Challenges and Solutions – A Case Report of Institutional Data Management at the Leibniz Centre for Tropical Marine Research (ZMT)

Birte Hemmelskamp-Pfeiffer, Philipp Gies, Alexandra Nozik, Finn Opätz, Sebastian Swirski

This work was supported by the German Research Foundation DFG under the grant agreement number 442032008 (NFDI4Biodiversity). The project is part of NFDI, the National Research Data Infrastructure Programme in Germany.

The ZMT use case #27 in the NFDI4Biodiversity focuses on institutional research data management. The aim is to identify challenges in dealing with research data, develop corresponding concepts and guidelines and make these available to the community.

Abstract

High-quality research data management (RDM) is becoming increasingly important at research institutions. This article discusses the challenges encountered by the Research Data Service (RDS) at the Leibniz Centre for Tropical Marine Research (ZMT) since its establishment and presents the solutions developed in response. Building on a demand-oriented approach and the development of research infrastructure over time, the article outlines ten key challenges and corresponding solutions for institutional data management at ZMT – ranging from collaboration and networking to storage and beyond.

Introduction

The Leibniz Centre for Tropical Marine Research (ZMT) conducts inter- and transdisciplinary research, in close cooperation with external partners to address critical challenges in tropical marine and coastal ecosystems. To foster equitable international collaboration, ZMT has established guidelines to ensure collaborations on equal terms with its partners (ZMT, 2025a) and ethical handling of transboundary data transfers (ZMT, 2022). Researchers at ZMT are supported by three science-enabling units – the diving centre, the laboratories and the Research Data Service (RDS), which together form an integral part of the institute's organizational infrastructure.

The RDS at ZMT was established in 2020 as an institutional initiative to advance best

practices in research data management (RDM[1]). Its foundation aligns with the overarching principles of the Leibniz Association, particularly in fostering openness, collaboration, and the responsible handling of research data. The RDS provides guidance, best practices, and infrastructure for secure and sustainable data storage along the whole research lifecycle, from project planning and data collection to ensuring long-term accessibility and re-usability of research data. Frameworks and guidance are provided to support the integration of the FAIR (data should be findable, accessible, interoperable, re-usable) (Wilkinson *et al.*, 2016) and CARE (Collective benefit, Authority to control, Responsibility, and Ethics) principles (Carroll *et al.*, 2020) into the management of institutional data across all disciplines, embedding them into workflows, training, and infrastructure. The RDS adopts a demand-driven approach, actively engaging with researchers to understand their requirements and address potential concerns regarding RDM. This is of major importance as individual researchers have different needs and requirements from RDM practices and policies compared to research institutions or funding bodies (Science Europe, 2021).

As high-quality RDM grows in importance, this paper aims to share focus points to consider when establishing research data infrastructures in institutions similar to those at ZMT. Based on the demand-driven approach and the development of the research infrastructure since the establishment of the RDS unit, this paper outlines 10 challenges of and solutions for institutional data management at ZMT:

1. Transparency, Visibility and Collaboration with Researchers
2. Internal and External Networks
3. Supporting Team: Disciplinary Diversity
4. Early-Stage Data Management Planning
5. Central Storage Solution and Organised Project Data Storage
6. Digital Tools and Automatised Transfer of Collected Data
7. Platform for Data-Intensive Computing
8. Incentives for Metadata Annotation
9. Searchable Database of Institutional Datasets
10. Research Data Offboarding

These points are not exhaustive, but have proven to be essential for the current state of

[1] RDM "is the process of providing appropriate labelling, storage, and access for data at all stages of a research project" (Harvard University, 2025).

development of the infrastructure unit.

1. Transparency, Visibility and Collaboration with Researchers

Challenge: Establishing an institutional unit dedicated to RDM entails a redefinition of the role and significance of RDM within an institute, and consequently necessitates adjustments in established research practices. This inevitably leads to challenges, scepticism, and in some cases rejection, but also curiosity and interest among researchers.

As a science-enabling unit at ZMT, the RDS is committed to providing demand-oriented services and tools tailored to the needs of ZMT researchers and their partners in tropical regions. In its initial phase, however, the group faced challenges in achieving institutional recognition and in communicating the value of its activities. In particular, it was necessary to foster an understanding of RDM as an integral and supportive component of the research process, rather than as an external or additional requirement.

Several measures have been implemented to enhance transparency and strengthen the visibility of the RDS within the institute. First, the dedicated presence on both the institute's homepage and intranet facilitates access to general information, while enabling researchers to obtain tailored consultation on specific questions. In addition, by providing timely and consistent advisory services, the RDS contributes to a supportive environment and helps to alleviate challenges commonly associated with data management. Furthermore, an internal workshop series on RDM, with particular emphasis on institute-specific practices, serves to regularly reinforce awareness and highlight the range of support available to researchers. The RDS workgroup members also engage with both formal and informal exchanges with researchers that are interested in data management, as well as active participation in research groups and programme area meetings. This further creates opportunities to identify cross-cutting themes and potential areas for cooperation. In turn, participating researchers contribute to the broader promotion of RDM by serving as multipliers within their respective groups.

2. Internal and External Networks

Challenge: The establishment of a new infrastructure unit such as the RDS does not automatically guarantee seamless integration with existing institutional (infrastructure) structures. Active efforts are required from all parties involved to identify points of intersection, foster networking, and build cooperative relationships that enable the development of efficient processes. Furthermore, it is crucial to establish and maintain networks with data management-related groups beyond the institute itself at local, national, and international

levels. Such networks provide valuable platforms for discussing shared challenges and exchanging knowledge on current developments in the field of RDM.

At ZMT, the RDS group leader has taken a pro-active role in establishing strong connections with other infrastructure units, including the IT department, laboratories, science management, and administration. Building such internal networks and fostering collaboration on change processes within existing structures requires carefully considered strategies, including sustained communication, trust-building, and gradual integration. These exchanges are essential for cultivating strong relationships and developing collaborative workflows. Demonstrating openness and respect toward established practices, in turn, fosters receptiveness and acceptance of change.

Beyond the institute, the RDS actively engages in networking within the Leibniz Association, the Deutsche Allianz Meeresforschung (DAM), the U Bremen Research Alliance (UBRA), and the National Research Data Infrastructure for Biodiversity (NFDI4Biodiversity), among others. These engagements ensure that ZMT's practices are aligned with national standards and benefit from community-driven developments. Additionally, regularly exchanging experiences within these networks makes it easier to find proven solutions for integrating the new infrastructure into institutional structures.

Expanding international networks is a strategic priority. Current activities include, among others, the integration of internal metadata systems with the Ocean Info Hub (OIH) and the development of research data management concepts for ZMT's tropical hubs.

3. Supporting Team: Disciplinary Diversity

Challenge: At a transdisciplinary institute like ZMT, RDM must accommodate a wide variety of methodologies, data formats, disciplinary cultures and growing demands. This requires support that is not only technically proficient, but also contextually aligned with the specific needs of different fields.

The RDS team addresses this challenge through a deliberately diverse team structure, led by experienced leadership that fosters strategic direction, coordination, and innovation. Two data managers – one specialising in RDM for the natural sciences and one for the social sciences – serve as direct points of contact for researchers, providing guidance in a discipline-oriented framework. The RDS team further includes a geodata specialist who is simultaneously engaged in a scientific working group. This dual role facilitates the testing of data collection tools in diverse research contexts, enabling validated tools to be incorporated into the RDS service portfolio. In addition, a research systems engineer supports the technical infrastructure, including the development and maintenance of globally accessible systems for ZMT

researchers and their partners. Finally, a student assistant supports the RDS in diverse topics, especially in programming and other technical tasks. The disciplinary diversity within the support team helps to bridge cultural differences across fields and ensures that the wide range of research approaches and data types at ZMT are met with tailored, discipline-appropriate solutions. Even as an interdisciplinary service unit, the RDS operates with limited resources and cannot take over researchers' manual tasks. Its role is to enable and promote best practices, both through collaborations and by responding to emerging demands.

4. Early-Stage Data Management Planning

Challenge: RDM is a dynamic field, shaped by evolving frameworks, principles, and policy requirements that researchers are expected to integrate from the outset of their project. Institutions must therefore remain flexible and continuously adapt their RDM guidance to emerging demands. One prominent example is the increasing requirement of Data Management Plans (DMPs), which have become mandatory in many funding schemes.

In response to this challenge, the RDS has implemented a range of measures to support the creation of DMPs:

- Embedding the requirement to prepare a DMP for all projects within the ZMT Research Data Policy (ZMT, 2025b);
- Providing guidance on data management planning through the ZMT intranet;
- Offering templates with illustrative examples;
- Conducting regular workshops on the preparation of DMPs;
- Offering individual and timely consultation;
- Integrating DMPs in the internal review process for third-party funding proposals.

Through this combination of policy anchoring, structured resources, and personalised support, the RDS ensures that DMPs are both systematically embedded in institutional practice and meaningfully aligned with the needs of researchers.

The ROMPi model (Della Chiesa, 2024) is currently being evaluated as a possible expansion for further improvement. In contrast to traditional DMPs, which are often perceived as primarily bureaucratic, the ROMPi model emphasises operational management and planning activities. Specifically, it focuses on integrating individual workflows, managing research outputs, and addressing the specific needs of researchers. This approach aligns closely with the RDS's objective of providing tailored, demand-oriented services at ZMT.

5. Central Storage Solution and Organised Project Data Storage

Challenge: Over time, research institutes often develop heterogeneous storage solutions shaped by preferences of individual research groups. Such practices frequently result in fragmented and inefficient management of institutional research data. To address this challenge, it is essential to provide reliable, scalable, and user-oriented storage and collaboration infrastructures.

At ZMT, the RDS has implemented a standardised, accessible, and secure storage system supported by comprehensive back-up procedures and an integrated collaborative cloud platform for ZMT researchers and their partners in the tropics. For each project, a dedicated folder with a pre-defined sub-folder structure is provided. The structure facilitates systematic organisation of project data while maintaining sufficient flexibility to accommodate disciplinary differences. Collaboration is enabled through the creation of partner user accounts or the option to share certain project folders or subfolders. This allows partners in the tropics to participate directly in collaborative research projects. Beyond research data, the cloud also supports the storage and dissemination of teaching and mentoring materials, thereby serving as a comprehensive cloud environment for all science-related activities at ZMT.

In line with the ZMT Research Data Policy, researchers are expected to use the storage and collaboration services provided by the RDS. The corresponding Data Management Guideline (ZMT, 2025c) defines procedures for accessing accounts and project folders, as well as for managing data sharing. To support their use, the RDS offers training workshops that highlight collaborative functions and practical applications of the cloud platform.

6. Digital Tools and Automatised Transfer of Collected Data

Challenge: Fieldwork in tropical coastal regions, typically undertaken by ZMT researchers, is often long-term, remote, and resource-intensive. It must be carefully prepared while remaining adaptable to changing conditions. When data is recorded on paper or stored locally on a device, it remains disconnected from the research data lifecycle, as integration and analysis occur only after fieldwork is completed. This delays collaboration with off-site project partners and increases the risk of data errors or organisational challenges becoming evident only weeks or months later.

The use of digital data collection tools with synchronisation and collaboration options facilitates the systematic structuring of data and metadata from the outset of a research project, improving both efficiency and data quality. However, digitalisation is not a universal solution: all data collection methods involve inherent challenges, and no single tool can meet

the diverse requirements of an interdisciplinary research environment such as ZMT. Careful assessment of project-specific needs is therefore essential. To support this process, the RDS provides guidance based on two principles:

1. Pre-fieldwork consultation: Early engagement with researchers to define dataset structures and identify suitable tools.
2. Accessibility and sustainability: Preference for free or low-cost tools that are accessible to international partners and supported by open documentation.

In practice, the RDS primarily recommends and trains researchers in the use of KoboToolbox, a survey tool for remote fieldwork, and QField for QGIS, which is tailored to projects involving geospatial data.

7. Platform for Data-Intensive Computing

Challenge: Scientific research frequently involves highly data-intensive workflows, ranging from the annotation of entire genomes to the modelling of complex natural phenomena and the application to AI-driven methods for species classification in ecosystems. These tasks demand computational power and storage capacities that far exceed the capabilities of standard office PCs or laptops.

To address this challenge, the RDS has established an integrated in-house computing environment designed to enable ZMT researchers and their partners in tropical regions to collaboratively process complex datasets, independent of geographical location.

The computing environment is built around three core components:

1. Remote Linux Desktop Environment: Provides worldwide access to a Linux-based desktop with a wide range of open-source tools for data analyses.
2. Distributed Computing Cluster: Allows the processing of large-scale datasets on local servers through demand-based scheduling of computational jobs.
3. Collaboration and Documentation Platform: Serves as a central space for project discussion, code documentation and the design of data processing pipelines.

Researchers work in project-based shared folders on a central storage system, enabling secure data management and exchange with external partners. While raw and final, fully processed datasets are preserved in the cloud, the computing environment functions as a *digital laboratory bench*: a temporary workspace for active data processing. Small-scale tasks can be

executed directly on desktop nodes, whereas computationally intensive jobs are submitted via command line to the attached cluster. A demand-based scheduling system allocates resources fairly, preventing individual users from monopolising the cluster while still permitting full access during idle periods. In addition, the environment supports collaboration through integrated chat, documentation of workflows and code, and local version control, effectively serving as a central digital lab notebook.

8. Incentives for Metadata Annotation

Challenge: Many researchers are still unfamiliar with metadata annotation and do not integrate it as an essential part of their workflow. Additionally, adding metadata is often considered time-consuming and tedious.

The ZMT Research Data Policy emphasises early linkage of documentation and metadata to research data (ZMT, 2025b). To support this, the RDS provides a comprehensive guide, discipline-specific recommendations, and established connections with data repositories such as PANGAEA and Qualiservice.

Although awareness of the importance of metadata annotation is increasing, its systematic collection remains challenging. To address this, the RDS provides a user-friendly in-house metadata editor that structures annotation through dedicated tabs ("Basics," "Authors," "Projects," "Events") and generates JSON files linking essential metadata to datasets. A central database under development will further streamline the process by enabling standardised selections, such as author or project names.

While metadata should ideally be created during the research process, it is often added only during the offboarding process or during repository submission. To promote earlier adoption, the RDS supports the annotation of raw laboratory data stored in the internal cloud systems and has extended its metadata editor to include discipline-specific fields. Planned developments focus on interoperability with publishers such as PANGAEA to enable automated metadata transfer, thereby reducing redundancy and inconsistencies. By establishing standardised procedures, offering targeted training, and fostering community exchange, the RDS advances metadata practices and contributes to the overall quality assurance of research data at ZMT.

9. Searchable Database of Institutional Datasets

Challenge: The increasing adoption of open science practices requires research institutions to systematically track, manage, and promote datasets produced by their researchers. This task is complicated by the dispersion of data across a variety of general-purpose and domain-

specific repositories, as well as by the persistence of unpublished data resulting from legal, ethical, or technical constraints (e.g. anonymisation requirements), or from limited awareness of data publication practices in the past.

In the absence of an institutional repository or a Current Research Information System (CRIS) in the past, the identification and monitoring of datasets – particularly those from earlier projects – has been a challenging task for the RDS. The resulting lack of systematic oversight reduces institutional visibility and constrains the reuse of data assets for knowledge exchange. To address this gap, the ZMT DataPortal has been developed as a secure, searchable database of datasets associated with ZMT researchers, independent of their storage location; i.e. data may reside on the internal cloud infrastructure or in external data repositories like PANGAEA. The portal is built on the open-source platform LinkAhead in collaboration with IndiScale GmbH. A significant portion of the metadata within the ZMT DataPortal has been harvested from PANGAEA through an automated crawler. Development is currently underway to extend this functionality to include metadata from datasets stored in the internal cloud system as well as other repositories.

10. Research Data Offboarding

Challenge: During the transition to a centralised storage solution, not all research data has yet been systematically transferred to the institute's central servers, and metadata is often missing. This challenge becomes particularly evident during the offboarding process, which includes a review by the RDS.

Researchers leaving the institute are required to complete a comprehensive offboarding procedure. As part of this process, they must ensure that all research data generated during their time at ZMT is deposited on the institute's internal servers and/or that access is provided to already published data. Because many of the related tools – such as the metadata editor – have only recently been introduced, they are often not yet fully integrated into the researchers' workflows.

The RDS is currently standardising its offboarding procedures in order to streamline the process. To support departing researchers and students, the RDS offers guidance through dedicated meetings and consultations. It is therefore strongly recommended that individuals notify the RDS as early as possible once their departure from ZMT is foreseeable. This allows for a collaborative identification of datasets requiring storage, the provision of advice on metadata annotation where necessary, and, by means of the metadata editor, the systematic registration of both published and unpublished datasets in the ZMT DataPortal.

Summary

This paper presents the challenges and solutions encountered in establishing and developing the Research Data Service (RDS) at the Leibniz Centre for Tropical Marine Research (ZMT). Drawing on a demand-driven approach, the RDS has implemented measures to promote transparency, provide discipline-specific support, and integrate best practices such as FAIR principles and early-stage data management planning into institutional workflows. The team's disciplinary diversity ensures tailored solutions for a wide range of research methods and data types, while centralised infrastructure and targeted training support the entire research data lifecycle. By standardising procedures, providing discipline-specific advice, and offering targeted training, the RDS contributes significantly to the overall quality assurance of research data at ZMT. Beyond supporting individual projects, the RDS strengthens the institute's overall research capacity, reputation, and compliance with evolving data policies, positioning ZMT within national and international research data networks. Looking forward, the RDS aims to further integrate emerging digital tools, expand engagement in national initiatives such as NFDI, and continuously adapt its services to evolving scientific and policy requirements.

References

Carroll, S.R., Garba, I., Figueroa-Rodríguez, O.L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J.D., Anderson, J. and Hudson, M. (2020). The CARE Principles for Indigenous Data Governance. *Data Science Journal*, *19*(1), 43. https://doi.org/10.5334/dsj-2020-043

Della Chiesa, S. (2024). Research Output Lifecycle - Guidance for the Research Output Management Planning (ROMPi) (3.1). Zenodo. https://doi.org/10.5281/zenodo.11394846

Harvard University (2025, September 05) What is Research Data Management. https://datamanagement.hms.harvard.edu/about/what-research-data-management

ZMT, Leibniz Centre for Tropical Marine Research (2022) Guidelines for Safeguarding Good Scientific Practice at the Leibniz Centre for Tropical Marine Research (ZMT). https://www.leibniz-zmt.de/images/content/pdf/Interne_Gremien/ZMT_GSP.pdf

ZMT, Leibniz Centre for Tropical Marine Research (2025a) Bremen Criteria (2.0). https://www.leibniz-zmt.de/images/content/pdf/OKE_Office_Knowledge_Exchange/Bremer_Kriterien_Deutsch_September_2025.pdf

ZMT, Leibniz Centre for Tropical Marine Research (2025b) ZMT Research Data Policy (1.0). Zenodo. https://doi.org/10.5281/zenodo.14672213

ZMT, Leibniz Centre for Tropical Marine Research (2025c) Research Data Management Guideline (1.0). https://intranet.leibniz-zmt.de/share/page/

Science Europe (2021), Practical Guide to the International Alignment of Research Data Management - Extended Edition. https://doi.org/10.5281/zenodo.4915861

Wilkinson, M., Dumontier, M., Aalbersberg, I., Appleton, G., Axton, M., Baak, A., Blomberg6, N., Boiten, J-W., Bonino da Silva Santos, L., 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. (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data,* 3, Article 160018. https://doi.org/10.1038/sdata.2016.18

Use of AI to polish written content.