Data stewardship goes Germany 2025 (DSgG 2025), 30 September - 1 October 2025,

Karlsruhe Institute of Technology (KIT) from

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

Abstract

Research Data Management (RDM) has become a vital tool of modern scientific practice. However, significant challenges persist, especially in interdisciplinary research projects such as Ruhr University Bochum (RUB)’s Collaborative Research Centers (CRC) 1280 "Extinction Learning", funded by the German Research Foundation (DFG). In CRC 1280, researchers from biology, psychology, medicine, and computational neuroscience produce heterogeneous data through diverse experimental methods including: electrophysiology, electroencephalography, microscopy, and functional magnetic resonance imaging, which are collected both from human and animal subjects. This diversity in the field of neuroscience results in a wide array of data lacking standardization in data formats and metadata, which hampers data exchange, interoperability, and reuse. These challenges make it difficult for researchers to align with the FAIR (Findable, Accessible, Interoperable, and Reusable) principles and fulfill the requirements of funding organizations.

To overcome these obstacles, the CRC 1280 developed a common data model and made the storage of data on a central network drive mandatory in its RDM policy. The model features a hierarchical folder structure that organizes data by subject, session, and modality, combined with a metadata inheritance strategy to ensure structured and consistent data management. To support interoperability, standardized vocabularies were introduced for CRC-specific metadata fields and aligned with established frameworks such as Dublin Core and DataCite. To enhance the incentive for the researchers for storing research data according to the CRC data model, we aimed at providing a platform which enables a faceted search according to the CRC’s common metadata schema and thereby enhance findability of data and cross-group collaboration on data within the CRC.

For the implementation of the repository platform, Samvera Hyrax 3 was selected for its flexibility, scalability, and robust community support. On this basis, the repository platform ReSeeD was implemented, which today serves also as the campus-wide institutional repository solution of the Ruhr-University Bochum. ReSeeD enables instantaneous data storage and displays data in a familiar hierarchical folder structure that reflects the complexity of neuroscience research. The implementation of this

hierarchical data model includes the presentation of all sub-folders of the data structure along with their associated metadata, which are (partially) prepopulated by inheritance across the folder levels. This supports structured data entry, retrieval, and publication, enabling researchers to formally archive and publish datasets using a customized infrastructure that improves visibility and ensures compliance with best practices in research data management.

ReSeeD facilitates the modification of files at any stage prior to publication or archiving, which is essential for CRC 1280 to promote regular and practical use of the repository. All published data in ReSeeD are associated with persistent identifiers (DOIs), which guarantee persistent access, enhances data citation, visibility, and promotes open science. Additionally, ReSeeD reflects CRC 1280's structure through a three-step roles and rights system, enabling fine-grained control over data curation and ensuring high-quality data publication. To reduce administrative burden and enhance metadata quality, ReSeeD also supports automated metadata population, pre-filling fields such as funding references (e.g., DFG project numbers), thereby improving interoperability and reuse across disciplines.

A major achievement in ReSeeD was the design of two data ingestion workflows: Workflow A enables bulk ingestion for completed studies, while Workflow B supports daily data entry through a researcher-friendly web interface, promoting early and consistent metadata capture during active projects. This dual approach ensures that ReSeeD accommodates both retrospective and prospective data management strategies, embedding RDM into daily research activities without disrupting scientific workflows.

INF provides comprehensive training through face-to-face lab workshops, workshops on reproducibility, data transparency, open tools, and annual "Lab Data Cleaning Days" which are hands-on sessions for organizing, documenting, and uploading data. Overall, these efforts helped researchers integrate ReSeeD into their daily routines and reflect broader CRC data stewardship strategies, where early engagement and continuous guidance are key to fostering a lasting culture of data stewardship.

Looking ahead, the RDM concept of the CRC will continue to evolve to meet future needs. In the third funding period (FP3) of CRC 1280, we will introduce the use of electronic lab notebooks (ELN) to support workflows for early-stage metadata capture and to improve traceability and data quality. To ensure a targeted and effective implementation, we will begin by selecting representative use cases from CRC research groups and focus on commonly used modalities such as MRI, behavioral experiments, electrodermal activity (EDA), and questionnaires. These use cases will serve as best-practice examples for a gradual and scalable ELN rollout across the CRC. For these, standardized documentation templates will be developed in close collaboration with the involved groups, allowing for consistent and machine-readable metadata entry from the beginning of the research process. The open-source software eLabFTW, already used across UA Ruhr, will be centrally operated and integrated with ReSeeD to enable

seamless workflows from experiment documentation to data publication. This approach will foster reproducibility, cross-group collaboration, and research transparency at CRC.