

Leibniz Centre for Tropical Marine Research (ZMT) GmbH

Open Data Policy

Agreed on 28 September 2015

Preamble

Part of the ZMT identity is to share scientific research results with partner institutions in the tropics and beyond. This idea was adopted during the evaluation of ZMT in 2013 and led to the recommendation to provide the **scientific community with research data of the institute by using Open Access**. ZMT follows this recommendation and increasingly aims at sharing its original data with partner organisations and other Leibniz institutions. Further plans include intensifying the use of information infrastructures available within the Leibniz Association. As a basis for the implementation of this recommendation, ZMT developed this Open Data Policy which was adopted in autumn 2015.

The publication of research results in digital and freely available form does not only include specialist publications but also original data.¹ Quality-checked research data are a basic pillar of scientific findings and can frequently provide the basis for further research projects, regardless of the original purpose for which they have been collected. Hence, the **sustainable storage and provision of research data** serves the purpose of gaining future research results.

Management of Research Data at ZMT

In 2012, ZMT introduced a central "**Archive of Scientific Data**" which can currently be accessed via the ZMT intranet. ZMT explicitly encourages its scientists to provide research data in this research data repository. In this way, the permanent availability of data is

¹ Cf. Leitlinie zu Open Access der Leibniz-Gemeinschaft, agreed on 23.11.2007; URL: http://www.leibniz-gemeinschaft.de/fileadmin/user_upload/downloads/Infrastruktur/OpenAccess-Leitlinie.pdf (last updated 19.08.2015)

guaranteed pursuant to recommendations of the Deutsche Forschungsgemeinschaft (DFG) (German Research Foundation) for ensuring good scientific practice.²

Among other things, the DFG recommends that "primary data as basis for publications **[shall] be saved for ten years [...] on durable and secured data carriers within the institution in which they were collected**".³ Furthermore, project proposals at the DFG and the EU require a specification regarding the planned data management.^{4,5}

A reliable management of research data must meet a variety of technical and organisational requirements which are defined and implemented by the scientific staff and the institutions supporting scientific work - particularly IT and library.⁶

Each scientist is responsible for storing the original data which form the basis of a project or publication. Any relevant meta data which are necessary to understand and process the data, must be specified. Ideally, a data management plan is developed.

During the project planning phase, the most important definitions for the handling of research data are summarised in a **data management plan**. This process starts with evaluating and classifying the amounts of data which will be generated. Furthermore, information must be provided on accuracy, integrity, authenticity and confidentiality of the data.

By entering the data into the "Archive of Scientific Data", the scientist grants ZMT the right to make the relevant meta data available online and free of charge. The publication of data can be withheld until expiry of disclosure deadlines.

Research data will be stored for an unlimited period of time. The data shall be entered into the archive of scientific data simultaneously with the release of the relevant publication, if possible, however, not later than on expiry of the project or on termination of employment at ZMT.

² Cf. DFG, Sicherung guter wissenschaftlicher Praxis, ergänzte und aktualisierte Denkschrift, agreed on 03.07.2013; URL: http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf (last downloaded 19.08.2015)

³ DFG, Sicherung guter wissenschaftlicher Praxis, ergänzte und aktualisierte Denkschrift, agreed on 03.07.2013, p. 21.

⁴ Cf. DFG-form: Leitfaden für die Antragstellung (54.01 - 06/2014), p. 5-6; URL: http://www.dfg.de/formulare/54_01/54_01_de.pdf (last downloaded 19.08.2015)

⁵ Cf. Guidelines on Data Management in Horizon 2020, version 1.0, 11.12.2013; URL: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf (last downloaded 19.08.2015)

⁶ Cf. <http://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten/grundsaeetze/> (last downloaded 13.08.2015)

Apart from research data collected during research work at ZMT, **additional data and products shall be made available to ZMT for unlimited and easy use.** In summary, the following products are intended for archiving and easy usage:

- specialist publications in PDF-form for internal storage
- meta data belonging to the research data of the publication
- research data not published, if any
- final papers (bachelor, master and Ph.D. theses)
- photographic material taken in connection with research work
- video recording prepared in connection with research work

The scientist is responsible for ensuring that the release of data does not infringe copyrights or on personal data and the interests of third parties, if any. For any questions, please contact the administration department.

Check list for the management of research data

Without defined work flows for data management, only small amounts of data with unclear quality can be secured. For the appropriate planning of a research project we therefore recommend to apply a checklist:

- **Planning and preparation***(organisation by the scientist)*
 - Have you documented all underlying data of the project (project name, funding organisations, term, project leader, goals)?
 - How can existing data be integrated/used subsequently?
 - What meaning do the data have for the project goals?
 - How can used/generated data be characterised (data types, formats, reproducibility)?
 - In what way do you collect / produce data?
 - How large is the amount of data / rate of production?
 - How do you control quality?
- **Choice and storage period** *(organisation by the scientist)*
 - What is the reason for storing which data?
 - Who selects the data, at what time and by what means?
- **Data input and assumption of responsibility***(Organisation by scientist upon consultation with the IT-department)*
 - At what time and in what way are data delivered/transferred?
 - Who collects what kind of meta data at what time?
 - How do we validate data and meta data?
 - How do we handle sensitive data?
- **Storage and infrastructure***(organisation by IT-department upon consultation with the scientist)*
 - Who is responsible for storing the data during and after the project?
 - What technologies and locations are used to store the data?

- Are there any special infrastructure requirements for data transfer, access and use? (network bandwidth, hardware etc.)
- **Preservation measures** (*organisation by the scientist in cooperation with the IT-department*)
 - Have you documented technologies used, dependencies, target groups and requirements for use?
 - Do you check regularly whether requirements, available technologies or dependencies have changed?
 - Do you check regularly whether the data are worth being stored?
 - Is there a succession plan if the storing institution changes?
- **Access and use** (*Organisation by the scientist in cooperation with the IT-department*)
 - Are we obliged to release data or to withhold release of data?
 - What means are used to realise publication, search options and access?
 - For what purpose will the data be used and what software will presumably be applied?
 - In what way will you enforce first rights of use, restrictions of use or license terms?
- **Management, organisation and policies** (*organisation by the scientists*)
 - Have all parties agreed and have their contributions been defined?
 - Have you described the data management procedure?
 - Have you described and estimated all resources?
 - Which requirements/provisions/policies must be observed and implemented by what persons?
 - In what way will you check compliance?
- **Legal aspects of research data** (*control by the scientist in consultation with the representative for research data, if appropriate, notification of IT-department*)
 - Are the data sensitive due to data protection laws or other reasons.
 - Will you use external data or software which are subject to any copyrights, patents or similar rights?
 - Have you clarified the terms of use and rights for your own intellectual property?
 - Do you have to observe any expiring terms of protection?
- **Meta data** (*Organisation by the scientist*)
 - What are the conditions for processing and understanding the meta data?
- **Identifiers and information objects** (*control by the scientist, in consultation with the representative for research data, if appropriate, notification of IT-department*)
 - What is the relation between the information objects?
 - Which information objects require permanent identifiers and which ones are used (DOI, URN, Handle etc.)?
 - Who is responsible for maintaining and updating identifiers?

For comparison, please refer to the checklist for management of research data of the German WissGrid-Project.⁷ More information is available in examples and check lists of the British Digital Curation Centre (DCC).^{8,9}

In compliance with this Open Data Policy, the web-based application of the research data base shall enable external access to research data in the future. The implementation will include access rights of different levels which will be entered by the scientist who supplies the data:

- Full access to all meta data and data without authentication.
- Access to meta data incl. description. Full access only upon authentication allowing for recording of who used which data at what time.
- Access to meta data incl. description. Full access only upon owner's approval. A corresponding request is possible upon authentication.

Transmission of confidential data to cooperation partners via Cloud services of external providers must be regarded critically. Preference should be given to the internal Cloud service ZMT-Owncloud (<https://zmtcloud.zmt-bremen.de/owncloud/>)

Structure of the Research Data Base

For archiving of research data, all relevant meta data must be specified. Depending on the research discipline this usually includes the following attributes:

- **Project title** (title of the project in which the data were collected, max. 200 characters in English)
- **Short title/acronym** (title of the project in which the data were collected, max. 200 characters in English)
- **Contact person** (name of the scientist coordinating the project)
- **Participating scientists** (names of the scientists involved in the project)
- **Department/work group**(name of the responsible department and work group)
- **Key words** (searchable descriptors to specify the contents of the data, max 10)
- **Taxon** (specification of the Latin name of species involved, max. 10)
- **Type of study** (specification of the type of study, e.g. experiment, survey, field study)
- **Location** (specification of the city, country, region)
- **Coordinate 1** (coordinate 1 of the location where the data were collected)

⁷ Cf. WissGrid: Checkliste zum Forschungsdaten-Management, 2013, S. 83-97; URL: http://www.wissgrid.de/publikationen/Leitfaden_Data-Management-WissGrid.pdf (last downloaded 19.08.2015)

⁸ Cf. DCC: Data plan guidance and examples, 2013; URL: <http://www.dcc.ac.uk/resources/data-management-plans/guidance-examples> (last downloaded 19.08.2015)

⁹ Cf. DCC: Checklist for a Data Management Plan, Version 4.0, 2013; URL: http://www.dcc.ac.uk/sites/default/files/documents/resource/DMP/DMP_Checklist_2013.pdf (last downloaded 19.08.2015)

- **Coordinate 2** (coordinate 2 of the location where the data were collected)
- **Start of study** (standardised date format)
- **End of study** (standardised date format)
- **Data context** (e.g. Ph.D. project, master thesis)
- **Abstract** (summary of the project in which the data were collected)
- **Methodological description** (description of all materials, measuring instruments and method parameters used so that the experiment can be reproduced on the basis of the description)
- **Files** (at least raw data together with meta data which are necessary to use the actual data: information structure, contents etc. and other useful data like GIS files etc.)
- **Publications** which are based on these data (upload as PDF-file is possible)

For the purpose of **long-term storage and transmission of archived data** storage should be effected in open standard formats, at least, however, in formats used and accepted by the majority of the research community.

For a choice of suitable file formats, please refer to the list on the portal Forschungsdaten-Bildung (establishment of research data)¹⁰. Depending on the type of file we recommend:

Quantitative data (statistics software formats)

Preferred formats:

- STATA (*.dta)
- R (*.R)
- SAS Transport (*.sas)
- proprietary formats of statistical packages, e.g. SPSS (*.sav), Stata (*.dta)
- text files (CSV) separated by tabulators, commas or columns with additional data definition file (setup, syntax, command file)

Accepted formats:

- OpenDocument table documents (*.ods), MS Excel (*.xls, *.xlsx), MS Access (*.mdb, *.accdb)
- CSV formats without additional data definition file
- Column binary formats or card image formats

Text files

Preferred formats:

- PDF/A (*.pdf)
- Plain text formats (ASCII)

Accepted formats:

- OpenDocument Text (*.odt)
- PDF (*.pdf)
- MS Word (*.doc, *.docx)
- RichTextFormat (*.rtf)
- HTML (*.htm, *.html)

¹⁰ Cf. Recommended file formats for transmitting research data files; URL:<http://www.for-schungsdaten-bildung.de/formate> (last downloaded 22.09.2015)

Images

Preferred formats:

- TIFF version 6 uncompressed (*.tif)

Accepted formats:

- JPEG (*.jpg, *jpeg), PNG (*.png), GIF (*.gif), BMP (*.bmp)
- PDF/A, PDF (*.pdf)

Audio

Preferred formats:

- MPEG-1 Audio Layer 3 (*.mp3)

Accepted formats:

- Wave Audio Format WAV (*.wav)

Video

Preferred formats:

- MPEG-4 (*.mp4), MPEG-2 (*.mpg)

Accepted formats:

- Audio Video Interleave AVI (*.avi)
- Windows Media Video WMV (*.wmv)

Data security of personal and sensitive data is guaranteed. These data are not visible for all users and can be made available on request, if necessary.

In the future, the "Archive of Scientific Data" will provide a Digital Object Identifier system, **DOI-system**, so that a long-term access to research data is guaranteed. The DOI identifies research data directly, is assigned to them permanently and is suitable for the indication of reference.

Contact persons at ZMT are Hauke Reuter (scientific support), Thomas Rau (IT support) and Christina Schrader (media support).