Bausteine Forschungsdatenmanagement
Empfehlungen und Erfahrungsberichte für die Praxis von Forschungsdatenmanager*innen

Development and Implementation of Research Data Management in the DFG-Funded Research Consortium SFB/TRR219 – Experiences from the Second Funding Period

Kseniia Dukkart[i]

2026

Zitiervorschlag

Dukkart, Kseniia. 2026. Development and Implementation of Research Data Management in the DFG-Funded Research Consortium SFB/TRR219 – Experiences from the Second Funding Period. *Bausteine Forschungsdatenmanagement. Empfehlungen und Erfahrungsberichte für die Praxis von Forschungsdatenmanager*innen* Nr. 1/2026. DOI: 10.17192/bfdm.2026.1.8963.

[i]IT Center RWTH Aachen University. ORCID: 0009-0008-1230-8041

1 Abstract

Research Data Management (RDM) is a crucial component of any Collaborative Research Centre (CRC). It facilitates interdisciplinary collaboration, promotes good research practice and supports reproducibility, and adherence to the FAIR principles (Findable, Accessible, Interoperable, Reusable) (Wilkinson et al., 2016).

This article presents the RDM approach of the Transregional Collaborative Research Centre SFB/TRR219 "*Mechanisms of Cardiovascular Complications in Chronic Kidney Disease*." It outlines measures such as data management plans (DMPs), storage and archiving in Coscine [1], electronic lab notebooks (ELNs), and internal guidelines. It also discusses challenges and resulting workflows. The experience of SFB/TRR219 demonstrates how RDM can be effectively integrated into an ongoing consortium and provides a model that can be adapted and reused in similar collaborative research settings.

Keywords: Research Data Management, FAIR principles, Data Stewardship, Electronic Lab Notebooks, Metadata, Coscine

2 Introduction

The SFB/TRR219 brings together researchers from medicine, biology, biochemistry, and bioengineering at RWTH Aachen University, the University Hospital Aachen (UKA), and Saarland University who work together to investigate the fundamental mechanisms underlying the increased risk of cardiovascular disease (CVD) in patients with chronic kidney disease (CKD).

The second funding period of SFB/TRR219 spans from 2022 to 2025. The introduction of structured RDM began in April 2023, when the Data Steward joined the consortium, when collaboration networks and workflows were already established.

At that time, the SFB/TRR219 consisted of approximately 110 participants across 20 subprojects. A comprehensive and interdisciplinary approach integrates molecular, cellular, and systemic studies, supported by specialized service units for animal models, mass spectrometry, proteomics, bioinformatics, and biobanks. The SFB/TRR219 generates heterogeneous data types, including mass spectrometry data, single-cell and bulk sequencing data (FASTQ, FASTA, GenBank), flow cytometry data (FCS), imaging data (DICOM), instrument-specific formats (e.g. Bio-Rad compressed and optical files), experimental analysis files (pzfx, wsp). Data originate from in vitro experiments, animal models, cell lines, and human samples. Sensitive patient-related data were handled separately and stored within the consortium's biobank infrastructure.

[1] Coscine https://about.coscine.de/en/

Although RDM was widely recognized as an essential component of successful research, much work remained to translate this understanding into daily practice. Implementing RDM required time for training, communication, and retroactive alignment of existing workflows.

The following sections describe how these challenges were addressed through technical infrastructure, policy development, and community-building activities.

3 RDM Implementation in the Second Funding Period (FP2)

Regular weekly coordination meetings were held between the Managing Director and the Data Steward to discuss current RDM needs, next steps, and ongoing developments. Although these coordination meetings were conducted without researchers, the Managing Director is also a principal investigator in several subprojects and therefore familiar with day-to-day research practices. This close collaboration ensured that new initiatives were aligned with consortium goals and that RDM received visible support from management.

The Data Steward often acted as a facilitator between researchers, IT infrastructure teams, and the central RDM services of RWTH Aachen University [2], ensuring that appropriate expertise was connected across levels.

During FP1, RDM activities were mainly handled at the level of individual working groups. In the FP2, specific RDM challenges included coordinating practices across two locations (RWTH Aachen University and Saarland University), elevating RDM from isolated group-level solutions to a consortium-wide level, and handling heterogeneous data types (e.g. experimental, computational, and clinical data). These challenges were addressed through the introduction of consortium-wide DMPs, the implementation of a shared ELN system, the development of metadata profiles, and the establishment of internal RDM guidelines. According to the wish of the consortium, none of mentioned documents or practices were published.

3.1 Data Management Planning

Data Management Plans (DMPs) were established for all 18 research subprojects within the SFB/TRR219 to ensure structured planning of data collection, documentation, storage, access, and sharing throughout the research lifecycle. For each subproject, coordination took place with the Data Steward, involved principal investigators

[2] Research Data Management at RWTH Aachen University https://www.rwth-aachen.de/cms/root/forschung/Angebote-fuer-Forschende/Forschungsdaten-und-information/~bpboai/Forschungsdatenmanagement-FDM/lidx/1/

(one to three PIs per subproject) and, where applicable, a nominated project member (e.g. a researcher or laboratory technician). Beyond meeting funder requirements, the DMPs served as a practical tool to promote good data organization, prevent data loss, and support reproducibility and future reuse.

DMPs were created using the Research Data Management Organizer (*RDMO*), which supports systematic planning and documentation of research data workflows (Anders et al., 2024). A dedicated "*Life Sciences / Medicine (with UKA reference)*" template was developed in RDMO and served as a common model for all subprojects. The template adds explicit questions on responsibilities, long-term archiving, institution-specific storage options, common biomedical data formats, linkage between laboratory notebooks and data, expected data volumes, and requirements for handling personal and sensitive data in compliance with the GDPR. These adaptations ensure that DMPs capture regulatory, ethical, and practical aspects specific to medical and life sciences research. Its development was carried out in close collaboration with the Managing Director, who provided the institutional and technical information necessary to adapt the template to the needs of medical and life sciences research.

The RDMO instance at RWTH Aachen University was used as a central platform. Researchers at the Saarland University were either invited as guest users or had their DMP's entered by the Data Steward, depending on local access conditions. The development of the DMPs for all subprojects took approximately four months.

At the end of the second funding period, all DMPs were updated to reflect changes such as new project structures, new principal investigators or research personnel. This ensured continuity and supported onboarding of new team members, contributing to sustainable data management practices and systematic data monitoring across the consortium.

3.2 Electronic Lab Notebooks

An ELN system, *eLabJournal*, was introduced to replace traditional paper-based laboratory notebooks and support structured digital documentation across SFB/TRR219 [3]. As a browser-based tool, *eLabJournal* enables secure and consistent access from laboratory environments and remote work settings, supporting modern research workflows in both institutes and clinical facilities. The ELN was already an institutional service at UKA, which made adoption within the consortium much easier.

The ELN system allows documentation at a project level rather than being tied to individual users, enabling multiple scientists to collaboratively document experiments and link research data directly to experimental entries. Each of the 18 SFB/TRR219 subprojects based at RWTH Aachen University adopted the ELN early in the second funding period, with the system subsequently introduced and fully implemented at Saarland

[3] eLabJournal https://www.elabjournal.com/

University during the fourth consortium year. This ensured consistent documentation practices across institutions.

To facilitate adoption, researchers received onboarding and training support. Regular bimonthly ELN user meetings were held, with an average participation of approximately 30 users per session. These meetings served as a platform for experience exchange, discussion of challenges, and presentation of practical solutions related to ELN usage.

External experts (e.g. Data Stewards from NFDI4Ing) as well as internal ELN-experienced researchers from SFB/TRR219 were invited to present best practices, featuring best-practice demonstrations and discussions on ELN organization, data linking, and collaborative workflows. These meetings were held in a hybrid format to include researchers from all partner sites and were recorded and stored within the consortium's internal drive, ensuring sustainability and efficient onboarding for new members.

In addition to internal training activities, the consortium actively engaged in knowledge exchange with other research groups. Joint events with partners such as the DFG-funded research consortium CRC1382 Gut-Liver Axis facilitated the exchange of best practices in areas such as the storage of samples in a centralized sample database, the use of advanced search and filters, the management and storage of protocols and SOPs, and the creation of experiment documentation.

While ELN adoption was not formally evaluated through quantitative metrics, feedback gathered via the questionnaire at the end of the FP2 indicated broad acceptance and increasing routine use among participants.

3.3 Policies

A bottom-up RDM guideline approach was central to SFB/TRR219's RDM strategy, ensuring that not only PIs but also early-career researchers were trained and well-aware of good RDM practices and could contribute to decision-making and tool development. A workshop driven by postdocs and PhD/MD students was held at the 2024 annual meeting to discuss practical RDM challenges and improvements, such as the metadata profiles creation, storage strategies, data organization and naming conventions, and version control. The discussions and ideas expressed during this workshop provided the foundation for further initiatives.

One of them was the development of an 'SFB-dedicated, Bottom-Up RDM Guideline Document' for the structured documentation of experiments as well as the storage of the associated *in vivo*, *in vitro,* and clinical data and the documentation of data analyses. This guideline was written in tight collaboration between the Data Steward, the postdocs, the PhD/MD students and the Managing Director of the SFB/TRR219 consortium. For this, an RDM working group with regular weekly meetings was established. The development of the guideline took approximately three months. The

document exists on the internal server, where it is accessible to all members of the consortium. Due to its strong focus on consortium-specific workflows, local infrastructure, and internal procedures, the guideline was intentionally maintained as an internal working document and was not published externally.

The guideline states that data should be stored securely in shared systems and organized using standardized, date-based file names without special characters, should include meaningful metadata in file names, and a clear, structured folder hierarchy per project that includes complete experimental documentation, raw and analyzed data, and separate handling of sensitive patient identifiers.

3.4 Data Storage & Archiving Infrastructure

Secure and long-term research data archiving is essential to comply with good scientific practice and enable future reuse of data (Deutsche Forschungsgemeinschaft, 2019). Within SFB/TRR219, the *Coscine* platform [4] served as a central solution for data storage and archiving. Personal patient data were excluded from Coscine and stored separately in the consortium's biobank infrastructure. *Coscine* provides controlled access, persistent storage, and FAIR-aligned metadata capabilities, supporting transparent and sustainable research practices across the consortium.

To ensure high-quality, interoperable metadata, project-specific metadata profiles were developed in line with Resource Description Framework [5] standards. The profiles were not based on a single standardized metadata schema but were designed in alignment with RDF and SHACL World Wide Web Consortium, 2017) concepts. This process required intensive collaboration with individual project groups.

These profiles captured essential metadata elements relevant to the consortium's research domains, including mass spectrometry, imaging, and RNA sequencing data. Profiles were created in collaboration with service units and domain experts, enabling harmonized terminology and consistent documentation across laboratories and institutions.

To streamline the archiving procedure, a custom automation script was developed. The script standardized data upload processes for, e.g., mass spectrometry data. It also extracted key metadata fields and automatically populated project-specific *Coscine* metadata profiles, significantly reducing manual effort for researchers and increasing metadata consistency. This framework improved (meta)data organization, sustainability, sharing, and metadata quality across the consortium.

[4] Coscine https://about.coscine.de/en/
[5] RDF; SHACL - Shapes Constraint Language https://www.w3.org/TR/rdf11-concepts/

3.5 Connection between *eLabJournal* and *Tick@lab*

Another initiative originating from the annual RDM Workshop involved exploring interoperability between the consortium's frequently used tools - *eLabJournal* and *Tick@lab*[6], which is a centralized management software system for laboratory animal research. The goal was to enable automatic transfer of animal-related information into corresponding experimental records within *eLabJournal*. If successful, such a connection could potentially be reused not only within SFB/TRR219 but also by other working groups using both systems.

Several RDM working group meetings and community consultations were held to define which data should be transferred, clarify the expected use cases, and specify the technical requirements. The Data Steward organized these meetings and coordinated exchanges between the *eLabJournal* and *Tick@lab* support teams. Although potential points of connection were identified, the technical implementation required extensive customization of both systems, and additional research effort for data validation and synchronization. No productive technical integration was implemented.

This connection attempt highlighted the importance of conducting early feasibility assessments and ensuring a close involvement of the Data Steward in researchers' day-to-day processes to realistically evaluate technical complexity, data integration efforts and workload implications.

3.6 Training & Community Building

Community engagement and continuous training were key components of the RDM strategy within SFB/TRR219. Several measures were implemented to raise awareness, strengthen collaboration, and promote the adoption of research data management practices across all project groups.

A dedicated RDM webpage [7] was created on the consortium's website to centralize information on upcoming events, available tools, and current news. Monthly RDM newsletters, were distributed to all members. These newsletters provided updates on RDM topics such as data structure, version control, FAIR principles, and journal data requirements, as well as practical recommendations for daily data handling.

A mandatory RDM course "Research Data Management – an Overview", offered by the RWTH Aachen University RDM team [8] was introduced for all consortium members, ensuring that every PI, postdoc and PhD student received foundational training. For

[6] A-Tune Software AG, Tick@lab – Animal Research Facility Softwar https://www.a-tune.com/products-services-software/animal-research-facility-software/
[7] SFB/TRR 2019 FDM https://www.sfb-trr219.de/fdm
[8] RWTH Aachen University https://www.rwth-aachen.de/cms/root/forschung/Angebote-fuer-Forschende/Forschungsdaten-und-information/~bpboai/Forschungsdatenmanagement-FDM/lidx/1/

postdocs and PhD students this course was mandatory, they proved their attendance by sending the certificate.

To provide more individualized support, biweekly consultation hours were established in the format of open hours, offering researchers the opportunity to discuss project-specific RDM questions. At the start, this consultation was used by consortium members mostly to ask questions about ELN usage and data organization in groups. In the final year of the FP2, these consultation hours were expanded into short presentations on selected RDM topics, such as file naming, Johnny decimal system, Coscine usage, and metadata profiles, fostering greater exchange and discussion within the consortium.

4 Lessons Learned and Conclusion

Implementing RDM in an ongoing consortium required both technical innovation and cultural adaptation. The consortium encompassed a broad diversity of data types (e.g. mass spectrometry, imaging, sequencing, and clinical data), data volumes, and disciplinary data and publication cultures, as well as varying requirements regarding data safety, sensitivity, and access restrictions. Establishing personal contact with researchers proved crucial for identifying real needs, addressing discipline-specific challenges, and overcoming initial scepticism toward RDM measures.

The Data Steward role was essential for connecting research practice with institutional infrastructure, facilitating communication across organizational levels, and ensuring that RDM activities were integrated into scientific workflows. In particular, the presence of a central contact person who was familiar with both technical infrastructures and domain-specific practices helped translate abstract RDM principles into practical, usable solutions.

Automated solutions – such as standardized metadata scripts and shared ELN environments – significantly improved efficiency and reproducibility. At the same time, the experience showed that technical integration between diverse tools remains resource-intensive and requires realistic planning, early feasibility assessment, and close coordination with software providers.

Feedback collected via a consortium-wide questionnaire and during consultations and training activities indicated increased awareness of RDM topics, greater acceptance of Data Management Plans, and broader and more advanced use of ELNs across project groups. These observations are primarily qualitative in nature and reflect practical experience rather than a formal, systematic impact evaluation.

Motivation to engage with RDM was supported through a combination of mandatory baseline training, easily accessible consultation formats, and visible practical benefits, such as reduced documentation effort, clearer data organization, and simplified data

sharing. Challenges specific to the SFB/TRR219 context included disciplinary heterogeneity, multiple locations, and differing local infrastructures. These were addressed by combining standardized consortium-wide approaches with flexible, project-level adaptations and continuous dialogue with researchers.

Looking ahead to the third funding phase, lessons learned include the importance of earlier integration of RDM planning into project proposals, further automation of metadata capture and archiving workflows, and stronger alignment of RDM measures with publication and data-sharing processes.

5 Acknowledgments

Parts of the research data described in this article were managed using the research data management platform Coscine. The storage resources were provided by the Research Data Storage (RDS) and DataStorage.nrw, funded by the Deutsche Forschungsgemeinschaft (DFG: INST 222/1261-1 and INST 222/1530-1) and the Ministry of Culture and Science of the State of North Rhine-Westphalia (MKW: 214-4.06.05.08-139057 and 214-76.01.09-7-7937).

6 Declaration of AI use

The author has used AI tools (DeepL, ChatGPT) for grammar, style, spelling, and translation that do not influence the substance of the article. The author retains full responsibility for the content.

References

Anders, I., Enke, H., Hausen, D. A., Henzen, C., Jagusch, G., Lanza, G., Michaelis, O., Peters-von Gehlen, K., Rathmann, T., Rohrwild, J., Schönau, S., Wedlich-Zachodin, K., & Windeck, J. (2024). The research data management organiser (rdmo) – a strong community behind an established software for dmps and much more. *Data Science Journal*, *23*, 28. https://doi.org/10.5334/dsj-2024-028

Deutsche Forschungsgemeinschaft. (2019). Guidelines for safeguarding good scientific practice (DFG code of conduct). https://doi.org/10.5281/zenodo.3923602

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B., et al. (2016). The fair guiding principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18