Metadata
in fundamental Research

Anke Wegman, Data steward CMM
May 2023

Content

- **What** is meant by Metadata?
- **Why** Metadata?
- **Examples** of Metadata documents & topics
- How to **collect/ produce** MetaData ***efficiently***?

What is meant by Metadata?

(in scientific research)

Metadata = Data about Data

= Information about your Research Data

≠ Content of your Research Data themselves

- **General** information (like title/ author / Keywords)
- Information needed to **understand (interpret)** the data
- Information needed to **reproduce/ reuse** the data

Why Metadata?

Metadata are essential in making your data:

see:

*"**FAIR Data in Fundamental Research (UBEC_UMC Utrecht)**"*

** available as video and info slide deck*

Why Metadata?

- To be able to **find** potentially **useful datasets** for your research (like Title, Author and Keywords)

- To **determine** in an **early** stage whether or not the specific **dataset** is really **valuable** for you

  - ✓ Well description of the content of the dataset
  - ✓ Allowed to reuse?
    - *which (if any) restrictions / requirements do apply?*
  - ✓ Standards (ontologies/ codes/ file formats) used (and are they useful for your research?)

Why Metadata?

In sum:

Benefits for you if others provide well described metadata:

Reducing costs, time and effort

→ Prevent generate data that are already there

→ Prevent to upload datasets which turn out not to be (that) useful for your research

However,

What are the **benefits** **for you** **if you provide** well described metadata?

Why producing Metadata yourself?

- **Your research is more Visable**

  - ✓ More **credit**  / Higher **impact**

  - ✓ **Future research opportunities**
    - → Sharing / cooperation with other researchers
      - → Taking **science** to a **higher level**

- Fulfil requirements of grant givers  / journals

  (= Being FAIR, see ***"FAIR data in Fundamental Research"***)

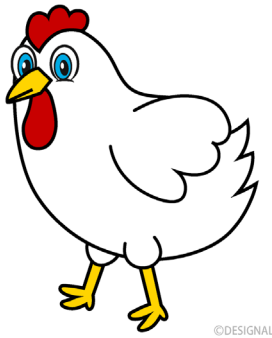
- Reducing costs, time and effort for others

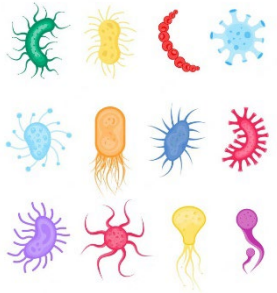
Examples of Metadata

2. Information needed to **interpret** the **data correctly**

2a) What are the research data about?

2a.1 Subjects/ Species





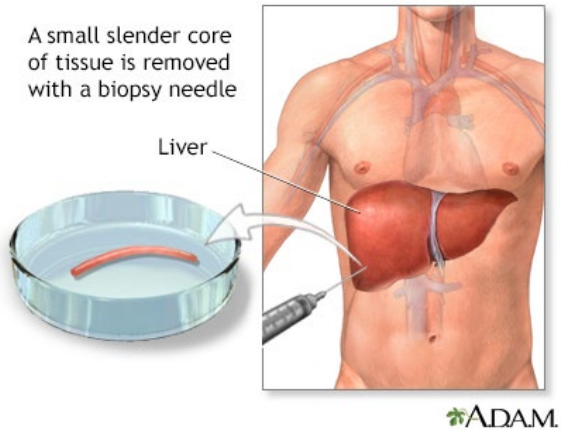
Examples of Metadata

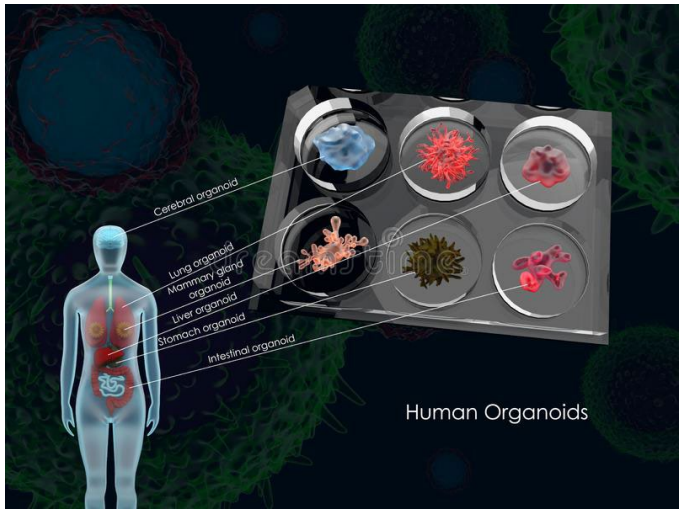
2. Information needed to **interpret** the **data correctly**

2a) What are the research data about?

2a.1 Subjects/ Species

2a.2 Sample Material





Examples of Metadata

2. Information needed to **interpret** the **data correctly**

2a) What are the research data about?

2a.1 Subjects/ Species

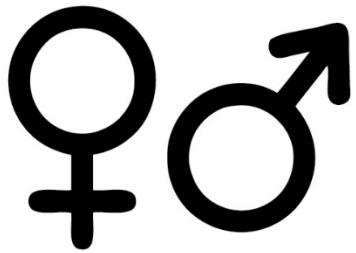
2a.2 Sample Material

2a.3 Other sample metadata – depending on the experiment

examples:

age

sex



health

Examples of Metadata

2. Information needed to **interpret** the **data correctly**

2a) What are the research data about?

2a.1 Subjects/ Species

2a.2 Sample Material

2a.3 Other sample metadata – depending on the experiment

2b) Meaning of the research data themselves

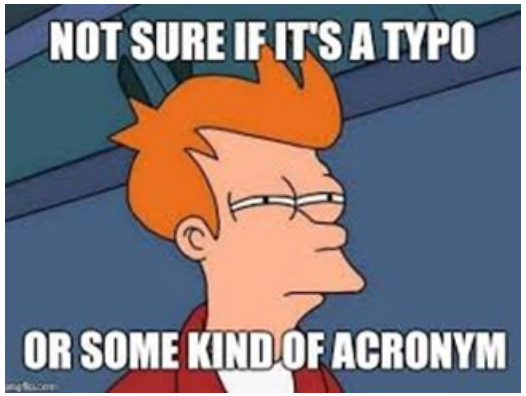
(see next slides)

Examples of Metadata

2b) Meaning of the research data themselves

2b.1 Codebook/glossery

- meaning of abbreviations



- what do the colors in the images represent?

2b.2 Units clearly stated? (examples: mg/ml, %, …)

2b.3 Reference data

(example: the reference genome that was used)

Metadata Resume (so far):

1. **General information** of the research **dataset enables you:**
   → **Find** potential useful datasets
   → **Determine** dataset is actually **valuable**
   → Know at early stage:
      - **allowed** to dowload dataset
      - **able** to open and read the data
2. Information to **interpret** the data correctly, like:
      - Sample characteristics
      - Abbreviations/ units clearly stated

Next - Examples of Metadata

3. <u>All Information</u> needed to **reproduce/ reuse** the **data**

a) **Description** of the (standard) **methods** used

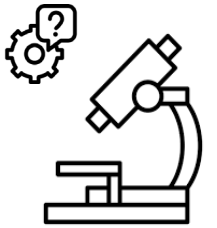
a) Which **settings** of which **device** used

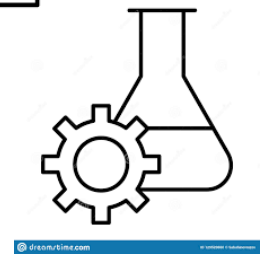
b) Conditions (temperature, concentrations, etc.)

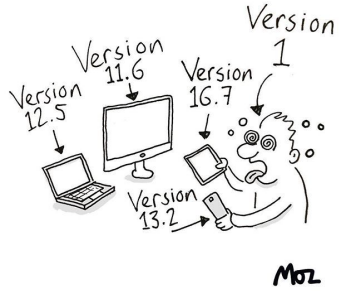
c) Already existing **software** tools used (exact **version**!)

d) Developed software tools

e) ….







Examples of Metadata-documents

Needed to archieve (at your organization)

a) Project proposal
b) Protocol
c) DMP (Data Management Plan)
d) Info in ELN
e) Legal documents:
   - Consortium agreement
   - Blank Informed Consent (!)
   - Requirements for use (cell lines)
   - …

How to collect/ produce Metadata efficiently?

a) **Determine as soon as possible** what **metadata standards** are used in:
   - your research field
   - the repositories you plan to store your data at the end of your research project

b) **Collect, note** and **save as soon as possible** your *metadata* + *ReadMe file* in your project folder

More detail: see next slide

How to collect/ produce Metadata efficiently?

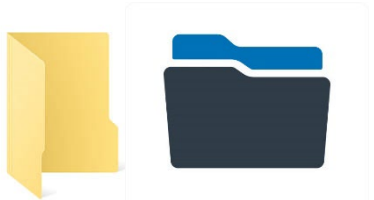
b1) **Collect, note** and **save** your metadata **as soon as possible** (device settings, software versions, projectplan, Informed Consent (**!**), DMP, …)

b2) Create a **ReadMe file**

a) What information is saved in which subfolder?
(general project info; raw data; analysed data; …)

b) How version-control of the files is regulated
(naming of files)

Store in your **project folder**



Considerations

The field of metadata keeps developing

Contact your **Data Steward !!!**
For the most updated Metadata-info

More info regarding
Research Data Management:
www.UBEC.nl