-roles/data-steward.html, accessed 22 July 2023

The story from “unstructured” research data management towards the establishment of the central organisation UR Data Hub

DOI: https://doi.org/10.5281/zenodo.10046889

Sophie Stolzenberger[1], Gernot Deinzer[1]

[1]University Library, University Regensburg

Permanent research data management (RDM) is anchored at universities in different ways. Facilities such as the university library or the computing centre often integrate RDM into their departments not least because of their roles within the university network. The (legal) basis on which RDM is operated and assessed mainly depends on the infrastructure of the university.

Long before RDM was introduced as a centralised task at the University Regensburg, researchers faced the challenge to fulfil funders' requirements on their own. If support was needed, first contact points were the library or the computing centre. The decision was made by the issue the researchers encountered: if storage was needed, they contacted the computing centre, if aspects like publishing or archiving were important, they contacted the library. Sometimes, however, the search for suitable solutions ended in a long odyssey, especially when areas overlapped and responsibilites were not clearly defined. First steps to concentrate RDM at one point, mainly at the library, did not solve this problem. It was not considered as RDM in its whole (when technical issues were addressed) nor for the entire university. To overcome this problem the Executive Board of the university decided to reorganise the RDM.

One milestone in the development of RDM was the passing of the research data policy where the principles for the responsible and sustainable handling of research data have been defined. These guidelines assist researchers, for example, in the preparation of data management plans and the storage of research data during and after the project phase, also regarding long-term preservation and accessibility. These guidelines are also addressed to research supporting staff working in a research environment and students.

With the introduction of the research data policy, the establishment of a central organisation for permanent RDM – the UR Data Hub – was requested, where all activities concerning RDM at the university are coordinated. The UR Data Hub acts as an interface between the university, the faculties and the researchers considering their needs and requirements. While the university's responsibility primarily lies in supporting researchers in meeting the requirements of RDM, for example, by providing and extending the technical and organisational infrastructure, the faculty's responsibility is, among others, to communicate the requirements of RDM to their researchers and students. Researchers are responsible for the careful handling of their own research data according to the principles of good scientific practice. Corresponding measures and offers are reconciled and accompanied by the UR Data Hub considering the individual subject cultures.

How can the UR Data Hub specifically assist researchers? Its support covers the entire life cycle of research data including the implementation of an RDM-infrastructure with needs-oriented services according to the FAIR-principles. This comprises the development of subject-related standards for RDM together with the faculties, the establishment of training programmes for quality assurance of RDM for researchers and early career researchers, and the support for creating data management plans considering the requirements of various research funding organisations. Recommendations suggested, for example, by national and multidisciplinary networks, are contemplated and implemented. By developing and operating a public information platform on the research activities of the faculties and other institutions of the university, the UR Data Hub supports the visibility of ongoing research.

Who is involved in the UR Data Hub? The central bodies are the Management Board, and the Steering Committee RDM. While the Management Board leads and coordinates the daily business of the UR Data Hub, the Steering Committee RDM monitors and evaluates the UR Data Hub's work concerning, for example, existing and future policies regarding data management, the professional and technical implementation, and the future development according to subject-specific standards and interdisciplinary cooperations. To guarantee a high level of quality for discipline-specific requests, the scientific participation and representation is of major significance. Therefore, the Steering Committee RDM is established to accompany the work of the UR Data Hub and to make strategic decisions. It consists of the Vice Presidents for Research and Support for Emerging Academics, and for Digitalization, Networks and Transfer, and one RDM representative from each faculty. Additionally, the data protection officer and IT-security officer are included as experts in legal and security questions.

The UR Data Hub is a linking point between the central infrastructure facilities, like library and computing cluster, and the faculties. Therefore, the members are Data Stewards and researchers chosen by their faculties. The Data Stewards are located in the library as the central point of the RDM. They are part of the IT and publishing department of the library and therefore experts in handling, storing, publishing, and archiving data. For computational support also members of the computer centre can become members of the UR Data Hub. They are dealing with storage, backup and handling large data. In addition, each working group, chair or project can propose members for the UR Data Hub. In this way, a close link is established between the infrastructure services and the researchers' requirements for RDM (internal and external).

Cooperation with other institutions within the university is a key task of the UR Data Hub and necessary to ensure a comprehensive RDM-infrastructure. As a composition of the university library, the computing centre, and the faculties, the UR Data Hub coordinates the establishment and expansion of the technical infrastructure that is required for RDM. The officers for data protection and IT-security are consulted on all legal aspects concerning, for example, the digital services offered. Another important task is the cooperation with Open Science officers acting in third-party funded projects: already during the project application process, interfaces in which project-internal structures could be embedded are discussed and potential synergies are identified.

In this contribution, we will give an overview of the process to change from unstructured to a well-formed organisational RDM. The way to establish a new organisational facility will be shown and the new structure with its tasks will be explained.

The NFDI4Culture Helpdesk – Consulting Service regarding Cultural Data

DOI: https://doi.org/10.5281/zenodo.10066604

Angela Kailus [1], Desiree Mayer [2], Martha Stellmacher [2]

[1]Philipps-University Marburg
[2]Sächsische Landesbibliothek — Staats- und Universitätsbibliothek Dresden (SLUB)

NFDI4Culture[1] is the consortium within the German National Research Data Infrastructure (NFDI[2]) that establishes a needs-based infrastructure for research data on tangible and intangible cultural assets.

The research landscape considered here ranging from architecture, art and music to theatre, dance, film and media studies, is characterised by great diversity. It includes university institutes, art and music colleges, academies, libraries, archives, museums and individual scholars as well as infrastructure and service providers within the cultural heritage sector. The consortium aims to establish a network that ensures services guided by the FAIR principles for all phases of the research data life cycle. A considerable challenge is the wide range of application areas, purposes and software that have to be taken into account in digitally supported processes, but also the heterogeneous levels of knowledge in our academic communities about relevant research data management procedures, standards and existing reusable solutions.

Profile of the Helpdesk The NFDI4Culture Helpdesk[3] is a consulting network that enables individual, direct consultation on concrete questions and takes into account the project contexts, individual circumstances and the clients' prior knowledge regarding research data management and subject-specific standards. It is one of the consortium's main services, along with guidelines, trainings and forums of diverse formats with which NFDI4Culture aims to exchange knowledge and promote community integration. The NFDI4Culture Helpdesk is available to all people in the above-mentioned domains, regardless of qualification level or institutional affiliation. We provide advice according to NFDI4Culture's main themes:

- Organisational and technical aspects of the digitisation of cultural assets (2D, 3D, audio, video, AR/VR).
- Data quality, standards, data curation
- Implementation of the FAIR principles
- Development, consolidation, operation and certification of sustainable, interoperable research tools and data services
- Publication processes, especially for multimodal publications and their archiving in repositories
- Data law and ethical issues, such as copyright, property and personal rights, dealing with open science or culturally sensitive objects
- Content, organisation and design of training courses, workshops, etc. on data and code literacy
- Information on eligible funding programmes and requirements for project applications
- Planning research data management

A team of about twelve consultants from different institutions is available to provide support, all of whose members have expertise in one or more of the above topics.

Process and Workflow The NFDI4Culture Helpdesk can be reached via a contact form on the Culture Information Portal. The incoming enquiry gets assigned to a member of the helpdesk team – the data steward - who coordinates the request and provides guidance throughout the entire consulting process. He or she is the contact person for the client and is always informed about all activities concerning this request. Usually, the data steward calls in one or two additional persons with relevant expertise from the consortium or from other interlinked research data management initiatives, service institutions or experts. Advice can range from a written response to long-term support of a project including several meetings.

[1]https://nfdi4culture.de/
[2]https://www.nfdi.de/
[3]https://nfdi4culture.de/de/dienste/details/nfdi4culture-helpdesk.html

Behind an enquiry is often the desire for networking. Therefore, the NFDI4Culture Helpdesk also sees its mission in bringing actors, projects and institutions into contact with each other in order to support the development of the entire domain of NFDI4Culture.

Needs-based orientation The results of the consultations are documented in our project management tool (OpenProject) and the file hosting cloud (Nextcloud) because they also serve to get to know the community and its challenges. In this way, further NFDI4Culture services can be tailored more closely to subject-specific, yet also cross-disciplinary needs. In regular meetings, the Helpdesk team works on optimising its services according to the requirements of quality-driven client-centred consultation practice.

Acceptance of the service The NFDI4Culture Helpdesk service has been well received. Since the project launch of NFDI4Culture at the beginning of 2021, more than 300 enquiries have been sent to us. So far, the majority of support requests have come from researchers who are in the process of planning or carrying out projects in an institutional context. However, there is also considerable interest from institutions that maintain existing resources and from service and infrastructure facilities. Cross-disciplinary enquiries or those from neighbouring academic disciplines working with similar data formats are also gladly accepted.

Increasingly, there is also cooperation with other NFDI consortia, the state initiatives on research data management (e. g. SaxFDM) and similar offers to establish effective and coordinated advisory services, so that our helpdesk contributes to foster increased networking between these initiatives.

NFDI4Culture is funded by the German Research Foundation (DFG) – 441958017.

References

[1] Reinhard Altenhöner, Ina Blümel, et al. NFDI4Culture - consortium for research data on material and immaterial cultural heritage. *Research Ideas and Outcomes*, 6:e57036, 2020. https://doi.org/10.3897/rio.6.e57036.

Enhancing Research Data Management and Reusability: A Case Study at UNU-FLORES

DOI: https://doi.org/10.5281/zenodo.10046951

Timothy Dube[1,2], Azin Zarei[1], Vera Greschner Farkavcova[1], Serena Coetzee[1,3], Edeltraud Günther[1], Kaveh Madani[1,4]

[1]United Nations University Institute for Integrated Management of Material Fluxes and of Resources (UNU-FLORES), Germany

[2]Institute of Water Studies, Faculty of Natural Sciences, University of the Western Cape

[3] Department of Geography, Geoinformatics and Meteorology, University of Pretoria, South Africa

[4] United Nations University Institute for Water, Environment and Health (UNU-INWEH), Canada

Research data management plays a crucial role in modern scientific endeavors. Effective data management practices ensure the long-term accessibility, usability, and reusability of research data. This paper presents a case study on research data management and the dedicated efforts undertaken at UNU-FLORES (United Nations University Institute for Integrated Management of Material Fluxes and of Resources) to enhance the reusability of research data. The institutional knowledge at UNU-FLORES is managed through a Science Knowledge Hub, which serves as a central repository for materials and documents supporting scientific work at the institute. To improve collaboration and document sharing, a new Document Management System was developed and implemented in cooperation with UNU FLORES members and the UNU ICT department.

In 2022, UNU-FLORES demonstrated its dedication to effective institutional project data curation and management by creating a Data Management Plan (DMP). The primary objective of the DMP is to ensure the visibility, transparency, accessibility, security, scalability, and sustainability of institutional data in line with the FAIR data principles (Findability, Accessibility, Interoperability, and Reusability). The DMP provides a guiding framework for effective data management practices, enhancing data discoverability, facilitating access and sharing, enabling interoperability across systems, and supporting data reuse. Furthermore, the DMP addresses knowledge gaps, ensures data continuity during staff transitions, and promotes data reproducibility. Since October 2022, it has been mandatory for all projects with a Pelikan entry at UNU-FLORES to develop a corresponding Data Management Plan (DMP). To ensure effective implementation and understanding of the DMP requirements, two workshops have been conducted thus far. Additionally, meetings, presentations, and surveys have been utilized to gather feedback on the DMP, allowing staff members to actively contribute their thoughts and suggestions. The careful analysis of survey feedback led to subsequent meetings with the entire UNU-FLORES staff, fostering effective communication and continuous improvement of the DMP process. To date, thirty-five project DMPs have been submitted for review by the Data Coach to assess their completeness. These initiatives demonstrate UNU-FLORES' commitment to integrating data stewardship into the institutional infrastructure and daily research practice. By establishing effective research data management workflows, providing necessary tools and support, and offering education, training, and consulting services by data stewards, UNU-FLORES aims to enhance the visibility, usability, and reusability of research data. These efforts align with the objectives of its Research Programme on Sustainability Nexus Analytics, Informatics, and Data (AID)[1].

In conclusion, this case study provides insights into the efforts made at UNU-FLORES to improve research data management, enhance data reusability, and establish effective data stewardship practices. The findings and experiences shared in this paper serve as valuable lessons for institutions seeking to integrate data stewardship into their own research infrastructure and workflows, ultimately fostering a culture of open science and maximizing the impact of research data. By prioritizing data management, UNU-FLORES sets an example for other organizations to follow, ensuring that research data is accessible and reusable and contributes to the advancement of knowledge and sustainable development.

[1]https://flores.unu.edu/en/news/news/resource-nexus-aid-enabling-the-nexus-approach-to-sustainable-development.html

Establishing discipline-specific Data Stewardship at the Data Science Center of the University of Bremen – One Year Review

DOI: https://doi.org/10.5281/zenodo.10047491

Sandra Zänkert[1], Heike Thöricht[1], Lena Steinmann[1], Rolf Drechsler[1]

[1]Data Science Center, University of Bremen

It has been a year since discipline-specific data stewards complemented the Research Data Management (RDM) support infrastructure at the University of Bremen in summer 2022. Now they want to share their experiences in establishing their services, lessons learned, mile stones and what hurdles had and still need to be conquered.

The data stewards are located at the University's Data Science Center (DSC), an interdisciplinary institute that acts as hub for data-intensive research. The DSC's main task is to promote excellent research, education, and interdisciplinary collaboration in the field of data science. For this purpose, the DSC provides researchers from all disciplines with a wide range of services, such as computing capacities, data literacy trainings, financial support as well as technical and methodological consultancy on the implementation of data science methods (e.g. machine learning). Sustainable and FAIR data management are an integral part of data-intensive research, therefore the DSC also provides data steward services.

The DSC's data stewards are discipline-specific experts acting as an interface between researchers and RDM-infrastructure. Their overarching goal is to reduce the barriers associated with RDM for scientists and to promote a cultural shift towards a FAIR-data-culture. According to their professional backgrounds, they foremost support the research domains (1) Social Sciences and Humanities, and (2) Natural and Health Sciences.

During their first year, the data stewards provided consultation for researchers and developed and conducted RDM and data literacy trainings, fostered cultural change, and engaged in local and national networks. The consultations included both support in writing DMPs and questions on implementation of RDM practices during the research process. Here, the data stewards were contacted by researchers across disciplines, including neuroscience, engineering, public health research and ecology but also sociology as well as communication, education and political sciences.

The data stewards further developed discipline-specific data management as well as tool-specific data literacy trainings that where successfully implemented. These cover topics like "How to draft a DMP in the Social Sciences" and a "Git for Dummies" course. The training concepts include online courses as well as in-person workshops. Feedback from participants was mostly positive and always constructive to make the trainings better for the next round. The trainings of the data stewards so far were addressing PhD students and postdocs. To foster cultural change towards FAIR-data, the DSC in general tries to approach a broader audience, e.g. with the low-threshold hybrid lunch-and-learn lectures called "Data Snacks". This snack on data is scheduled for 30 minutes with room for discussion afterwards and cookies and coffee for on-site participants. The first Data Snacks series provided an overview on data management plans and requirements by funding agencies and a second series on tools and standards. The data stewards also explore new ways to teach data management skills in an engaging way, for example with a LEGO workshop [1] on data documentation techniques. This workshop was first implemented for the "Research Data Day 2023" and attracted participants from different status groups and career stages. The data stewards also got invited by different research groups and institutes to share their expertise. For example, they organized a World Café in the humanities for early career researchers and presented their work during the "Brown Bag Session" of the public health institute. Furthermore, they were requested to give introductions to RDM in the context of research classes for university students in in ecology and health sciences.

Cultural change and networking go hand in hand; therefore, part of their work is professional exchange and community engagement with the local data steward network, e.g. during the bi-monthly virtual meeting or more informal in person get-togethers. Moreover, they also participate in national conferences, workshops, and network events to exchange knowledge and experience with a broader RDM support community.

While progress has been made, it's important to acknowledge that there are still challenges to overcome and opportunities for improvement in the coming years. This comprises the need for more visibility of the data steward services in general and the ongoing task of raising awareness for a more open and sustainable data culture. Exciting ideas are buzzing around, such as moderated peer-to-peer learning formats or collaborative research community-driven approaches. The data stewards of the DSC are looking forward to exchanging and discussing concepts, ideas, questions, and experiences with the data steward community during this meeting.

References

[1] Mary M. Donaldson and Matt Mahon. LEGO® Metadata for Reproducibility game pack, 2019. https://www.doi.org/10.36399/gla.pubs.196477.

Research outcomes and scholarly publishing: FAIR principles as quality assessment

DOI: https://doi.org/10.5281/zenodo.10047551

Izadora Silva Pimenta[1], Kevin T. Logan[1], Michaela Leštáková[1], Peter F. Pelz[1]

[1]Chair of Fluid Systems, Technische Universität Darmstadt

The scholarly publishing process has undergone notable transformations in recent years. To increase reproducibility of research outcomes, new methods for publishing research are on the rise [1], such as the possibility to link journal articles to the corresponding data and software and to publish research objects other than articles (e.g. datasets). These practices are gaining more importance and popularity also through initiatives like the Joint Declaration of Data Citation [2].

Publishing datasets as research objects fosters transparency, traceability, comprehensive documentation, and long-term data usability. Hence, as a measure of research data management, it is considered good scientific practice. Still, the role published datasets play in awarding scientific credit needs to be established. This further raises the question of quality assurance in published datasets [3] e.g. in how adequate natural language dataset descriptions and machine-readable descriptions with standardised metadata are provided. While several data journals have emerged to fill this gap, there remains a need for additional measures to promote fairness and objectivity in the publishing process, such as open peer review [3] and preprint publication. Open peer review increases transparency, while preprint publication reduces the time for making research findings available. Developing universally applicable standards that accommodate diverse academic communities, each with unique requirements, also poses a significant challenge [2].

These challenges are addressed in the journal "ing.grid - FAIR Data Management in Engineering Sciences"[1]. ing.grid was founded in order to offer a platform for discussion on and recognition for good scientific practice in generating research data, developing reusable tools for processing data and curating data to make them findable, accessible, interoperable and reusable (FAIR).

By publishing datasets as best-practice examples for well-documented FAIR data, ing.grid drives forward the discussion on standards and assessment criteria for data quality in the field of engineering. This also requires considering further subjects such as data literacy, data management, and data infrastructure. The discussion further benefits from interdisciplinary exchange of experiences beyond the engineering community for which the journal is also open.

Given that data publication promotes transparency and enables its reuse, comprehending its structure and significance [4] emerges as fundamental requirement for openly shared and trustworthy data [5]. For engineers to meet this requirement, tools need to be developed that enhance workflow efficiency and compliance with FAIR principles. These data management-related software tools are equally considered relevant and qualify for publication in ing.grid. Accordingly, the journal accepts three different types of submissions: manuscripts, datasets and software. For each submission type, supplementary material of the other types is encouraged and considered in peer review.

ing.grid works with minimal requirements for ensuring FAIRness in the submissions, such as persistent identifiers, relevant metadata, standardised vocabulary and licence statement. These requirements are checked by the reviewers. Authors are offered clear guidelines to consult this information and adhere to best practices.

The standards applied to the submission for alignment with the FAIR principles also considers domain-relevant community standards of FAIR data. During the reviewer check, this measure is a quality requirement that gains the same weight as novelty, originality and relevant literature presented in the study. The review happens directly in the preprint server following an open single-blind review method, ensuring the openness and transparency of this evaluation. Applying the FAIR principles as a mandatory quality assessment, ing.grid secures the reusability of this material and offers scientific credit for this effort, which is a further step in the transformation of scholarly publishing.

[1]https://www.inggrid.org/

Moreover, it promotes new ways of communicating research results and building applicable community standards.

References

[1] Council of the European Union. Council conclusions on high-quality, transparent, open, trustworthy and equitable scholarly publishing. https://data.consilium.europa.eu/doc/document/ST-9616-2023-INIT/en/pdf, 2023.

[2] M. Martone (ed.). Data Citation Synthesis Group: Joint Declaration of Data Citation Principles. FORCE11, 2014. https://doi.org/10.25490/a97f-egyk.

[3] Leonardo Candela, Donatella Castelli, et al. Data journals: A survey. *Journal of the Association for Information Science and Technology*, 66(9):1747–1762, 2015. https://doi.org/10.1002/asi.23358.

[4] Ingrid Torres Schiessl and Milton Shintaku. Artigos de dados: conceitos e tendências no cenário editorial. *InCID: Revista de Ciência da Informação e Documentação*, 13(2):88–106, 2022. https://doi.org/10.11606/issn.2178-2075.v13i2p88-106.

[5] David Carlson and Hans Pfeiffenberger. Editorial: Science, data and society. *Earth System Science Data*, 15(2):617–619, 2023. https://doi.org/10.5194/essd-15-617-2023.

A Pragmatic Approach on the Provision and Visualization of Formalized Data Provenance

DOI: https://doi.org/10.5281/zenodo.8391406

Arne Rümmler [1,2]

[1]Sächsische Landesbibliothek — Staats- und Universitätsbibliothek Dresden (SLUB)
[2]Technische Universität Dresden

FAIR Principles, R1.2: (Meta)data are associated with detailed provenance: "[...] Include a description of the workflow that led to your data: Who generated or collected it? How has it been processed? Has it been published before? Does it contain data from someone else that you may have transformed or completed? Ideally, this workflow is described in a machine-readable format."[1]

As stated above, data provenance is considered a vital part of FAIR data. It fosters understandability, transparency, reproducibility, and thus, trust in research data. On the other hand, widely known and used research data repositories do not promote and describe metadata fields suitable for describing data provenance in a formalized way. Repositories mostly rely on free text fields or bundle provenance information among related documents (Table 1). Additionally, there is no obvious way to convey the so provided provenance information in a coherent picture, e.g., through visualization of the resulting provenance graph. As a result, the provision of data provenance information is generally lacking, and inconsistent across datasets if available [1], [2]. Although the research data infrastructure landscape is shifting, and future infrastructures will probably integrate the required capabilities, pragmatic solutions are required for the road ahead.

The problem is addressed by the Python package provo[2]. The package implements the starting point terms as defined by the PROV Ontology (PROV-O)[3] as Python classes and methods. PROV-O is a semantic data vocabulary and W3C Recommendation to describe the provenance of any object (on the Web). The package is designed for an audience of data scientists who use Python scripts for data processing and analysis, but do not have in-depth knowledge of PROV-O or semantic web technologies in general. Since the PROV-O classes and properties are defined as actual Python classes and methods, users of the package are assisted in the creation of provenance graphs by their IDE. Additionally, the constructed provenance graphs are checked for validity, and users are prompted with meaningful error messages on violation. Provenance graphs can be created pro- and retrospectively in singular scripts; or in an ad-hoc manner by integration in the data processing and analysis scripts, which allows the capturing of provenance information from the runtime environment. As PROV-O is a semantic data vocabulary, the package provides an interface to rdflib[4], which is a "[...] Python package for working with RDF." The provenance graphs constructed with the package can be exported as RDF documents (targeting machines) or as mermaid-flowcharts[5] (targeting humans). Both of these resources should be provided, when publishing research data. An example of the graphical output is depicted in Figure 1.

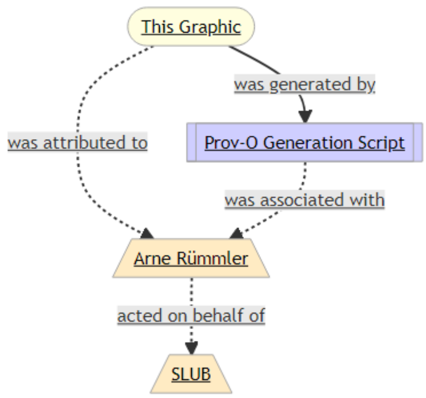

Figure 1 Image of an example provenance graph. The graph and the source code that was used to generate it are published at figshare[6].

[1]https://www.go-fair.org/fair-principles/r1-2-metadata-associated-detailed-provenance/
[2]https://github.com/rue-a/provo
[3]https://www.w3.org/TR/prov-o/#description-starting-point-terms
[4]https://rdflib.readthedocs.io/en/stable/
[5]https://mermaid.js.org/syntax/flowchart.html
[6]https://doi.org/10.6084/m9.figshare.23561661

Table 1 Selection of generalist repositories and their capabilities to express a research dataset's provenance in its metadata.

Repository	Fields that are most suitable for data provenance description
Dryad	Free text fields Methods and Usage notes in the Data description block
Figshare	Selected Relations in the Related Materials block
Harvard Dataverse	Free text fields Description or Notes at dataset level, and free text field Provenance Description or PROV-DM file upload at file level
Zenodo	Selected Relations in the Related/alternate identifiers block
Mendeley Data	Selected Relations in the Related links block

References

[1] Pierre Maillot, Olivier Corby, et al. Metadatamatic: A web application to create a dataset description. In *Companion Proceedings of the ACM Web Conference 2023*, WWW '23 Companion, page 123–126, New York, NY, USA, 2023. Association for Computing Machinery. https://doi.org/10.1145/3543873.3587328.

[2] Julia Fischer, Lukas Egli, et al. Approaches and tools for user-driven provenance and data quality information in spatial data infrastructures. *International Journal of Digital Earth*, 16(1):1510–1529, 2023. https://doi.org/10.1080/17538947.2023.2198778.

The BonaRes repository data publication workflow, tools and support system

DOI: https://doi.org/10.5281/zenodo.10053510

Maureen Fonji Atemkeng[1], Nikolai Svoboda[1], Abraham Tula[1], Xenia Specka[1], Kristin Meier[1]

[1]Leibniz Centre for Agricultural Landscape Research (ZALF)

The BonaRes Repository publishes research data from soil- and agricultural research. All data are described with standardized metadata according to the BonaRes metadata schema[1]. The schema combines the two most popular standards INSPIRE[2] (Infrastructure for Spatial Information in the European Community) and DataCite[3] with discipline-specific metadata elements to foster scientific reuse. The data publication workflow is detailed below.

Data providers (authors) upload datasets via the submission tool. Before uploading data, the author can perform a data quality check through a provided tool kit. Each uploaded dataset is assigned a ticket number through the Redmine ticketing system. Information necessary for publication is stored in Redmine (e.g. internal storage organization, processes involved, as well as updates on the status of each dataset. The system administrator examines the data for malware, backs up the data, then follows with data analysis for gaps and importing data into the SQL database. If the data is fit for publishing, data analysis and import information is entered for the ticket via the Redmine tool, and the data is later assigned to a data steward for metadata collection and review.

The upload[4] tool lets data providers download word metadata templates prepared by data stewards. Through the upload tool filled out metadata templates and any supplemental materials can be uploaded. After data evaluation, import, and backup, the data stewards download a metadata XML file (including part of the metadata from the ticket system), import it into the metadata editor, and finish it with information from the templates. Controlled vocabularies are recommended for data interoperability and reuse. The metadata template of the BonaRes Repository[5] contains links to the AGROVOC[6] and GEMET[7] databases, where keywords can be selected. After the metadata editor receives the whole metadata, a second review is done before making it public. The data provider is informed when missing metadata components are identified. The dataset with complete metadata required for map service setup is then assigned to the geo-informatician via the ticketing system.

Map services are URLs that transfer vector and raster data from desktop systems or local databases to the online. ArcGIS with BonaRes Database Access configures these map services. Data type determines map service configuration workflow. BonaRes repository publishes three data types. Data Type 1 (non-geographical data) is data without a spatial reference or that should not be released (data protection). Geodata Type 2 has two subcategories. Data Type 2.1 for geodata in table columns and Data Type 2.2 for geodata files (coverage, shape, gdb, KML). Data Type 1 (non-geographical) uses a fake map service of German administrative borders. Latitude and longitude (x, y) coordinates are used to create a geo-data type (feature class) for Data Type 2.1's map service. A map service is created using Data Type 2.2 feature classes. The metadata editor couples the dataset to the map service URL and layer name after creating it. After map service the geo-informatician gives the data steward the dataset for final evaluation.

The metadata editor and repository functionalities (downloads, map service) related to the dataset to be released are then reviewed. Each dataset requires a data steward review template. In a control centre linked to all datasets, the dataset's ticket number is entered and the DOI from DataCite registered and stored as a draft. During registration of the DOI in the control centre, DataCite is fed with dataset metadata. Data steward reviews the metadata and changes the DOI status from draft

[1]BonaRes Metadataschema: https://doi.org/10.20387/BonaRes-5PGG-8YRP
[2]INSPIRE: http://inspire.ec.europa.eu/
[3]DataCite: https://schema.datacite.org
[4]Upload Tool: https://upload.bonares.de/
[5]BonaRes Repository: https://maps.bonares.de/mapapps/resources/apps/bonares/index.html?lang=en
[6]AGROVOC: https://agrovoc.fao.org/browse/agrovoc/en/
[7]GEMET: https://www.eionet.europa.eu/gemet/en/about/

to findable. Finally, the data steward updates the dataset status as published in the Redmine tool.

The BonaRes Repository homepage provides authors with central assistance. Data stewards track and respond to queries or reassign them to experts. Until the inquiry is closed, the system updates automatically.

ToMeDa: A Tool for Metadata Management for Computational Sciences

Maximilian Sander[1]

[1]Center for Information Services and High Performance Computing (ZIH), Technische Universität Dresden

In the sphere of computational research, which includes fields such as Computational Fluid Dynamics (CFD), Finite Element Method (FEM) and Performance Engineering, effective data management is crucial. The FAIR (Findability, Accessibility, Interoperability, and Reusability) principles are key to this process, but their implementation can be challenging due to lack of standardized tools, complex data relationships, and fragmented data sources. As an example, computer simulation configuration files and results exist but are disjoint and scattered across files. To address this, we present ToMeDa, a Python-based tool designed to simplify the creation and maintenance of FAIR-compliant metadata.

ToMeDa's process begins by customizing a schema to suit the specific needs of a project, building on an established schema for engineering tasks. Utilizing the robust data validation capabilities of the PyDantic Python package, ToMeDa enables a simplified definition of schemas and supports various entity relationships. The schemas defined within ToMeDa can be exported as user-friendly JSON or blueprint configuration files and be implemented directly in a workflow. This feature reduces implementation errors, ensuring the output aligns with the original intention.

Understanding the importance of capturing all relevant information, ToMeDa distinguishes between required and optional metadata fields. If a required field is omitted, the data gathering process is aborted, whereas if an optional field is overlooked, the user is warned. This approach ensures the capture of all essential metadata and encourages the inclusion of additional beneficial information.

A standout feature of ToMeDa is its ability to effortlessly generate a metadata definition file that is compatible with Dataverse [1], a widely adopted repository software. This greatly simplifies the implementation of custom metadata schemas within Dataverse, thus facilitating a seamless data management experience.

ToMeDa is particularly suitable for computational science engineers, who may not be familiar with the detailed aspects of FAIR principles [2]. By providing intuitive fields and guiding users towards adherence to these principles, ToMeDa makes metadata management more accessible and less daunting. Despite its robust capabilities, ToMeDa offers a flexible, lightweight solution for common metadata management steps, rather than a monolithic system that mandates the use of specific frameworks and additional software infrastructure. It is designed to complement comprehensive data management systems like Dataverse or COSCINE [3] and can function as a preprocessor, ensuring the FAIR compliance of supplied metadata.

The inclusion of pre-filled definition files of widely-used computational software like OpenFOAM within ToMeDa opens the door for the community to contribute further templates, enhancing the tools versatility and applicability. These contributions enable ToMeDa to adapt to a variety of data management needs, broadening its applicability across diverse computational research scenarios and allowing the reuse of proven templates provided by other community members. Importantly, these templates exist as external dependencies, ensuring that the core ToMeDa remains clean and focused.

ToMeDa also aims to create an ecosystem where computational science engineers can become active participants in the wider scientific data community. ToMeDa has been designed with extensibility and adaptability in mind, reflecting its open-source nature. Users are empowered to modify and expand its functionality to meet their specific needs or to accommodate evolving metadata standards. This adaptive capability ensures that ToMeDa remains not only a useful tool but a dynamic platform that can continually meet the shifting requirements of the computational research field. Just as ToMeDa can adapt to meet the evolving needs of the computational research field, it also encourages active adaptation and contribution from its user community.

In conclusion, ToMeDa represents an important contribution for the practical metadata management for computational research. By ensuring adherence to FAIR principles, allowing for easy schema definition, seamless integration with Dataverse and emphasizing a community-driven approach, ToMeDa positions itself as a potent and invaluable tool for researchers and data professionals. As ToMeDa continues to evolve, it seeks to further enhance user experience and adaptability, fostering the open, shared, and reusable knowledge that is pivotal to the future of computational research. By endorsing the broader scientific community's objective of shared and reusable knowledge, ToMeDa positions itself as a pragmatic yet impactful instrument in driving forward metadata management in the computational research.

References

[1] Gary King. An introduction to the dataverse network as an infrastructure for data sharing. *Sociological Methods & Research*, 36:173–199, 2007. https://doi.org/10.1177/0049124107306660.

[2] Mark D. Wilkinson, Michel Dumontier, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1):160018, 2016. https://doi.org/10.1038/sdata.2016.18.

[3] Ilona Lang and Marcel Nellesen. Coscine - FAIR play integrated right from the start. In Tobias Hamann, Mario Moser, and Michael Zeng, editors, *NFDI4Ing Conference 2022 - Conference Contributions*, pages 1–25, 2022. https://doi.org/10.18154/RWTH-2023-02360.

Analysis and evaluation of Data Management Planning Tools

DOI: https://doi.org/10.5281/zenodo.8383753

Carina Becker[1], Carolin Hundt[2], Claudia Engelhardt[1], Johannes Sperling[2], Moritz Kurzweil[2], Ralph Müller-Pfefferkorn[1]

[1]Center for Information Services and High Performance Computing (ZIH), TU Dresden, Dresden, Germany

[2]Directorate – Library and Digital Services, Leibniz Institute for the History and Culture of Eastern Europe (GWZO), Leipzig, Germany

Data Management Plans (DMPs) are crucial for a structured research data management and often a mandatory part of research proposals. A DMP should contain information about the creation, management, sharing and preservation of research data. The manual creation of DMPs can be very time-consuming, since many researchers have to start from scratch, are unsure about the required content and may run the risk of not meeting the funder requirements. By using tools, DMPs can be effectively developed and managed. There are a variety of tools to support the development of DMPs: from discipline-agnostic tools, which can be used to generate a generic draft DMP, to discipline-specific tools, which support the creation of a DMP in a specific research field, such as psychology, biodiversity, engineering, or the life sciences. The project aims to develop a state-wide DMP service for Saxony, ideally building upon an existing tool. Therefore, the objectives of this work are: (1) the identification of requirement parameters to evaluate existing DMP tools and (2) the evaluation of DMP tools based on the identified parameters. The evaluation results will be used to support the development of a quick and easy to use DMP service for members of Saxon research institutions to generate data management plans.

In total, we evaluated 18 DMP tools, 13 of which provide open access. Seven of the tools have a discipline-specific focus, the other eleven follow a discipline-agnostic approach. Eight tools were developed and hosted in Germany. The remaining DMP tools originate from other European countries, the USA and Australia. Based on the findings of 19 expert interviews and a subsequent discussion among the project partners, we defined 32 requirement parameters for the evaluation of existing DMP tools. The parameters were grouped into three main categories: basic functions, technical aspects and user-friendliness. Because the purpose of the evaluation was to identify the tools that could easily be reused as the basis for the DMP service to be developed by the project, we mainly focused on the technical parameters. We aimed at identifying DMP tools, which are easy to host and maintain in order to ensure their adaptability to the specific needs of researchers, institutions, and funders. To further prioritize, a weight factor between zero (not relevant) and three (high priority) was assigned to every parameter. The weight factor was determined individually by each member of the research team, and afterwards the arithmetic mean was calculated. The DMP tools were rated by two different researchers independently according to a fixed rating scheme from zero (poor) to ten (excellent). In a next step, we calculated the arithmetic mean for each requirement parameter. To calculate the final score, the score for each parameter was multiplied by the weight factor. Then, the sum of the rating scores was calculated per main category and for the total score. Furthermore, we wanted to determine, how many of the 32 requirements each of the tools fulfilled. This was done by using a threshold of greater equal five, i.e. a certain requirement was considered to be fulfilled if a minimum score of five was obtained.

The evaluated tools satisfied between three and 28 requirement parameters. 11 tools covered at least half of the parameters. When looking at the total scores, it should be kept in mind that the evaluation was carried out against a set of parameters that was specifically developed to fit our purpose. The highest total rating scores were attained by Data Stewardship Wizard (733.5, DSW), DMPTool (645.5) and RDMO NFDI4Ing (579.5). In the main category 'basic functions', DSW also performed best (269.5) followed by easyDMP (235) and DMPTool (230). The three best performing tools in terms of 'technical aspects' were DSW (190), DMPTool (190) and RDMO NFDI4Ing (181). The most user-friendly ones were DSW (274), DataPLAN from NFDI4plants (262) and DMPTool (225.5). Experience from RDM consultancy at TU Dresden showed that researchers find pre-fabricated text passages very helpful, which are automatically generated by the DMP tool based on their input. Although such a text might need some refinement by the researchers, it can serve as a first draft of a DMP. Accordingly, the corresponding requirement parameters are of high importance to us. The only tools generating pre-fabricated text passages while also providing the tool's source code were DSW, DataPLAN from NFDI4plants and the DMP tool from the TU Dresden Service Center Research Data. Moreover, we consider the machine-actionability of the DMP as an important requirement, because it can facilitate data findability, reusability, automated evaluation and monitoring. A machine-actionable DMP is machine- and human-readable and aims to be interoperable, automated and standardized. Seven DMP tools fulfill the requirements of being open source and providing machine-actionable DMPs.

In the light of recent developments in the area of DMP tools, this study provides an up-to-date evaluation of 18 DMP tools according to 32 parameters covering basic functions, technical aspects and user-friendliness. Our results show that, evaluated against our set of criteria, DSW, DMPTool and RDMO NFDI4Ing received the highest total scores and can be recommended for researchers and institutions as flexible tools for hosting, which provide numerous functionalities. Our results can support tool developers to identify potential improvements and hosting institutions to select a tool suited to their specific needs. In a next step, we will check the feasibility of adapting each of the three tools according to our needs and estimate the respective workload. The most suitable DMP tool will then be customized for our requirements.

A further perspective on data stewardship: Experiences and challenges of "RDM stewards" in a collaborative project in Rhineland-Palatinate

DOI: https://doi.org/10.5281/zenodo.8363542

Johannes Putzke [1], Manuela Richter [2], Thomas M. Schimmer [2], Anett Mehler-Bicher [2]

[1] Trier University of Applied Sciences
[2] Mainz University of Applied Sciences

Recently, scholars had increasingly paid attention to research data management (RDM) practices (cf. the literature reviews by Donner 2022 [1]; and Perrier et al. 2017 [2]). In this context, particularly the role model and tasks of data stewards have been subject to research (e.g., Curdt et al. 2021 [3]; Gruber et al. 2021 [4]; Peng 2018 [5]; Rothfritz 2019 [6]; Seidlmayer and Dierkes 2022 [7]; Verheul et al. 2019 [8]; Wendelborn, Anger and Schickhardt 2023 [9]; Whyte et al. 2018 [10]). Despite the great contributions of all these works to our understanding of data stewardship at higher educational institutions, we have still a limited understanding of the tensions and conflicts that data stewards are exposed to in their daily work. And that aside from the difficulties of defining the term "data stewardship" respectively "data steward". This poster aims to fill in this gap. It answers the following questions: Which tension and conflicts are data stewards experiencing in their daily work, and how can these conflicts be mitigated?

To answer this question, the poster will be structured as follows: First, it portrays the embeddedness of so called "RDM stewards" in a RDM project in Germany. Second, the effects of the relationships of the universities within the federal states as well as the employment conditions in the German science system are examined. Third, it portrays the role understanding and tasks of the "RDM stewards". The poster closes with a vision for the future, to overcome the obstacles.

The project under examination (Research Data Management at Universities of Applied Sciences in the State of Rhineland-Palatinate) is funded by the German Federal Ministry of Education and Research (BMBF) for a period of three years. In the project, eight Universities of applied Sciences (UAS) in Rhineland-Palatinate intend to establish an institutional RDM, which includes the development of competencies and consultation services for researchers and universitary staff as well as sustainable sensitisation for RDM and its context. As basic infrastructure, the project draws on eight "RDM scouts" (8 x 0,5 Full Time Equivalent, FTE) who are placed at the eight UAS, and who act as primary contact persons for the researchers at their institutions. Their main aims are to identify needs of RDM, raise awareness for RDM and provide a first level support for researchers. For discipline-specific questions, they can refer the researchers to one of four "RDM stewards" (4 x 1 FTE) who are in charge of one of the following scientific domains: (1) life sciences and natural sciences, (2) informatics / computer science, (3) engineering, and (4) humanities and social sciences. Despite the cross-institutional nature of the work, the RDM stewards are officially located at four UAS. After the project, the positions are supposed to be merged into eight solitary, sustainable RDM officer positions at the respective institutions (8 x 1 FTE). The RDM stewards' embeddedness in the project comes along with several strengths: Above all, it facilitates discipline specific services to researchers despite scarce resources and funding. At the same time, stewards face general and project-specific challenges. (1) A fundamental problem is the lack of awareness of the RDM topic among researchers at UAS, which they must address (together with the RDM scouts) through various measures. (2) It is also conceivable that the establishment of RDM stewards at certain UAS might lead to reservations on the part of researchers at other UAS. Especially when advising on proposals, sensitive information about research projects is at stake. Researchers may be reluctant to share these with "outsiders" and in view of more or less subtle

This research and development project is funded by the German Federal Ministry of Education and Research (BMBF) within the funding measure Reuse and Management of Research Data at Universities of Applied Sciences (funding number 16FDFH104A) and financed within the Recovery and Resilience Facility of the European Union. The authors are responsible for the content of this publication.

competitive relationships between the UAS. (3) Currently, RDM competency development takes place almost universally at the beginning of the project period in designated work packages, because there is a pronounced shortage of knowledge and competencies in the RDM field. Without a continuation perspective, which is not given at the moment, the skills, knowledge and competencies built up in the project will be lost again to the RDM community afterwards. However, this problem is not limited to the project, but is a general challenge for the RDM sector. This was explicitly addressed at the Herrenhäuser Conference on "Designing Data Spaces in Germany and Europe - Impulses from Science". In addition to these organizational shortcomings, the job description and associated tasks of the RDM stewards are vaguely worded. Therefore, there is a broad scope for the individual understanding of the role as well as for the activities to be carried out. Nevertheless, the stewards have the following tasks: (i) subject-specific competence development, (ii) raising awareness and establishing educational offers, (iii) discipline-related consulting and support, (iv) networking with the RDM community and (v) identification of necessary services (with RDM scouts). Other tasks and responsibilities will become apparent as the project progresses. These newly upcoming on-the-job experiences are essential for shaping future job descriptions and therefore need to be well documented. This is not only important for the project, but also for the emerging profession of "data steward" in principle. In this context, the poster also asks for the prerequisites for science policy and the necessary course to be set to establish the profession of "data stewards" in the long term.

References

[1] Eva K. Donner. Research data management systems and the organization of universities and research institutes: A systematic literature review. *Journal of Librarianship and Information Science*, 55(2):261–281, 2023. https://doi.org/10.1177/09610006211070282.

[2] Laure Perrier, Erik Blondal, et al. Research data management in academic institutions: A scoping review. *PLOS ONE*, 12(5):e0178261, 2017. https://doi.org/10.1371/journal.pone.0178261.

[3] Constanze Curdt, Jens Dierkes, et al. Data Stewardship im Forschungsdatenmanagement - Rollen, Aufgabenprofile, Einsatzgebiete. *Bausteine Forschungsdatenmanagement*, 4(3):70–81, 2021. https://doi.org/10.17192/bfdm.2021.3.8347.

[4] Alexander Gruber, Hermann Schranzhofer, et al. Kompetenzen von Data Stewards an österreichischen Universitäten. *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare*, 74(1):12–32, 2021. https://doi.org/10.31263/voebm.v74i1.6255.

[5] Ge Peng. The State of Assessing Data Stewardship Maturity – An Overview. *Data Science Journal*, 17(7):1–12, 2018. https://doi.org/10.5334/dsj-2018-007.

[6] Laura Rothfritz. *Data Stewardship als Boundary-Work*. Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin, Berlin, 2019.

[7] Seidlmayer, Eva and Dierkes, Jens. Forschung unterstützen: Empfehlungen für Data Stewardship an akademischen Forschungsinstitutionen - Ergebnisse des Projektes DataStew. Zenodo, 2022. https://doi.org/10.5281/zenodo.7327371.

[8] Ingeborg Verheul, Melanie Imming, et al. Data stewardship on the map: A study on tasks and roles in Dutch research institutes. Zenodo, 2019. https://doi.org/10.5281/zenodo.2669150.

[9] Christian Wendelborn, Michael Anger, and Christoph Schickhardt. What is data stewardship? Towards a comprehensive understanding. *Journal of Biomedical Informatics*, 140:104337, 2023. https://doi.org/10.1016/j.jbi.2023.104337.

[10] Angus Whyte, Jerry de Vries, et al. Skills and Capability Framework, 2018. https://www.eoscpilot.eu/content/d73-skills-and-capability-framework, Accessed: 2023-06-20.

Incentivising and Professionalising Data Stewardship – How networking, targeted course offers, and decentral adoption strategies facilitate data stewardship at ETH Zurich

DOI: https://doi.org/10.3929/ethz-b-000610622

Julian Dederke[1], Andres Bucher[1]

[1]ETH Library, ETH Zurich

ETH Zurich is a data-driven research institution. As such, the demand for expertise in topics of Open Research Data (ORD) and Research Data Management (RDM) has grown within the institution in recent years. This demand is driven by the adoption of binding Guidelines for Research Data Management at ETH Zurich, as well as the growing specificity of requirements from funding agencies, which emphasize transparency, openness, and reproducibility in research.

The poster focuses on the implementation of data stewardship at ETH Zurich and highlights the role of networking, targeted courses, and decentralised adoption strategies in facilitating effective data stewardship practices. In all data-driven disciplines, increasing data volumes and a large number of tools for data handling require establishing and professionalising support roles like data stewards. The poster summarises and presents a project and the measures that have been initiated for implementing data stewardship at ETH Zurich. The action plan defined for the university focuses on starting the process of recruiting and professionalising experts in ORD support roles and/or data stewardship at ETH Zurich. Key activities during the initial stages of the project include: (1.) onboarding of data stewards into a university-wide data stewardship network; (2.) including data stewards as a target group into existing RDM trainings, (3.) development of new training formats for data stewards as a target group; (4.) identification and promotion of data stewardship roles existing at the university.

The poster shows the combination of centralised and decentralised elements in the university's data stewardship approach. Consulting and training structures established by ETH Library and ETH Scientific IT Services build a fundament for extending services to the target group of data stewards. Furthermore, the presenter argues that a lack of adoption of data stewardship roles on a decentral level within the institution can only partly be remedied by centralised data stewardship support. As a specific example, the presenter provides insights from workshops and trainings with data stewards that focused on drafting data management strategies for labs and research groups.

A key component of ETH Zurich's action plan is to establish a network of data stewards within the institution. This network focuses at connecting embedded experts and RDM support staff that is already employed in departments, research groups and labs with each other as well as connecting them with central RDM support. It focuses on providing a platform for exchange, offer and share RDM courses and training material within the network, inform about developments regarding data stewardship, and further develop and strengthen offers for RDM support and data stewardship.

To summarize, this poster provides an overview of the action plan for data stewardship implementation at ETH Zurich. It outlines the project's objectives, measures, and initial challenges encountered during the implementation phase. The presented goals and activities are aligned with ETH Zurich's strategy and the Swiss national strategy regarding research data infrastructures and open research data. Finally, the goals and measures taken within the university are reflected in light of the institutional and national strategy on research data infrastructures and open research data.

Institution not anonymised, structure partly inspired by ChatGPT July 20 Version

An institutional framework for sustainable and centralized research data management and the role of Domain Data Stewardship in tool adoption

DOI: https://doi.org/10.5281/zenodo.10046636

Angela Ariza [1], Rukeia El-Athman [1], Jörg Rädler [1], Thilo Muth [1]

[1] Bundesanstalt für Materialforschung und -prüfung

The Bundesanstalt für Materialforschung und -prüfung (BAM, German federal institute for materials science and testing) contributes to research data standardization efforts in various domains of materials science and engineering through its work in several NFDI consortia. We aim to implement a research data management (RDM) platform (BAM Data Store) that enables the central storage of (meta)data generated along the data lifecycle and thus promotes adherence to the FAIR principles.

In 2021, the BAM Data Store was successfully tested with five research groups from the institute. The corresponding framework has since been further developed, starting with an institute-wide rollout in May 2023, with the aim of integrating research workflows and data from up to sixty scientific units of BAM over a period of three years.

In this abstract, we provide insights into the project and discuss the relevance of defining the role and tasks of Domain Data Stewards (DDS) to achieve tool adoption within an institutional RDM culture.

The BAM Data Store: an institutional RDM framework based on the open source software openBIS At BAM, research, development, and technical services are conducted with high societal impact by a scientific staff of more than 1,000 employees. No central RDM or electronic laboratory notebook (ELN) software solutions were established prior to the introduction of the BAM Data Store, resulting in institutional data silos that hampered the generation of FAIR [1] datasets. To close this gap, the BAM Data Store was introduced as a pilot project in 2021. The BAM Data Store is based on openBIS [2], [3], an open source software for RDM developed at ETH Zurich (ETHZ) that includes plugins for the digital representation of laboratory inventory, and an ELN for the documentation of experimental processes. The openBIS data model consists of a hierarchical folder structure and flexible, user-defined metadata schemas (so-called meta-metadata or "Masterdata") that allows for the storage of data in conjunction with both generic and domain specific metadata. openBIS further offers interface for automatized data upload, data analysis, and data publication, as well as the possibility to develop custom plugins.

In addition to the implementation and rollout of a central openBIS instance for BAM, the BAM Data Store project defines work packages to offer training and support, implement governance structures, develop a data policy and guidelines, and to engage different institutional stakeholders for sustainable RDM (Figure 1).

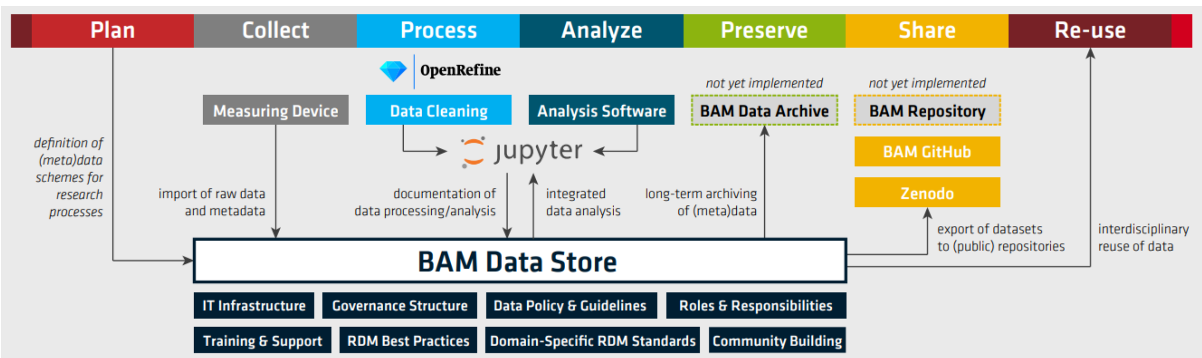

Figure 1 The BAM Data Store supports RDM along the phases of the data life cycle.

Pilot Project: Testing openBIS in five heterogeneous research groups At the beginning of the pilot phase, five pilot groups from different domains of materials science and engineering were chosen to define generic and domain specific RDM requirements and to implement and test openBIS individually for each group. Each pilot group needed to assign one or two researchers to take on the role of instance admins and lead the data model and Masterdata definition efforts for their group. The entire group was then trained to use the software with support of openBIS experts from the ETHZ. While the onboarding process was time-consuming, especially for the instance admins, most of the RDM requirements from the groups could be fulfilled by openBIS. Researchers were not only enabled to document their research in an ELN but also used the framework and its interfaces to automatize their workflows along several stages of the data life cycle. The pilot phase also showed that the successful tool adoption of a complex and highly customizable tool such as openBIS requires close collaboration with domain experts who need to receive tailored RDM training.

Rollout, onboarding concept and definition of "Domain Data Stewards" Based on the positive results of the pilot phase, the institutional rollout phase of the BAM Data Store started in May 2023 with the aim to store and document all research data generated in the institution. The BAM Data Store framework is continuously reviewed and revised to meet the infrastructure scaling challenges, as well as onboarding, training, and support concepts to ensure tool adoption within an evolving RDM culture.

To address some of these challenges, the BAM Data Store team is developing educational materials tailored to the needs of researchers. The onboarding concept aims to enable researchers to become DDS, who play a fundamental role within their research group and within the institution in ensuring the adoption of RDM tools and the implementation of generic and domain specific RDM best practices.

The onboarding concept includes a "communication module" to learn about the RDM needs of each individual research group and to keep stakeholders informed throughout the process.

The "hands-on module" includes four webinars for teaching specific openBIS functions. DDS get to know the functions of the software and customize the inventory and ELN for the group. Optional webinars cover further openBIS functions, e.g., the integration of Jupyter Notebooks for data analysis.

The "Masterdata module" aims to train researchers with any background knowledge of metadata and data model concepts, with a focus on the definition of openBIS Masterdata.

The "community-building module" includes a webinar to communicate the roles and responsibilities of the project team, the DDS, and other users. Finally, DDS and users can share success stories and challenges while contributing to a growing RDM community at BAM (Figure 2).

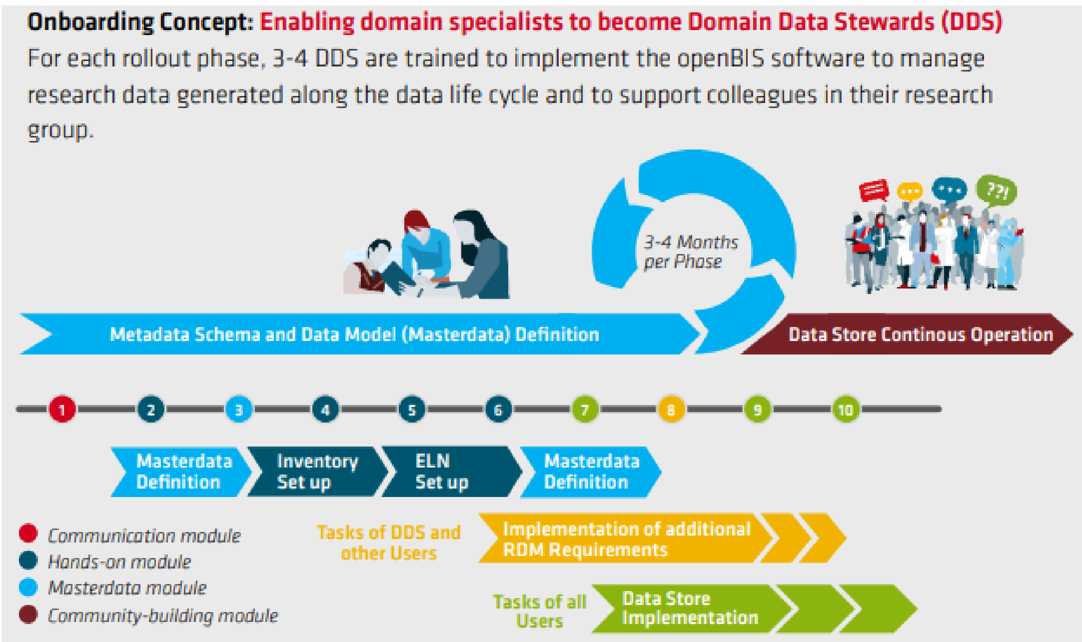

Figure 2 BAM Data Store Onboarding Process.

Roles and tasks definition for an institutional RDM Culture To create an RDM culture among researchers, we consider it essential to clearly define the role and responsibilities of DDS, as this informs group leaders about the time resources required, motivates researchers to take on the role, improves the quality of RDM processes and helps to identify gaps in RDM training.

To achieve this in a standardized way and to ensure the quality of teaching concepts, we applied the RDM learning matrix which lists learning objectives in six RDM topics for various skill levels, including professional data stewards.[4]

We selected learning objectives that we believe would enable researchers to become DDS and excluded those that align with the roles of other RDM actors within the institution. So far, we have identified 53 out of 150 learning objectives that are likely to have a greater impact on tool adoption and therefore should be prioritized as DDS tasks. This approach highlights tasks from four RDM topics:

1. Basics and overarching concepts of RDM
2. Working with data
3. Documentation and metadata
4. Supporting structures.

The BAM Data Store team aims to reach a consensus with institutional RDM actors on this topic and to complete the definition of the role and responsibilities of DDS. Our goal is also to identify the content of training modules that should be offered to researchers to enable them to perform the role of a DDS and be recognized as such in the institution. Additionally, we aim to distinguish which RDM topics require training material tailored to materials science and engineering, and which can be covered by reusing existing RDM training materials or outsourced to partner institutions.

References

[1] Mark D. Wilkinson, Michel Dumontier, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1):160018, 2016. https://doi.org/10.1038/sdata.2016.18.

[2] Angela Bauch, Izabela Adamczyk, et al. openBIS: a flexible framework for managing and analyzing complex data in biology research. *BMC Bioinformatics*, 12(1):468, 2011. https://doi.org/10.1186/1471-2105-12-468.

[3] Caterina Barillari, Diana S. M. Ottoz, et al. openBIS ELN-LIMS: an open-source database for academic laboratories. *Bioinformatics*, 32(4):638–640, 2016. https://doi.org/10.1093/bioinformatics/btv606.

[4] Britta Petersen, Claudia Engelhardt, et al. Lernzielmatrix zum Themenbereich Forschungsdatenmanagement (FDM) für die Zielgruppen Studierende, PhDs und Data Stewards (Version 2). Zenodo, 2023. https://doi.org/10.1093/bioinformatics/btv606.

Training program for research data management in chemistry

Ann-Christin Andres[1], Benjamin Golub[2], Daniela Hausen[3], Pascal Scherreiks[4]

[1]Johannes Gutenberg-University Mainz
[2]Technical University Braunschweig
[3]RWTH Aachen University
[4]Friedrich-Schiller University Jena

NFDI4Chem was formed as a consortium for chemistry within the national research data infrastructure (NFDI). The vision of NFDI4Chem is to seamlessly digitalise the entire workflow in chemical research. Starting at the bench with the provision of open-source electronic lab notebooks (ELNs), through developing standards, interfaces, and tools, NFDI4Chem strives to remove the analogue gaps from the digital data life cycle [1], [2]. To raise awareness of all RDM aspects intensive dissemination and training activities for chemists at all levels are required.[1]

In this poster, the consortium briefly introduces itself and sets out its strategy for chemistry-specific training and support. It is a significant building block in driving digitization in chemistry. For this purpose, various interactive workshops are offered regularly and on request - training on electronic lab notebooks, general RDM training for up to 25 participants or for institutes, CRCs, ... These trainings are part of NFDI4Chem's roadshow concept and are offered in person or online. The target groups are researchers and data stewards. The high demand of the RDM workshops as well as the feedback of the participants of past workshops pushed us to change our concept. In this regard, we had to consider the following key points:

Include topics such as communication at institutes, in CRCs, with data stewards ... Communication is a very important element when it comes to collaboration, starting a project or joining a team. Certain ground rules need to be established in the beginning and expanded over time. We discuss possibilities like wikis, policies, Git issues, ... with the participants and look for the best possible solution in the respective situation.

Design the workshop for an increased number of participants (50-90 persons) When upscaling a workshop concept designed for 25 participants, the challenge with an interactive workshop is to adjust the exercises and time. While the presentation part remains the same, new activities have to be considered or a unifying element has to be designed.

Workshop in a shortened version for a brief introduction (max. one day) The challenge in reducing the time is to shorten the content in an appropriate form or to balance the presentation parts and the exercises. What is important to us when teaching? It is also possible to integrate platforms for self-learning in order to gain resources and not to shorten the content too much.

Include offers and further development from the NFDI4Chem consortia There is a balance to be struck in teaching general, chemistry-specific, and applied RDM skills. At the same time, it is important for the trainers to stay up to date. They need to be informed about changes in e.g. publishers or new developments of the consortium. RDM tools are constantly being developed and new features are added. These must also be integrated into the training in a timely manner. A work-around must be developed for this purpose.

Design a best practice workshop with consortia from similar interests In research data management, best practice examples have been sought for a long time. Slowly, it is becoming clear to many researchers that there are some best practice examples, but that they still need to be adapted to the respective situation. In order to present a broad spectrum to researchers, we have developed a series of workshops with related disciplines (chemistry, materials science, physics). In a two-hour format, these workshops will briefly introduce the researchers to the respective problems and provide discussion space for implementation possibilities.

[1]Website: https://www.nfdi4chem.de

Workshop to foster programming skills in chemistry In chemistry, too, it is no longer enough to be able to produce good syntheses. Many calculations support chemists in their everyday laboratory work and show alternative (synthesis) paths, some of which work better. To educate chemists in this area, NFDI4Chem and NFDI4Cat offer a programming course over 3 days to learn Git and Python for chemists.

References

[1] Sonja Herres-Pawlis, Oliver Koepler, and Christoph Steinbeck. NFDI4Chem: Shaping a Digital and Cultural Change in Chemistry. *Angewandte Chemie International Edition*, 58(32):10766–10768, 2019. https://doi.org/10.1002/anie.201907260.

[2] Christoph Steinbeck, Oliver Koepler, et al. NFDI4Chem - Towards a National Research Data Infrastructure for Chemistry in Germany. *Research Ideas and Outcomes*, 6:e55852, 2020. https://doi.org/10.3897/rio.6.e55852.

Hit by Reality – How Socio-Technical Challenges Shift the Design of an Information System in an Interdisciplinary Research Environment

DOI: https://doi.org/10.5281/zenodo.10047721

Steffen Strohm[1], Hartwig Buenning[2], Matthias Renz[1]

[1]Department of Computer Science, Kiel University
[2]Department for Ecosystem Research, Kiel University

Within our archaeology-driven Collaborative Research Center (CRC), lots of different researchers are brought together to work towards the mutual goal of answering overarching research questions focused on human activity in prehistory. Without claiming completeness, these differences are to be found in people's personal backgrounds, scientific profiles, political interests and general demands towards their respective (working) environment.

Therefore, interdisciplinary research always involves bridging gaps in knowledge, matching vocabularies and balance very different expectations. Especially with technical solutions, expectations may easily be misled or over-the-top. This may happen due to a lack of time investment or a long-term accumulation of misunderstandings.

In our CRC the idea of how an effective information system should look like, such that it supports the overall research goals of the project had been too fuzzy and unattended for too long, which led to a long process of gathering requirements and defining necessities in a phase, where all sides involved already expected more progress. Suddenly, questions needed to be answered very precisely. On the one hand basic questions like:

- How do we present the CRC's results?
- Which software components should be used?
- Who should have access to which extent and when?

On the other hand, questions that aren't (easily) answered by effective decision-making needed answers as well:

- To which level of detail should which kind of data be represented?
- After we brought the data together, what do we want to do with it and what can we do with it?

Unfortunately, the latter questions are tough to answer and will only be answered alongside the journey to an integrated information solution, where data from very different research perspectives and interests are brought together. And even though these answers shape the information solution, they cannot be answered sufficiently before a user-friendly tool exists that enables interdisciplinary exploration of the data sets presented.

In the first phase of the CRC, the information system was thought of as a fully integrative solution, which allowed to decompose any arising structured data to then identify common attributes and allow for a wide range of database queries and analyses. While the phase ended with a proposition of a technical solution on how to create a tool using such a flexible data model, the solution was far from being implemented in a practice-ready state.

In the technical department of the project, the switch from Phase 1 to Phase 2 was dominated by a full change in personnel on all levels, which made an effective knowledge transfer impossible and slowed down progress heavily in the beginning of the second phase. After a year-long phase of catching up and trying to find out if the beforementioned questions had already been answered, the design of the information system had become clearer, involving new components that had not been thought of before (see Figure 1).

In a first step a repository-like functional component was needed to collect the data which had already been created or was in use in the CRC. Some of this data is openly accessible, others had to be protected through restricted access, which required at least a simple self-built solution within the project and involved a process on defining an adequate but still usable meta data catalogue, access rights, answers on licensing questions and lots of more small but relevant details.

The data collections accessible through the catalogue are then meant to be used for a record level integration, similar to what was thought of in the first phase of the CRC. However, a lot of detailed questions occur, when data from roughly 15 sub projects with very different background and perspectives are brought together. What common attributes are left when all data of the project are compared? How deep is the integration of such different data sets coming from archaeology, botany, ecology, geophysics and so on? Without spoiling too much: the integrational concept remains rather shallow, reduced to few dimensions (space, time, type), however, the need to create a common backbone also in the values of these dimensions introduces new challenges, not just on the technical side but also considering roles and processes on how to effectively discuss these topics.

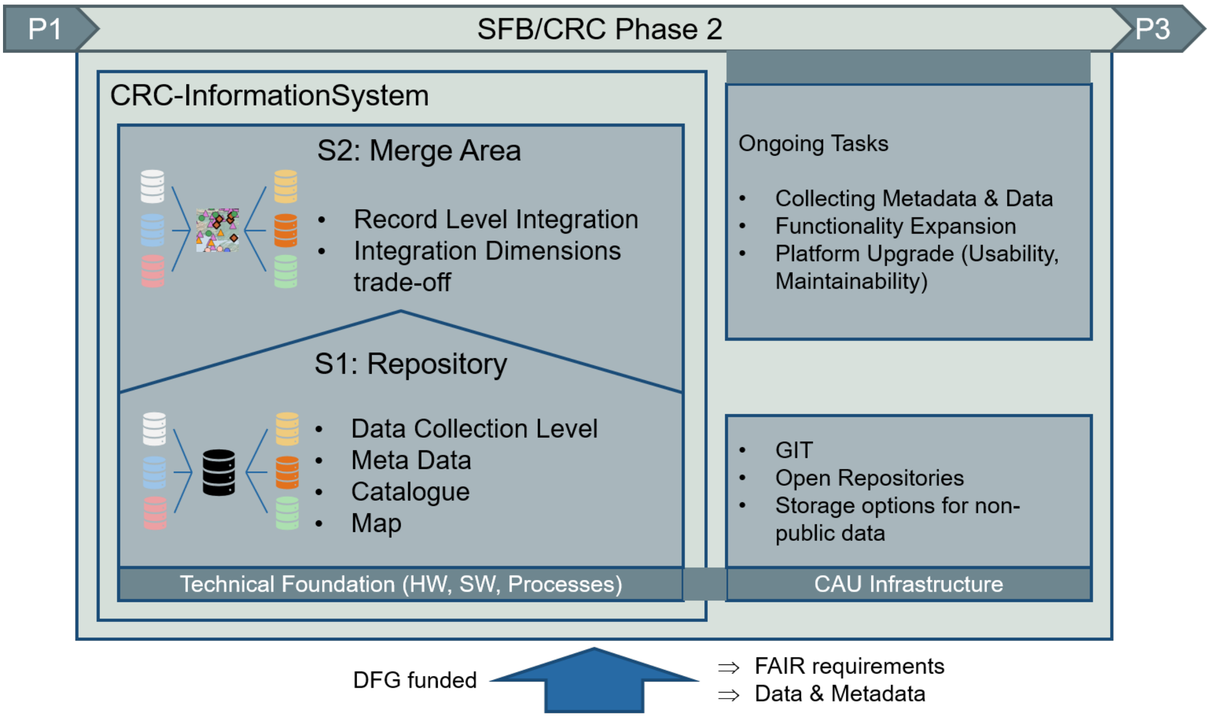

Figure 1 Concept of the CRC-Information System after Reconsideration in the second project phase.

Introducing a new research support service for Electronic Laboratory Notebooks (ELN) using service engineering methods

DOI: https://doi.org/10.5281/zenodo.10069278

Adienne Karsten [1]

[1]University of Münster (WWU)

This poster presents the implementation of the Electronic Laboratory Notebook (ELN) software, "elabFTW" at the University of Münster. It provides an overview of the efficient design and optimization of the ELN implementation using the concept of service engineering while taking into account all stakeholders involved in a university setting.

The goal of establishing the ELN service at the University of Münster is to provide researchers with an ELN software with ongoing operation. The service aims at enabling digital laboratory documentation to meet the FAIR principles and offer a replacement for traditional analog lab notebooks. The University and State Library of Münster (ULB) is responsible for coordinating the process, as well as providing support, consultation, and training for the subsequent cloud-based deployment and operation by the IT Department of the University (WWU IT).

elabFTW is a free, interdisciplinary open-source ELN with large global community both regarding its user base and its active contributors. The ELN can be installed on local servers (on-premises), ensuring data sovereignty and GDPR compliant data processing. It ensures the data will be processed by the university's own IT. The application is web based and allows for mobile access. An outstanding feature of elabFTW is the Timestamp Authority (TSA) based on the RFC 3161 standard, which utilizes the timestamp server provided by DFN (German Research Network). This enables the generation of defined revision points for laboratory records to ensure data integrity.

The implementation of the ELN service at the University of Münster is based on the Five Phase-Model[1] by Meiren/Barth [1]. This model consists of procedure models to structure service development as five phases: idea phase (1), concretization phase (2), conception phase (3), implementation phase (4), and operation phase (5), with no separate end-of-life phase. After each phase, the service draft is discussed and approved before it can progress to the next phase. The product owner (in this case, the data steward) conducts these evaluations in collaboration with an expert group. This iterative approach ensures a gradual and detailed service design, with constructive feedback being essential in each decision point.

Since the ELN service at the University of Münster is still in the early stages, this poster focuses on the first three phases, briefly outlining phases (4) and (5). Specifically, it emphasizes the idea phase (Phase 1), but primarily highlights the concretization phase (Phase 2) and the conception phase (Phase 3). Techniques and methods employed during these phases included user stories as well as qualitative interviews for requirement analysis. These interviews, conducted as part of a Master's thesis [2] , provided a detailed overview of researchers' expectations and requirements for an ELN and how these can be met through an interdisciplinary ELN solution.

Another key aspect of the poster is the involvement of additional stakeholders in the ELN implementation process, that highlights contexts such as data protection, IT security, and employment law. This involvement occurs during the conception phase and includes consultations with the university's data protection officers as well as works councils. The need for coordination also encompasses identity management and backup concepts on the technical level.

The poster is aimed at informing data stewards, researchers, and IT professionals at research institutions who are interested in implementing an ELN system. It provides a visual representation of the various phases involved in the implementation, specifically focusing on the elabFTW software. The poster demonstrates how the needs and requirements of different stakeholders can be considered to enable more efficient research processes, improved data integrity, and enhanced collaboration within the university environment, both within the service-receiving research groups and in terms

[1]The ULB adheres to this framework, which has been modernized, critically examined, and adjusted to meet the current needs of the educational institution.

of research support provided by data stewards.

elabFTW is a vital component of a research support service portfolio. lt enhances research data management by standardizing processes, improving the quality of laboratory documentation, facilitating collaborative work, and ensuring FAIR data treatment. The adoption of the Five-Phase-Model during the software selection process identified elabFTW as a solution that does not require custom software engineering nor development. As an open-source solution, it aligns with the university's values and empowers researchers to carry out and optimize their activities independently.

References

[1] Thomas Meiren and Tilmann Barth. *Service Engineering in Unternehmen umsetzen*. Fraunhofer IRB Verlag, Stuttgart, 2002.

[2] Maike Sommer. Einführung von Elektronischen Laborbüchern (ELN) an der WWU Münster, 2023. [unpublished Master's thesis in the course of study Digitales Datenmanagement (DDM), Humboldt-Universität zu Berlin (HU Berlin) and Fachhochschule Potsdam (FHP)].

Data stewardship and research data management tools for multimodal linking of imaging data in plasma medicine

DOI: https://doi.org/10.5281/zenodo.10069368

Mohsen Ahmadi[1], Robert Wagner[1], Philipp Mattern[1], Nick Plathe[1], Sander Bekeschus[1], Markus M. Becker[1], Torsten Stöter[2], Stefanie Weidtkamp-Peters[3]

[1]Leibniz Institute for Plasma Science and Technology (INP), Greifswald, Germany

[2]Leibniz Institute for Neurobiology (LIN), Magdeburg, Germany

[3]Heinrich-Heine-Universität Düsseldorf, Center for Advanced Imaging (CAi), Düsseldorf, Germany

Cold physical plasmas are used in the field of plasma medicine by employing clinically approved devices to treat non-healing and infected wounds and skin diseases. Preclinical research focuses on cancer treatment and applications in dentistry using gas plasma technology. Exploring plasma applications in medicine holds great potential but also represents a major challenge. Therefore, in-depth studies often require a combination of various investigation methods, with bioimaging playing an important role. By combining data from different imaging modalities, such as optical imaging, fluorescence imaging, and mass spectrometry imaging, researchers could gain a more detailed understanding of the effect of plasmas on biological systems. This, however, requires the implementation and use of sophisticated research data management (RDM) solutions to incorporate the influence of plasma parameters and treatment procedures as well as the effects of plasma on the treated targets. In order to address this, RDM activities on different levels and from different perspectives are started and brought together (Figure 1).

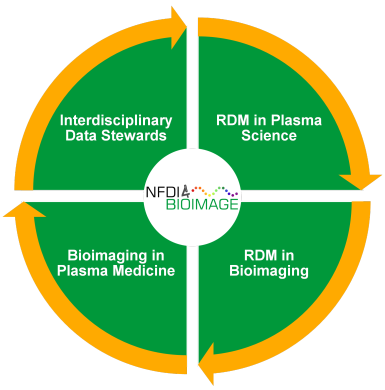

Figure 1 Participant Use Case in NFDI4Bioimage[1]

- Researchers from the field of plasma medicine familiarize themselves with RDM, collect researchers' requirements, work as data stewards, and bring demand-driven RDM solutions into the labs.
- RDM experts in the field of plasma sciences bring in existing RDM tools and processes that meet the needs of plasma sciences, such as domain-specific metadata schemas and research data repositories, and further develop them according to the specific requirements of imaging techniques.
- RDM experts in bioimaging support adopting RDM standards and tools for imaging methods and integrate them with RDM solutions from the plasma domain to address the specific needs of imaging in plasma science.
- Cross-institutional data steward teams with mixed expertise collaborate to enable broad community support for imaging in plasma science and other research fields.

This results in activities, aiming at the devise of FAIR (findable, accessible, interoperable, reusable) image objects (FAIR-IO) suitable for multimodal data linking and at enabling data stewardship for imaging across institutional boundaries. The present contribution elaborates on the concept of a cooperative project involving data stewards at the interface between RDM infrastructure developments and topical research. It is discussed how domain-specific (plasma medicine), generic and method-specific (bioimaging) RDM tools like INPTDAT [1], Adamant[2], eLabFTW[3], and OMERO[4] are considered to be integrated into a comprehensive RDM workflow enabling the extensive application of data-driven research methods in plasma medicine.

The work is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) –

[1]NFDI4BIOIMAGE – a consortium in the National Research Data Infrastructure: https://nfdi4bioimage.de/en/start/, retrieved in 2023-07-21.

project number 501864659.

References

[1] Markus M. Becker, Lucian Paulet, et al. INPT-DAT – a new data platform for plasma technology. Zenodo, 2019. 72nd Annual Gaseous Electronics Conference (GEC), College Station, Texas, USA; https://doi.org/10.5281/zenodo.3500283.

[2] Ihda Chaerony Siffa, Jan Schäfer, and Markus M. Becker. Adamant: a JSON schema-based metadata editor for research data management workflows. *F1000Research*, 11:475, 2022. https://doi.org/10.12688/f1000research.110875.2.

[3] Nicolas CARPi, Alexander Minges, and Matthieu Piel. eLabFTW: An open source laboratory notebook for research labs. *Journal of Open Source Software*, 2(12):146, 2017. https://doi.org/10.21105/joss.00146.

[4] Chris Allan, Jean-Marie Burel, et al. OMERO: flexible, model-driven data management for experimental biology. *Nature Methods*, 9(3):245–253, 2012. https://doi.org/10.1038/nmeth.1896.

Sharing, storing and analysing cell image data in a Collaborative Research Centre using OMERO

DOI: https://doi.org/10.5281/zenodo.10069503

Soni Lama[1,2], Johannes Koch[1],Nina Schulze[2], Stephanie Rehwald[1]

[1]Research Data Services, University Library, University of Duisburg-Essen
[2]Center of Medical Biotechnology, Faculty of Biology, University of Duisburg-Essen

A growing trend in the biomedical sciences is the emergence of cross-disciplinary research approaches and systematic collaborations combining highly specialized methodologies to address complex biomedical research questions. As a result, the complexity of data generated is increasing and demands sophisticated knowledge exchange. This requires not only careful organization and shared archiving of data, but also extensive documentation and cross-linking of data to exploit the full potential of valuable research results.

Collaborative Research Centre (CRC) 1430 is an interdisciplinary consortium of biologist, molecular oncologists and chemists which consists of 23 collaborative projects in the field of cell proliferation, cell cycle progression and cancer cell plasticity. The project produces huge amount of data that needs to be processed, shared and stored subsequently. Therefore, research data management (RDM) becomes a vital and essential process to ensure data accessibility for collaborative studies and to facilitate innovative research. In order to facilitate research data infrastructure and management in CRC, we provide a well-designed and structured approach to manage, process and share scientific data originating from CRC projects by ensuring data storage and sharing via Nextcloud within the CRC according to the FAIR data principles (Findable, Accessible, Interoperable, Reusable) (Wilkinson et al., 2016, [1]); implementing electronic laboratory notebook (eLabFTW) for documentation of laboratory data, processes and database; introducing an open source software platform Open Microscopy Environment Remote Objects (OMERO) as a central repository and searchable database for image data; building a central information platform Dataverse to integrate all research data and their documentation enriched with metadata.

The majority of CRC projects are working on cell imaging using state-of-the-art microscopy techniques to investigate molecular mechanism of cell state transitions. The CRC projects generate a tremendous amount of cell image data which needs to be analysed, shared and stored. Therefore, to prevent data loss and enable its reusability, secure storage and long-term archiving and handing requires a suitable tool to manage, visualize and analyse heterogenous image datasets recorded in various file formats. To fulfil those requirements, we have opted for an information system – the open source software platform OMERO, provided by Open Microscopy Environment (OME) initiative for the management of imaging data.

OMERO is an image data management and storage software which is a platform for visualizing, imaging and annotating scientific image data (Burel et al., 2015, [2]). It is a secure central repository for handling images that facilitates in viewing, organizing, analysing and sharing data from anywhere with internet access. OMERO supports over 150 image file formats, including all major microscope formats. Thus, OMERO can display image data generated with different biological imaging techniques and instruments. This demonstrates the potential of OMERO to facilitate management of image data ensuring accessibility and interoperability. Therefore, OMERO has been set as a standard tool in NFDI4Bioimage. Various annotation tools are available in OMERO, making research data findable and reusable. Nevertheless, the linkage of imaging data with the corresponding information about sample preparation, experimental procedures and metadata remains a challenge. It is currently addressed by the adaptation and implementation of interfaces between OMERO and eLabFTW among others by NFDI4Bioimage and adapted in CRC's infrastructure.

References

[1] Mark D. Wilkinson, Michel Dumontier, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1):160018, 2016. https://doi.org/10.1038/sdata.2016.18.

[2] Jean-Marie Burel, Sébastien Besson, et al. Publishing and sharing multi-dimensional im-

age data with OMERO. *Mammalian Genome*, 26(9):441–447, 2015. https://doi.org/10.1007/s00335-015-9587-6.

Report on a pilot study to implement OMERO for managing imaging data

DOI: https://doi.org/10.5281/zenodo.10103316

Silke Tulok[1], Gunnar Fabig[1], Thomas Müller-Reichert[1]

[1]Faculty of Medicine Carl Gustav Carus at TU Dresden, Core Facility Cellular Imaging

The **FAIR** (Findability, Accessibility, Interoperability, Reproducibility) concept of data management is of relevance for all current areas of research. Moreover, a statement about the handling of scientific data of is often required when applying at national and/or international funding agencies. For imaging data obtained from both light and electron microscopy, data management is associated with a number of challenges related to the size, complexity, dimensionality, visualization and quantitative analysis of the acquired data. Quite recently, we have started to use the Python-based and open-source image data management software, OMERO. This software package has been introduced by the Open Microscopy Environment developers of the University of Dundee. OMERO is an encouraging option to manage microscopy data. It is installed on a virtual machine and connected to a central data server to manage the storage of images in a multi-user environment. In addition to long-term data storage, it provides possibilities for saving important metadata in an efficient manner, thus avoiding multiple copies of data. It can also handle open source-based processing tools for image analysis, thus allowing effective image analysis workflows. The Core Facility Cellular Imaging (CFCI) at the Faculty of Medicine Carl Gustav Carus at TU Dresden is currently running a pilot project for testing the use and handling of the OMERO software. This is done together with interested users of the imaging facility and a research group. Currently, we are pushing forward this pilot study on a small scale without any data steward. However, we cooperate with the I3D:bio-Team (Information Infrastructure for BioImage Data)[1] in learning all necessary skills and administrative issues. Over the last months, we faced some challenges in using this professional software for data management on a daily basis. Our experiences argue for giving data management issues into the hands of dedicated personnel not fully involved in research projects. As funding agencies will ask for higher and higher standards for implementing FAIR-data principles in the future, this will be a relevant topic for the whole research community. During the workshop, we thus want to introduce a convenient solution, which could be applicable for many users within the DRESDEN-concept research alliance. Our poster will demonstrate how to establish and manage OMERO. Furthermore, we will report on technical, administrative and research-specific challenges by showing examples of our first user cases. Last but not least, we will also discuss the use of OMERO for teaching students within the scope of e-learning courses.

[1]https://gerbi-gmb.de/i3dbio/

MorpheusML Model Repository: World-wide Open Repository for Modeling and Simulation in the Life Sciences

DOI: https://doi.org/10.5281/zenodo.10103445

Diego Jahn[1], Jörn Starruß[1], Robert Müller[1], Andreas Deutsch[1], Lutz Brusch[1]

[1]Center for Information Services and High Performance Computing, Technische Universität Dresden

MorpheusML[1] is an innovative, globally adopted open-source modeling and simulation framework for the study of multi-scale and multicellular systems. It enables the modeling of complex biological systems at the cellular level, taking into account spatio-temporal and mechanical aspects. The XML-based declarative modeling language MorpheusML ('Morpheus Markup Language') provides symbols, a vocabulary, and a biomathematical syntax to describe the dynamics and coupling of the various model components, allowing scientists to focus on the 'what', i.e., the desired outcome, rather than the 'how', i.e., explicitly programming the step-by-step instructions to simulate the behavior of the biological system.

Building on this, free and open-source simulation software such as Artistoo[2] or Morpheus[3] further allows the integration of experimental data and provides a whole set of analysis tools to extract and visualize relevant features during the simulation.

The models defined in this way describe our quantitative understanding of dynamic processes and interactions, e.g. the kinetics of cytotoxic T cells in the human immune system[4], and contain all the necessary mathematical equations and parameter values.

MorpheusML is now being used worldwide[5]. The new model description language has proven that multicellular models can be developed cooperatively and independently of a specific simulation platform and can be freely exchanged between simulators such as Morpheus or Artistoo. This enables a new quality of cooperation on more complex models and has been practiced since. All project results are open source and embedded in the worldwide COMBINE community for the standardization of data exchange in the life sciences[6].

A key element in establishing a sustainable software infrastructure for multi-scale modeling and simulation of multicellular biological systems is the provision of a public, easily accessible, collaborative and citable platform for publishing, sharing and archiving MorpheusML-based models. Therefore, to make the results of the researchers accessible to all in accordance with the FAIR principles, the MorpheusML Model Repository[7] has been developed. It is divided into three categories, (1) simple didactic example models that serve as a starting point for users' own further developments and are explained in detail, (2) global user-contributed models for complex or interesting effects of self-organization in biological systems, and (3) models from publications after successful peer review of the corresponding paper. We also use the Elixir infrastructure[8] to assign each model its own permanent MorpheusModelID, which can be cited in publications. Researchers also often submit their models to us prior to the publication of the associated paper (and can set an embargo themselves until their paper is published) in order to be able to cite a MorpheusModelID in the data availability statement of their own paper. Currently, over 80 models are available in the repository and more are published every week.

Building on the distributed version control system Git, the static site generator Hugo and the DevOps platform GitLab, the repository was developed with a deliberate focus on the utilization and integration of established, widely used open source tools. The goal was to use existing solutions for the requirements of this kind of repository, such as complete history and full version-tracking capabili-

[1]MorpheusML project homepage: https://morpheus.gitlab.io
[2]Artistoo simulator: https://artistoo.net/converter.html
[3]Morpheus simulator: https://gitlab.com/morpheus.lab/morpheus
[4]Example model of the kinetics of cytotoxic T cells in the human immune system: https://identifiers.org/morpheus/M9495
[5]MorpheusML user map: https://morpheus.gitlab.io/#user-map
[6]COMBINE organization for worldwide standardization of multicellular models:https://fairdomhub.org/projects/90
[7]MorpheusML Model Repository: https://morpheus.gitlab.io/model
[8]MorpheusML namespace in the Identifiers.org central registry: https://registry.identifiers.org/registry/morpheus
[9]Course materials: https://morpheus.gitlab.io/#courses

ties, and to integrate them as seamlessly as possible with our existing educational resources[9]. For all steps, from submission and editing by the users to the review process by the repository maintainers to the publication of citable MorpheusML models, the above mentioned tools were combined[10].

Along with the continuous extension and improvement of the functionality of the MorpheusML standard, we also want in parallel by the further development of the MorpheusML Model Repository to make published research results more accessible and reproducible to all, and on the other hand to give young researchers the opportunity to work independently on the exciting topic of modeling of biological systems through an comprehensive and constantly growing collection of informative didactic models.

We here present our continuously growing citable MorpheusML Model Repository for multi-scale and multicellular systems in the life sciences featuring models from peer-reviewed papers as well as user-contributed examples and didactic example models.

In addition to demonstrating this concrete application for the multicellular modeling community, the project should also serve as an encouragement to approach research software problems not necessarily by developing mostly new own code, but by making intensive use of existing open source tools and combining them into new software solutions in a practical way.

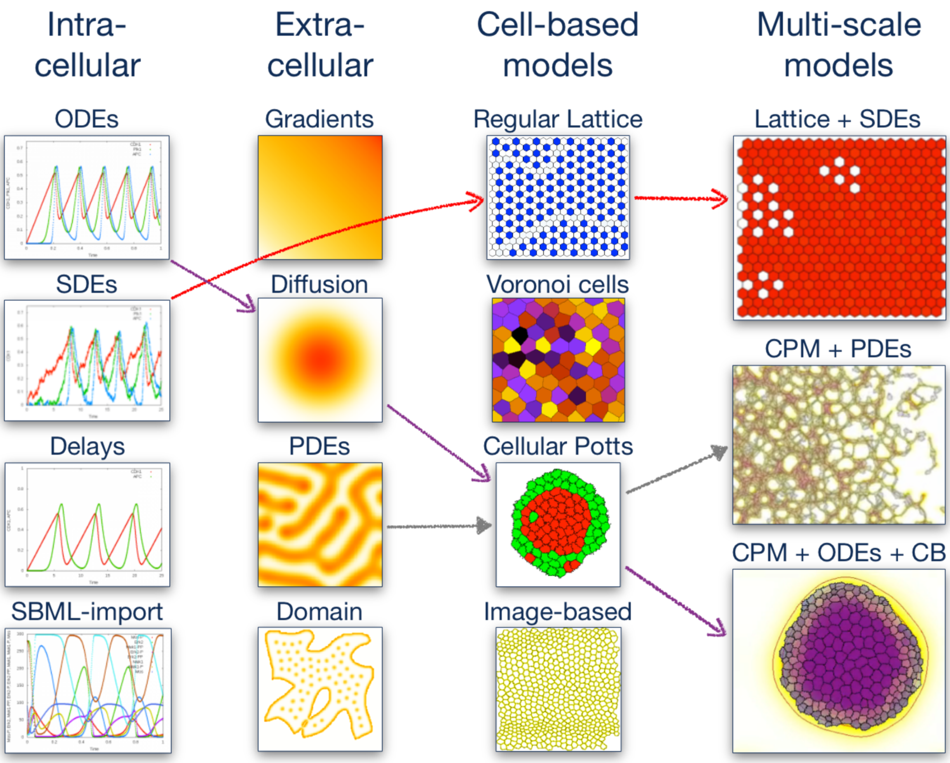

Figure 1 Modular assembly of multi-scale and multicellular models

[10]MorpheusML Model Repository source code: https://gitlab.com/morpheus.lab/model-repo

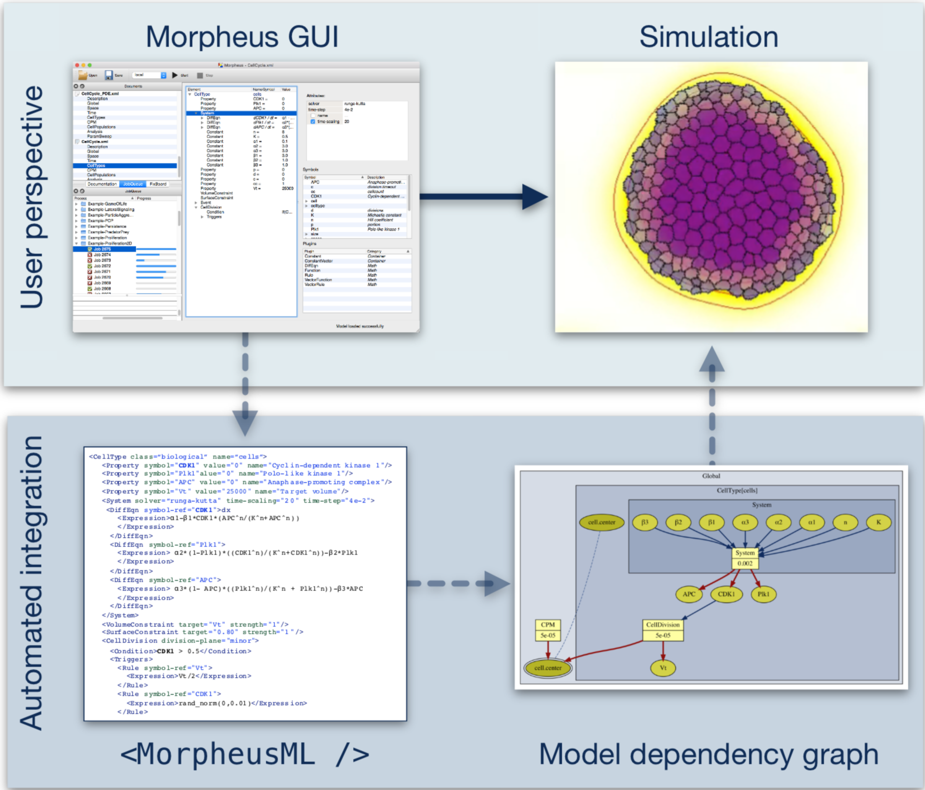

Figure 2 User-friendly modeling without programming

Figure 3 MorpheusML user map

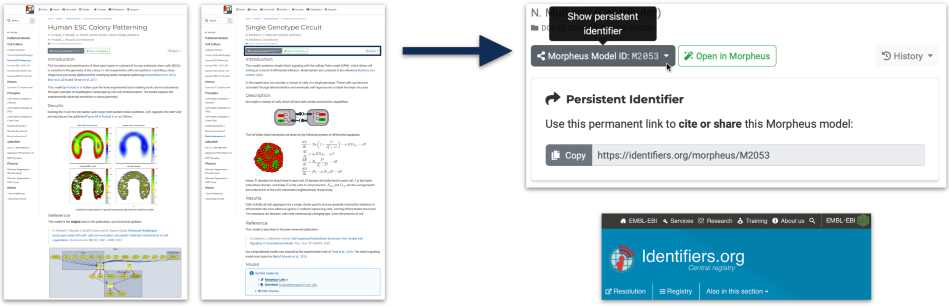

Figure 4 Citable MorpheusML models

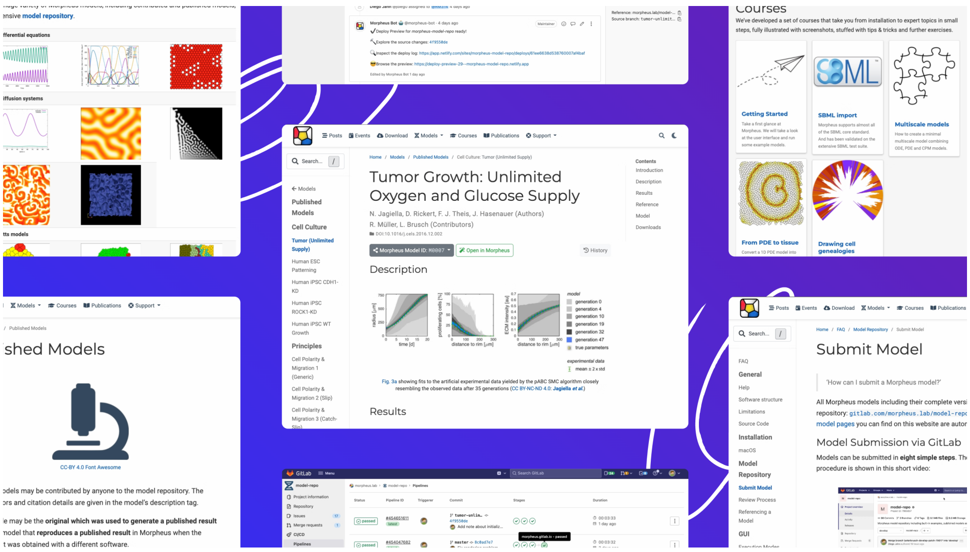

Figure 5 MorpheusML Model Repository

Tools to support data handling

DOI: https://doi.org/10.5281/zenodo.10103582

Maike Kleemeyer[1], Karolina Mader[1], Tobias Bengfort[1], Sebastian Nix[1]

[1]Max Planck Institute for Human Development

Given the increasing amount of (digital) research data as well as the strive towards openness and transparency, professional research data management (RDM) has become an essential factor for excellent and sustainable research. In addition, research institutions particularly in the humanities are confronted with strict guidelines regarding the processing of personal data since the European General Data Protection Regulation (GDPR) became effective. Dedicated positions are created as, e.g., data stewards to support researchers when it comes to managing their data effectively, but also to ensure legally compliant data handling. However, the complexity of the task often coupled with the unavailability of critical information challenges RDM staff in providing adequate support. We present two tools that we developed in the recent past to help RDM staff on the one hand to keep track of studies acquiring data, thus potentially needing support and, on the other hand to ensure GDPR compliant data handling.

1. A **study-registration** to provide one central overview of studies conducted at a research institution. By collecting a (rather small) set of relevant meta-information, the tool enables RDM staff to approach researchers in the very beginning of their research projects and to provide custom-tailored support along the data life cycle (e.g., defining a data structure before data collection starts, or converting to standardized formats such as BIDS for MRI datasets). By setting the stage right from the beginning, we hope to markedly reduce the effort of making data and code openly available in the end. As incentive, the tool assigns studies with badges for openness (preregistration, open data, open materials) to acknowledge the use of Open Science practices.
2. **Castellum** [1] is a turnkey open-source web application for the data protection-compliant management of participant related data. The application provides a clearly defined structure for handling the data of all study participants. While contact information (e.g., name and postal address), recruitment characteristics (e.g., age and education level) and process information (e.g., existing consents and current availability) are stored in Castellum, scientific data is explicitly stored elsewhere but connected via its pseudonym service. Further features of Castellum include participant recruitment, managing study appointments, and coordinating the study procedure. The development took place in close coordination with scientific staff and the data protection officer of the Max Planck Society. Thus, Castellum takes into account relevant aspects of the rules of good scientific practice as well as data security.

With these central tools, we are finally able to accompany studies throughout their life-cycle and can (semi-)automatically approach researchers with our services when reaching important cornerstones. In the long run, this shall help ensure that researchers can rather easily share data and analysis code according to the FAIR principles together with manuscripts containing the research results. Likewise, we can anonymize data easily, react to participants' rights regulated by GDPR (e.g., the right to information under Article 15 GDPR and the right to erasure under Article 17 GDPR). By presenting these tools to a broad community of RDM professionals we cannot only share our experiences when designing these tools and corresponding workflows, but more importantly benefit from valuable feedback and see how we can further improve.

References

[1] Tobias Bengfort, Taib Hayat, and Timo Göttel. Castellum: A participant management tool for scientific studies. *Journal of Open Source Software*, 7(79):4600, 2023. https://doi.org/10.21105/joss.04600.