BioData.pt

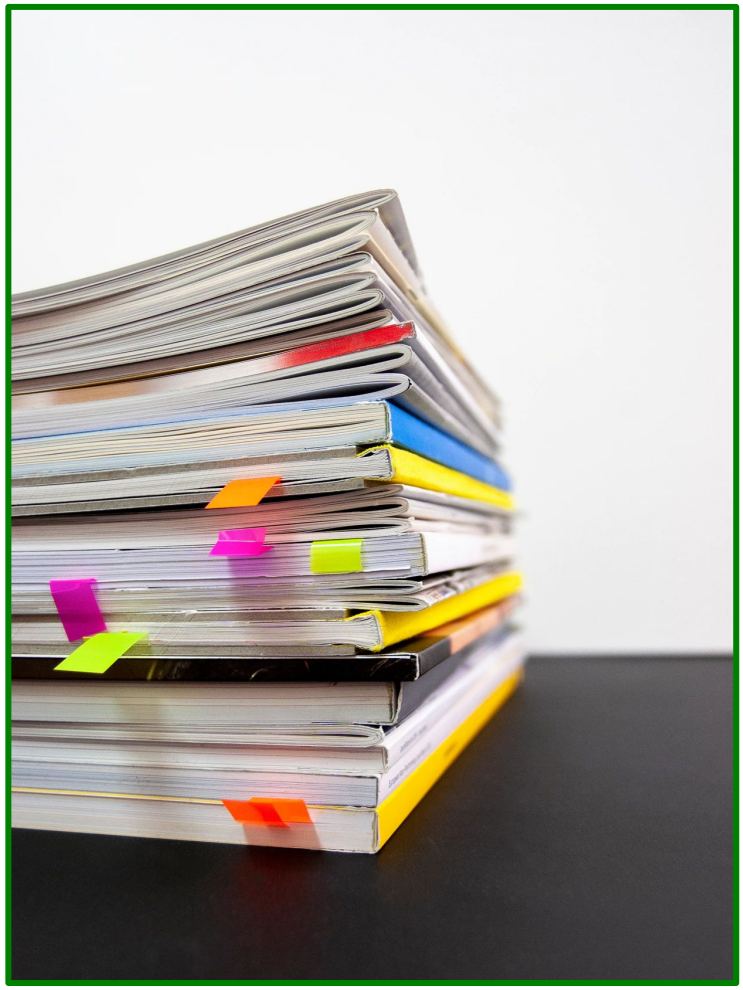
Ready for BioData Management?

Demystifying Data Management Plans

João Cardoso

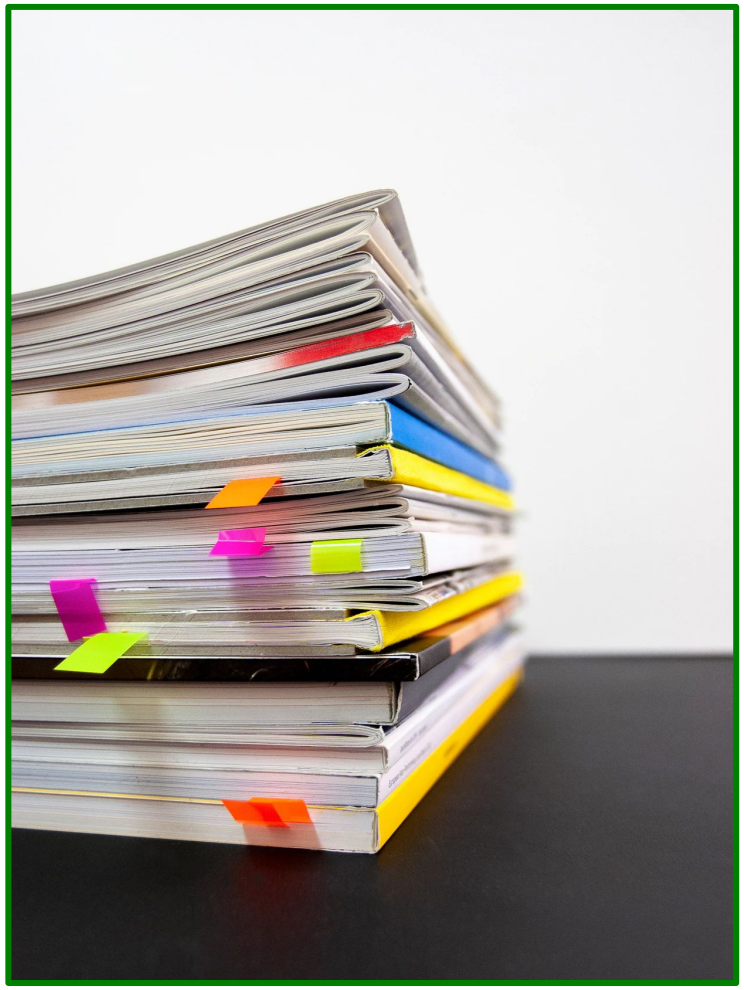
What is a Data Management Plan (DMP)?

- A **DMP** is a formal document used to support **Data Management.**
- It describes **techniques, methods and policies** on how data is to be:
  - Created / collected,
  - Documented,
  - Accessed,
  - Preserved,
  - Disseminated.



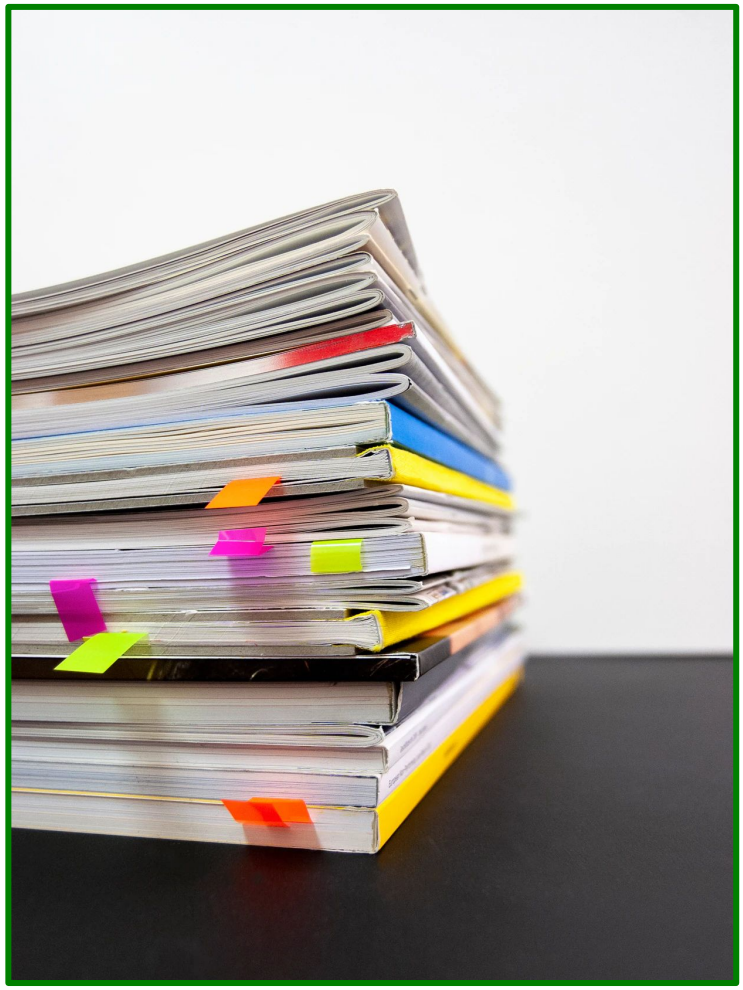
What is a Data Management Plan (DMP)?

- With the goal of ensuring an a**dequate allocation of resources**:
  - Funds,
  - Storage,
  - Man-hours.
- To the **managing of data** during the exercise of a:
  - Research project,
  - Research facility (e.g. sequencing),
  - Research lab or institution.



What is a Data Management Plan (DMP)?

- A DMP is a **living document**
  - It should be updated as needed during the exercise of its subject.
- It is the **latest of a series or artefacts** to tackle this issue over the years:
  - Data Handling Plans,
  - Data Sharing Plans,
  - Operations Manuals.



Why Do We Need DMPs?

- The carrot:
  - DMPs are **invaluable tools in the planning** of research activities to ensure the necessary resources are devoted to data management.
  - They can set-up compliance with the **FAIR data principles.**
- The stick:
  - Many **funding agencies** now require that **grant proposals** be **accompanied by a DMP**
  - **Monitoring** of the **quality** and **execution** of these DMPs is **still light**, but expected to tighten

What should be in a DMP?

https://en.uit.no/ub/forskningsstotte/art?p_document_id=473665

Administrative & Legal Aspects

- Which **institutions** and **people**:
  - Are involved in the research activities, and what are their roles,
  - Are responsible for the execution of the DMP, and what are their roles.
- Who are the **contact persons or institutions** for the DMP.

Responsibilities & Duties

- Which institutions (and people):
  - **Own** the data,
  - Are responsible for **data collection,**
  - Are responsible for **data protection** (if applicable),
  - Are responsible for **data security** (if applicable).

Costs & Resources

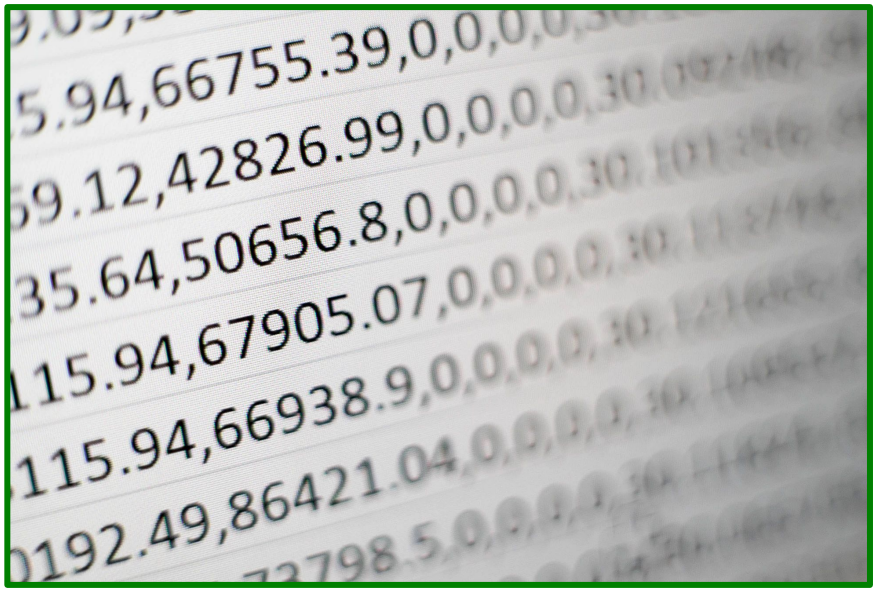
- What will it cost to:
  - **Analyse** data (hardware, software, man-hours),
  - **Prepare/clean/curate** data (man-hours, maybe software),
  - **Store** data (hardware, maybe man-hours),
  - **Publish** data (publication fees)
  - etc.

Project Description

- A **summary** of the project proposal:
  - Goals,
  - Experiments,
  - Methods.

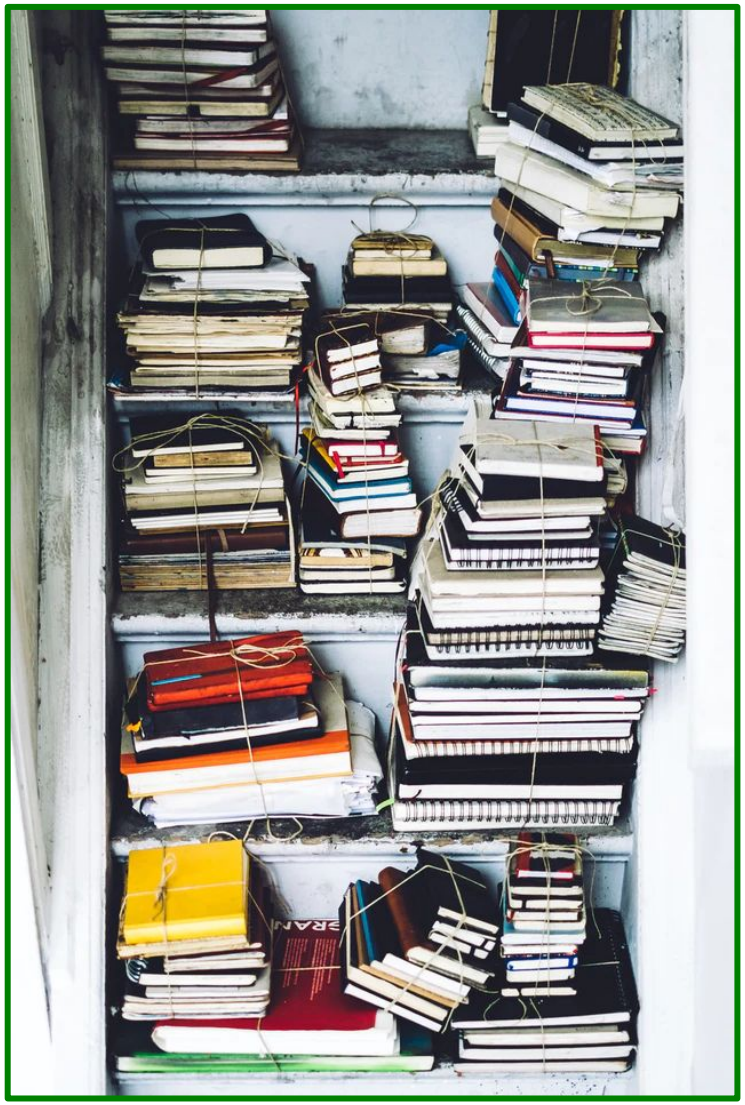
Data

- **Existing** data (if applicable):
  - Data source(s)?
  - Usage licence(s)?
  - Volume of data?
- Data to be **created** or **generated**:
  - How?
  - What types?
  - Volume of data?



Data

- Data **organisation**:
  - How data will be described (**metadata**)
  - How data will be structured (**data formats**)
  - How data will be interconnected (**data structures**)
  - Where data will be **stored** during the project



Data

- **Archiving, sharing** and **publishing** data:
  - Archiving:
    - Where?
    - For how long?
  - Sharing / publishing:
    - Where / how?
    - License?
    - Who can access?
    - Privacy & security?

The Typical DMP Scenario

- In current practice, DMPs are mainly seen as:
  - A **bureaucratic hassle**
  - **Static documents** or **rarely updated**
  - Just **human** readable
  - **Not published** or publicly accessible
- Which leads to poor quality DMPs, of low practical value
- This is aggravated by the fact that different funding bodies use **different DMP templates**

The Future of the DMP

- In a **perfect world** the DMP **should**:
  - Have a **human** and **machine-readable** representation
  - Be **shareable**
  - Comply with a **common standard**
  - Be a **living document**
  - Essential part of **data management**

The Future of the DMP

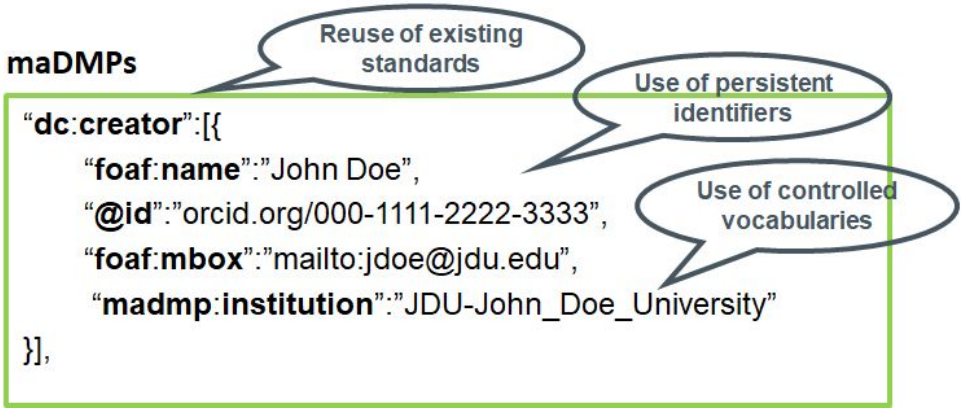
- The **Machine-Actionable DMP (maDMP)** concept was introduced to **extend** the concept of DMP.

Current DMPs

```
<admindata>
    <question>Who is the Principle Investigator?</question>
    <answer>The PI is John Doe from the JDU</answer>
</admindata>
```

- The **main points** of maDMPs:
  - **Machine** and **human** readable **descriptions**
  - **Automated** policy enforcement
  - **Shearable**
  - **Interoperable** DMP version
  - **Extensible**

maDMPs



The Future of the DMP

- The **RDA DMP Common Standards Working Group** was created to focus on the **standardization of the knowledge** contained in a DMP.

- Its **objective** was to establish a **metadata application standard** that defines a **core set of elements** for a DMP.

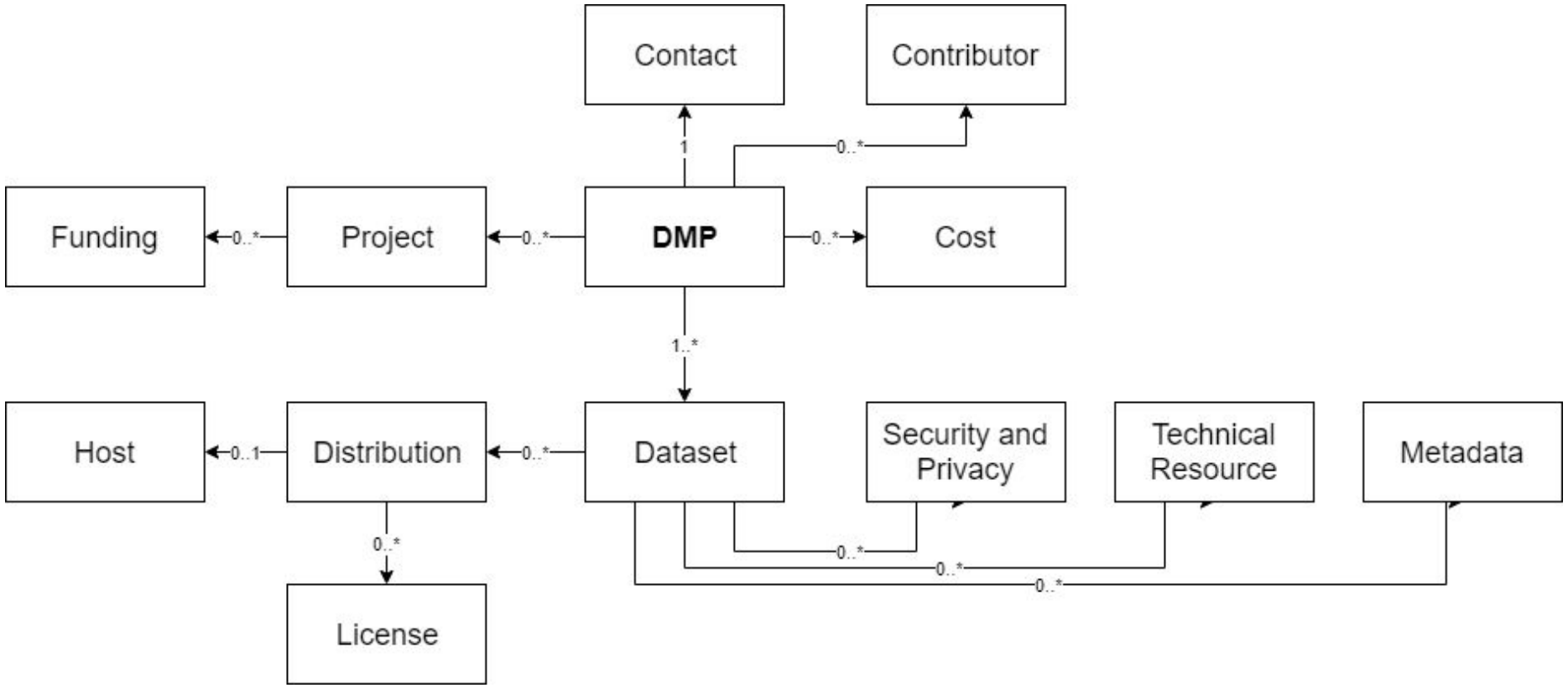
- The **metadata application standard** is **modular in design**, and allows for extensions.

RESEARCH DATA ALLIANCE

Scan for more!

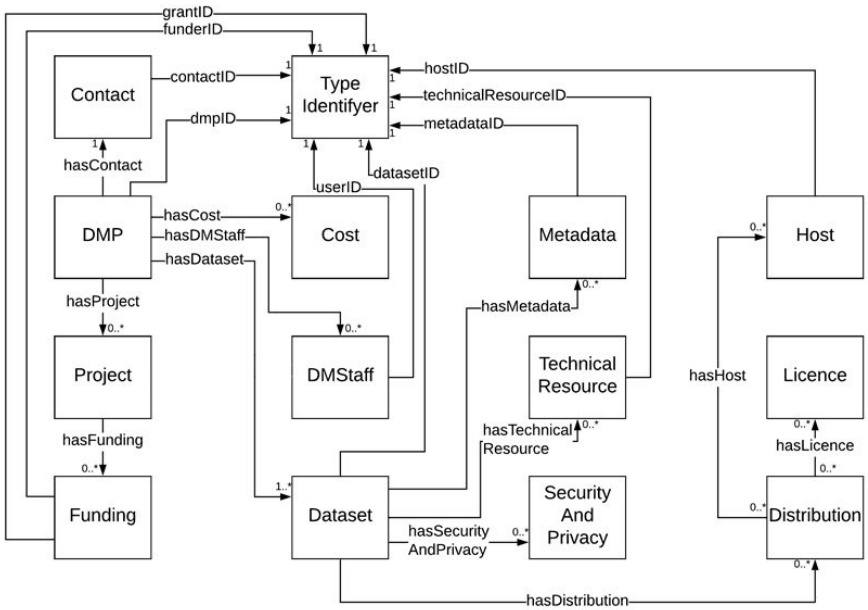
The Future of the DMP

- A **minimum set of universal terms** to ensure basic **interoperability** of systems using DMPs.



The DMP Common Standard

- The DMP Common Standards metadata application standard has several **machine-readable representations**.
- The **DMP Common Standards Ontology (DCSO)** is a **semantic representation** of the metadata application standard using **OWL**



The Future of the DMP

- Possible **applications** of a **maDMP**:
  - One DMP for all templates.
  - DMP maturity model.
  - Automation in both creation and monitoring during the project's life-cycle.

The Take Home Message

- The benefits of DMPs:
  - **Promote Data Management**
  - Assist in compliance with **FAIR data principles**
  - Ensure **adequate allocation of resources** to manage research data
  - **Accountability**
  - Compliance with **grant application requirements**
- The benefits of maDMPs:
  - **Automation** (creation, validation, policy enactment)