**What is metadata?**

Metadata are data about data. They give context to data, for instance:

- Who created the data?
- What are the data about?
- When were the data collected and why?

- Metadata can be created for many different resources (e.g. data, software or real world entities such as people or organizations.
- Ideally, metadata are structured and defined in a machine readable way.
- Metadata is key to make your data FAIR (findable – accessible – interoperable – reusable).

**Will you remember or will others understand**

- How you have processed your data?
- What variables name or values in your spreadsheet mean?

**Metadata as solution for common concerns**

<table>
<thead>
<tr>
<th>Concern</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inappropriate use of my data due to misunderstanding of research purpose or parameters</td>
<td><strong>METADATA</strong> Provide rich context information in form of metadata to make your data understandable</td>
</tr>
<tr>
<td>Security and confidentiality of sensitive data</td>
<td><strong>METADATA</strong> Metadata do not contain the data. Use constraints to specify who may access the data &amp; how</td>
</tr>
<tr>
<td>When others use my research data, they won’t acknowledge my original effort</td>
<td><strong>METADATA</strong> Specify a required data citation within the use constraints</td>
</tr>
</tbody>
</table>

»More than 70 % of researchers have tried and failed to reproduce another scientist’s experiments, and more than half have failed to reproduce their own experiments.«

(Baker 2016)
**Benefits**

- **Research community**: Enabling reprocessing of data
- **Researcher**: Promote the formation of new collaborations
- **Public**: Gain citations for (meta)data publications
- **Research community** and **Researcher**: Support the emergence of new research questions
- **Researcher** and **Public**: Enhance efficiency of scientific workflows
- **Researcher** and **Public**: Extending longevity of data
- **Researcher** and **Public**: Efficient use of public resources
- **Public**: Simplified interdisciplinary cooperation
- **Public**: Support integrity, transparency, and reproducibility of research
- **Public**: Public trust in science

---

www.helmholtz-metadata.de
Let’s face it: You have to invest time and financial resources to create, manage, and maintain accurate and robust metadata for your research data. We recommend to include costs for data documentation in your grant budget and incorporate metadata creation into the data development process. Many funding agencies require it anyway.

**STEP 1**
**Planning**

Start documenting metadata as early as possible! The effort to generate metadata retrospectively and attach it to a set of data is often very high. Therefore, the recording of metadata should always happen alongside or close to the generation of the research data itself.

Starting early will also ensure that information about your data is fresh in your mind. This in turn saves you time and enables you to provide more details about your data leading to a higher-quality metadata record.

**DIVE DEEPER**
Investments of research funds in data stewardship:
https://www.howtofair.dk/how-to-fair/metadata/#metadata-standards-and-ontologies

**STEP 2**
**Timing**

A metadata record captures critical information about the content of a dataset. In its simplest form, a metadata record can be a readme file attached to your dataset. Think about what others would need in order to find, evaluate, understand, and reuse your data:

- provide all context information that is applicable to your data (who-when-what-where-why-how you have processed your data?)
- get others to check the metadata to improve quality

The more complete your metadata record is, the less is the risk of your data being misinterpreted. If in doubt, ask the data management team of your institution.

**DIVE DEEPER**
Documentation and metadata:

**DIVE DEEPER**
How to write “readme style” metadata
https://data.research.cornell.edu/content/readme

**Metadata tools**
https://www.dcc.ac.uk/guidance/standards/metadata/tools
STEP 4
Use standards and vocabularies

- **Standards**
  Do not reinvent the wheel – use existing (domain specific) and machine readable standards to enable interoperability.

- **Controlled vocabularies**
  Where possible metadata used for the description of a dataset should be aligned with an acknowledged and controlled vocabulary. This ensures that the captured descriptions remain generally understandable and interoperable with other datasets ability.

**DIVE DEEPER**
What is a metadata standard
https://www.howtofair.dk/how-to-fair/metadata/#metadata-standards-and-ontologies

How to find a metadata standard
https://rdamsc.bath.ac.uk/

STEP 5
Publish metadata

Make your metadata discoverable and accessible by publishing in a searchable resource (e.g. trustworthy repository) even if the data themself are not publicly available. Assign a persistent identifier (PID) such as a DOI to your metadata record to make it findable and citable.

**DIVE DEEPER**
How to find a suitable repository
https://zenodo.org/record/3900922#
YOwuc25CQ1I

Publishing your metadata may increase your citations: »… up to 25,36 % higher citation impact on average for articles linking to research data«
(Colavizza et al. 2020)

Contact
HMC Office
at GEOMAR Helmholtz Centre for Ocean Research Kiel
Wischhofstr. 1–3
24148 Kiel | GERMANY
hmc-info@geomar.de

If you would like to create (better) metadata, if you are interested in activities of the Helmholtz Metadata Collaboration (HMC), or if you have any questions, we are happy to help you.

**DIVE DEEPER**
Find more information and our sources on
https://www.helmholtz-metadaten.de/en