

# Research Data Policy at MARUM – Center for Marine Environmental Sciences of the University of Bremen

Adopted by the MARUM Center Council on 27.10.2021

## Scope

This policy applies to all publicly funded research projects, hosted by MARUM, including student projects, which have or will generate and/or work with research data. It extends the corresponding guidelines of the University of Bremen. Where conflicting guidelines are provided by specific funding bodies, the more specific guidelines shall apply.

## Objective

The aim of this policy is to advance the handling of research data at MARUM in the context of the digital strategy of the University of Bremen, in particular to preserve future use, ensure reproducibility of generated data, and facilitate value-added access and analysis of publicly funded research results. The policy should ensure the findability, accessibility, interoperability and reusability principles (FAIR) [1] and quality [2] of the collected research data, in order to meet the requirements of national and international publication bodies and research funding agencies [3, 4, 5] as well as the German National Research Data Infrastructure (NFDI).

## Definitions

- Research data refers to all raw data, primary data and data products that represent added value for research. All research data must be provided with descriptive metadata.
- Raw data is data obtained directly from sensors, devices, or manually, unchecked, un-encoded, uncompressed, unformatted and not subject to error correction.
- Primary data are processed data obtained from raw data, calibrated, quality- and error-tested, which form the basis for further research.
- Data products are data derived from raw or primary data, which may be combined with other data and/or have undergone further quality and calibration steps.
- Metadata is structured data about data, i.e. information about characteristics of other data or descriptions of data. Metadata or meta-information enables the archiving and retrieval of data. All data collections and databases require metadata for structuring.
- Research software: research data also includes software code that transforms, processes or analyses data. To make research data fully reproducible, research software should be archived and versioned in relevant open long-term repositories.
- MARUM data sets - research data is considered a 'MARUM data set' if it is publicly funded and collected, procured or prepared by MARUM members during their regular work, a project, training or further qualification at the MARUM or by external users and associated partners, students and guests, substantially using the resources/infrastructure of the MARUM for the purpose of their own research. The latter concerns the entire data life cycle from raw data, including metadata, primary data to derived data products.

## Principles

MARUM is committed to long-term, sustainable preservation and reuse of research data for science and society. It recognizes the necessity of the additional effort required for research-data management and will support this structurally (e.g. through the U Bremen Research Alliance Data Stewardship Network) as well as through scientific recognition. MARUM is committed to the principles of open access to research data [6, 7] and the FAIR data principle [1].

This research-data policy applies to all MARUM data sets. Irrespective of the rights of disposal and use regulated in this data policy, persons who have produced research data are entitled to moral and intellectual rights of use of such data. It must be possible for them to use the data in a way that secures scientific recognition of their work, and to complete their projects (e.g. qualification work) within a reasonable time period. Decisions on the dissemination of research data are made with due regard to the conventions of the relevant subject area by the researchers. Once a decision has been made to make research data publicly available, researchers describe them completely and comprehensively in accordance with the DFG Guidelines for Safeguarding Good Research Practice, Code of Conduct [6].

Research-data management in the framework of this policy concerns the handling of MARUM data sets from data generation, quality control and analysis, documentation and archiving. This includes that MARUM data sets generally have to be archived in long-term repositories (e.g. [www.pangaea.de](http://www.pangaea.de)) in accordance with the DFG Guidelines for Safeguarding Good Research Practice, Code of Conduct [6] and the FAIR data principles [1] and made citable by data publication. Furthermore, data and processes in the context of their generation must be documented in a subject-specific manner and provided with metadata. Clarity regarding the provenance of the data must be ensured by describing data transformations. **Data management plans** (DMP) ensure the documentation of these processes and the description of the data. MARUM supports its organizational units and projects in developing standardized data management plans at the level of its Research Areas.

## **Responsibility**

The responsibility for the scientific quality and the careful handling of the data of all research activities and infrastructure facilities at MARUM lies with all persons involved in generating or using research data. The guidelines and subject-specific recommendations of the DFG or other funding bodies apply. The MARUM Center Council (Zentrumsrat) is responsible for creating the framework conditions for safeguarding, providing and fair allocation of research data. This includes the appreciation of the publication of quality-controlled research data as a scientific achievement, as well as agreeing on rules for dealing with research data at MARUM and ensuring their implementation.

In daily operations, responsibility for the quality, usability and preservation of research data is assumed by individual researchers. This applies to all scientific projects, dissertations etc. Subject-specific DMPs are provided by heads of the involved working groups.

With the acceptance of data by relevant recognized long-term repositories (e.g. PANGAEA), the responsibility for their preservation can be passed on to them. The researcher is responsible that the terms and conditions of the repository are compatible with this research-data policy.

If data are archived, but not made initially publicly available (embargoed), the disposition/determination of the data is transferred from the submitting party to the MARUM director [7].

Researchers are obliged to make all unpublished research data accessible to project and working group members at least three months before they leave their employment. If research data remain in personal directories, the currently valid service agreement on the use of MARUM IT applies.

## **Archiving and Publication**

In accordance with national and international organizations for the promotion and execution of research, MARUM is committed to long-term archiving and open access to research data from publicly funded research [5, 7]. Following current DFG guidelines, MARUM strives to publish at least the primary research data, including essential metadata as quickly as possible according to subject-specific recommendations.

All primary data, data products and MARUM data sets must be deposited in a publicly accessible, citable, long-term repository no later than two years after their generation, with a standardized license. The deposited data may be subject to an embargo for a maximum of two further years. After expiry of the embargo period, the data must be made public immediately and actively following the FAIR data principles. This rule also applies, as far as possible, retroactively to all primary data collected at MARUM prior to the adoption of the research data directive.

Deviating embargo periods from data generation for exclusive first use and for scientific validation and quality control can be set in project-specific data management plans with appropriate justification. This also applies if the confidentiality of research data, at least temporarily, is an indispensable prerequisite for later commercialization. Details are regulated by the specifications of the respective funding programs. Commercialization of research data is the exception; it must be applied for at an early stage, and a utilization plan must be submitted to the MARUM board of directors. When setting embargo periods, legal regulations are required, considering scientific interests and contractual agreements with cooperation partners and, where appropriate, exploitation interests. Embargo periods deviating from the 2 plus 2 years rule must be defined per project in the DMP and approved by the MARUM board of directors.

In determining and justifying the duration of the embargo periods, MARUM is guided by the corresponding specifications of professional societies, large research associations and funding bodies.

### **Quality Assurance**

For the traceability and re-use of research data it is necessary to record metadata, to document the context in which the data was generated, and the tools or software that was used to generate, calibrate or otherwise transform the data. Accordingly, in the generation of data, in addition to the description of the procedure of data generation itself, framework parameters (as metadata) must be recorded, which enable statements to be made about the origin, transformation and quality of the data in a standardized form (2). Which metadata must be collected for this purpose depends on the specific research plan/project in each case and should, in any case, comply with the FAIR data principles. Information for quality assurance and evaluation can be recorded in the form of laboratory books, but also in the documentation of data-generating processes in jointly created "Standard Operating Procedures" (SOPs) that are continuously coordinated within the discipline-specific research community also at the international level.

Within the framework of modern ("digital") science, preference should be given to the explicit and machine-readable coding of this information, and this should be implemented in accordance with the **FAIR data principles**.

Information on data formats used must be part of an extensive metadata set. If possible, open and free data formats should be used, since data should remain usable even after the originally used application has been discontinued. For subsequent use, especially with digital methods, care must be taken that quality-assuring metadata is stored alongside the research data, if possible in digitized form and programmatically accessible.

### **Scientific Recognition**

The publication of quality-assured research data is a central and indispensable service in the research process, which benefits science - and indirectly society as a whole. MARUM will promote a sustainable recognition of the associated scientific achievements both nationally and internationally.

The generation of research data is still often considered to be secondary to their analysis. This differentiation is no longer appropriate in view of the skills needed to generate and process data. Improved recognition of scientific achievement, which is expressed in the generation of

research data, therefore begins with the development of a new view of all work processes and the people involved in each of them, who can only realize excellent research results in a cooperative effort. MARUM recognizes this additional effort as part of its research performance and will promote this at the national and international levels.

When using data records of third parties, the obligation to cite and, if applicable, to offer co-authorship applies, based on the DFG rules of good scientific practice. Authors of scientific publications are always jointly responsible for their content. However, only those who have made a significant contribution to a scientific publication are considered to be authors. So-called “honorary authorship” is excluded in accordance with DFG rules of good scientific practice [6, 7].

### **Long-Term Availability**

MARUM data sets must be archived and published in suitable, sustainably operated, long-term repositories. When archiving and, if necessary, making research data accessible, the implications arising from applicable law or contractually based rights of third parties must also be considered.

If for plausible reasons, the recommended PANGAEA repository cannot be used, the alternatively selected repositories must be comparably qualified and interoperable with the standards and practices of national and international research disciplines and communities. The basis for archiving should be certified (e.g. Core Trust Seal) data repositories that can guarantee citation via a complete data citation with handle/DOI since this can be used as a standardized procedure for the application of one of the most important evaluation criteria in the scientific community. These repositories should comply with the FAIR principles.

### **Archival of data transformation and processing steps (Software)**

A key requirement to ensure reproducibility of research data is the full documentation of processing and transformation steps. This increasingly includes the use of specific software models or code, and/or scripting steps that should be versioned and archived. Thus, research data management also includes suitable archiving of software, scripts, and other data transformation and processing steps. Documentation, long-term preservation and citeability should be ensured through deposition in suitable open long-term repositories (e.g., the Zenodo repository run by CERN).

### **Qualification**

The working groups and infrastructure areas at MARUM aim to train personnel as Data Stewards. MARUM will support the necessary qualification. The scientists and the science supporting staff inform themselves comprehensively about qualification and reputational aspects in relation to data and use the options of data publications and citable archives as a contribution to sustainable science in research, infrastructure, teaching and transfer.

### **Legal Issues**

Any kind of disposal of research data and, if applicable, the mere handling of data makes it necessary for those responsible to consider the relevant legal context. In case of doubt, the Chief Digital Officer and/or the Data Protection Officer of the University of Bremen should be consulted.

Both making research data accessible in the sense of Open Science [8] and its commercial use require at least an examination of the necessary authorization to dispose of the data. In doing so, the labor law, the German Employee Invention Act and the Constitution (“Freedom of Science”), among others, must be observed. Questions of ownership law may also include whether the content to be archived is protected by copyright, whether it is a trade or business secret, or with which license a given content may be exclusively passed on or published. Further legal frameworks can result from fields as diverse as data protection or export controls. In general, there is an obligation to provide each data record with license information (licensing

or usage regulations). It is recommended to use the Creative Commons licenses [9], and especially the CC0 license for metadata and CC-BY license for data.

## References

1. Wilkinson, M. D., M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3:160018.
2. Rat für Informationsinfrastrukturen (RfII), Herausforderung Datenqualität 2019, <http://www.rfii.de/download/herausforderung-datenqualitaet-november-2019/>
3. Deutsche Forschungsgemeinschaft (DFG), Umgang mit Forschungsdaten, aufgerufen 2021, [https://www.dfg.de/foerderung/antrag\\_gutachter\\_gremien/antragstellende/nachnutzung\\_forschungsdaten/](https://www.dfg.de/foerderung/antrag_gutachter_gremien/antragstellende/nachnutzung_forschungsdaten/)
4. Digital Curation Center Data Policies, aufgerufen 2021, <https://www.dcc.ac.uk/guidance/policy/funders-data-policies>
5. forschungsdaten.org Data Policies, aufgerufen 2021, [https://www.forschungsdaten.org/index.php/Data\\_Policies](https://www.forschungsdaten.org/index.php/Data_Policies)
6. Deutsche Forschungsgemeinschaft (DFG), Wissenschaftliche Integrität, aufgerufen 2021, <https://wissenschaftliche-integritaet.de>
7. Guidelines to ensure Good Scientific Practice and prevent scientific misconduct at MARUM – Center for Marine Environmental Sciences of the University of Bremen, 2020, <https://www.marum.de/Binaries/Binary20323/GSP-Guidelines-MARUM.pdf>
8. Bundesministerium für Bildung und Forschung, Open Access in Deutschland, aufgerufen 2021, [https://www.bmbf.de/upload\\_filestore/pub/Open\\_Access\\_in\\_Deutschland.pdf](https://www.bmbf.de/upload_filestore/pub/Open_Access_in_Deutschland.pdf)
9. Creative Commons Lizenzen, aufgerufen 2021, <https://creativecommons.org/>