

Fieldwork



Before...

Data management already starts before data collection in the field. A **data management plan (DMP)** is a good way to think through and document the data life cycle, including a sampling strategy, anticipated data formats, possible storages for long-term preservation, a publishing strategy and a data backup plan during data collection and processing. Prepare standardized metadata sheets before going to the field. This will make it easier for you to integrate metadata documentation into your workflow.

During...

Fieldwork time is limited - thorough preparation pays off. Fieldwork is the phase of data collection, however already thinking about the next steps in the data life cycle can reduce workload and save extra work back home. Making use of **best practices** for sampling may help you to make your results comparable to already published research data. The usage of standardized metadata sampling sheets will prevent you from forgetting to document important information. Since it is easy to lose data in the field, a regular backup of all data (incl. field notes) and metadata is crucial.

After...

Back home, data processing and analysis are the next steps. At this stage, sharing data with your colleagues from your research project and converting the data in **open data formats** (formats which can be read without a proprietary software) will enhance their usability and usage. Do not forget to set up a backup system and use versioning while processing data.

Long-term data preservation and data publication are the last two major steps. Data repositories assign persistent identifiers to published datasets, which allow tracking the use of your data. The relevance and value of your data might rise enormously when used by your colleagues and further researchers. Embedded in a different scientific context, your research data may help answering scientific questions well beyond the scope of the initial research project.

Buzzwords and services

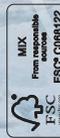
FAIR	Findable, Accessible, Interoperable, Reusable
DCC	Digital Curation Centre
DMP	Data Management Plan
DOI	Digital Object Identifier
OpenAIRE	European Open Research Data Pilot
RDA	Research Data Alliance
Re3data	Searchable online catalogue for data repositories



Cartoons by Auke Herrema
CC by www.ms.gerry.com/www.aukeherrema.nl

Photo by: Stephan Bernberg, Aarhus University

CC BY-NC-SA



Links and References

Digital Curation Centre (DCC): <http://www.dcc.ac.uk/resources/how-guides/develop-data-plan>

DCC checklist for a data management plan: <http://www.dcc.ac.uk/sites/default/files/documents/resource/DMP/DMP-checklist-flyer.pdf>

European Commission (EU-C) (2016): Guidelines on FAIR Data Management in Horizon 2020: https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

FAIR data: Wilkinson, M., et al., 2016: The FAIR guiding principles for scientific data management and stewardship. Nature Scientific Data, 3: 160018, DOI: 10.1038/sdata.2016.18. <https://www.nature.com/articles/sdata201618>

INTERACT: www.eu-interact.org

OpenAIRE: <https://www.openaire.eu/>

Research Data Alliance (RDA): <http://rd-alliance.github.io/metadata-directory/standards/>

Re3Data: <https://www.re3data.org>



INTERACT

Pocket Guide

Data Management

WHAT it is, WHY we need it, HOW to do it

Anna Irrgang¹, Esther Hemmens¹ & Øystein Godøy²

¹ Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Potsdam, Germany

² Norwegian Meteorological Institute, Norway



Norwegian Meteorological Institute

Data management is needed in each research project in order to keep track of data handling, modification and storage. Data evolution in a project is commonly described as the **Data Life Cycle**.



The **Data Life Cycle** describes the different stages of data within a research project. Data practices should be guided by the **FAIR data principles**, meaning that the data is Findable, Accessible, Interoperable and Re-usable (Wilkinson et al., 2016).

The **FAIR data principles** are acknowledged by scientists all over the world as guidelines for data handling.

Already in the **Planning** phase of your research project, it is helpful to think through the life cycle of your data, starting from how to collect them and ending on how to preserve and publish them. You can do this by creating your own **Data Management Plan (DMP)**. It will help you to organize your data in advance and keep track of them during the course of the project. Guidance material for making a DMP is provided for example by the Digital Curation Centre, which also provides a **checklist for creating a DMP**.

During data **Collection or Creation**, it is important that you thoroughly describe your data by creating metadata. Metadata are data describing data. The diversity of data collected at INTERACT stations is very high. The **best practices** for data collection and metadata standards differ with respect to the data type and scientific discipline.



Data Processing implies its modification and quality control. Processing steps need to be documented in the metadata so that you keep track of data evolution. We recommend to use a **self-explanatory versioning system**, to make the alteration of the data transparent.

A **regular data backup** during processing is crucial. This should include storing the data on two independent systems (e.g., PC and institutional server).

The **Preservation** of your research data is a decisive step in the data life cycle. The best practice is to store your data and metadata in a **trusted data repository**. **Re3data** is a search engine for data repositories which could help you find a suitable repository for preserving and publishing your data.

Data Publication is the key to make your data accessible for the public. You can either publish your data and its corresponding metadata, or just the metadata. Use **open data formats** (formats which can be read without a proprietary software) and follow **metadata standards** to ensure the findability, accessibility, interoperability and reusability of your data. Embargos and licensing give you control over data access and usage. Publishing your data makes your research

more trustworthy and may help increase your scientific reputation. Data publication will not only help others to find your data but it will also help you to find the final version of your data after your research project ends!

For the **Reuse** of data, you need to make them findable, accessible, interoperable and document them in a self-explanatory way. Having **rich and standardised metadata** increases the value of your data and the likelihood of its reuse.

Providing access to your data keeps them relevant beyond the scope of your project and will help to exploit their full potential.

Data management within INTERACT

INTERACT is a Horizon2020 project and thus funded by the EU. Horizon2020 projects comply to the Open Data Pilot called **OpenAIRE**. Thus, by default all scientists using INTERACT stations are obliged to provide **free and open access to data collected/created within the framework of the INTERACT network**, or at least to research data which validate the results in scientific publications. If you can provide solid reasons for keeping these data closed, you can opt out of this obligation.

The European Commission released some additional guidelines on FAIR data management in Horizon 2020.

What are Metadata?

Metadata provide the context for data and contain descriptive information about the dataset (e.g., author, location, time, method of data collection, processing steps, etc.). To ensure findability and reusability of data, it is important to use established data and metadata standards (e.g., ISO 19115 compliant). The used metadata standard depends on the data type and the scientific discipline. You will be requested to provide metadata during the data submission process to a trusted data repository.

A list of general and discipline specific metadata standards are provided by the **Research Data Alliance**.

How to find a trusted repository

The search for a trusted data repository usually starts with exploring the repository of your institution or data repositories used by your research community. You can use the Re3data platform to search for a suitable data repository for long-term storage and publishing of your data. Re3data is a searchable online catalogue with entries for more than 2000 data repositories.

The data repository of your choice should fulfill the following criteria:

- Complies to the FAIR guiding principles
- Supports relevant metadata standards and data formats
- Provides persistent identifiers (e.g., DOI)
- Uses data licensing (e.g., creative commons)
- Ensures long-term data preservation.

READ MORE ABOUT INTERACT AND
DATA MANAGEMENT AT:
www.eu-interact.org