



Ready for BioData Management?



# Demystifying Data Management Plans

João Cardoso, Daniel Faria



# What is a Data Management Plan (DMP)?

## Learning Outcome 1:

Recognize the purpose of Data Management Plans

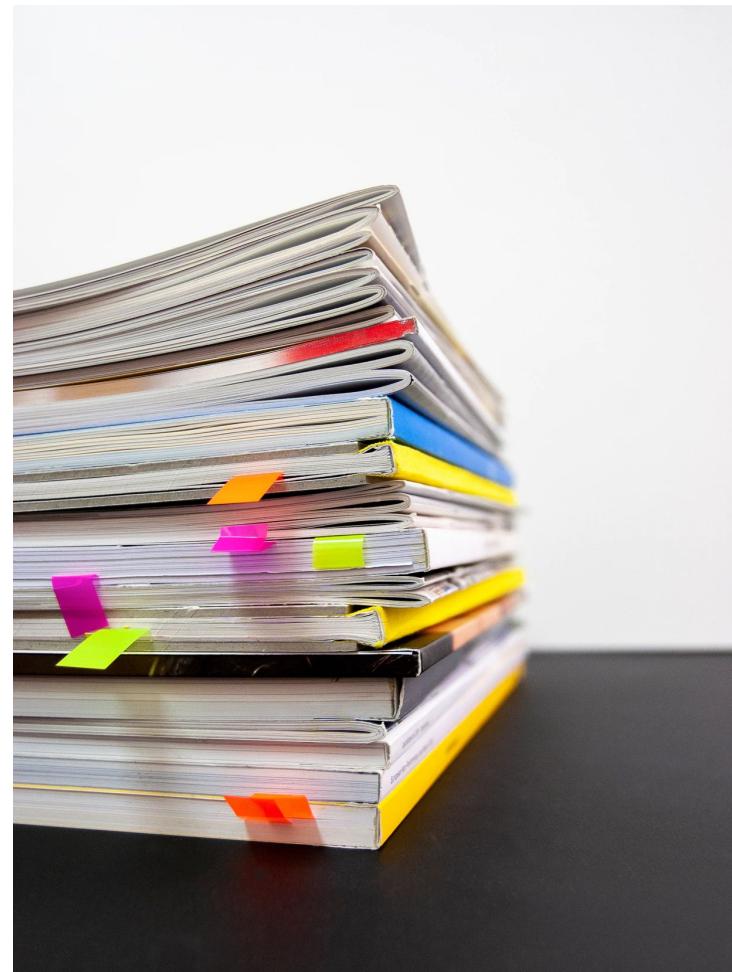
# What is a DMP?

- A DMP is a formal document used to plan and support data management activities by anticipating needs and requirements in a (research) project, facility or institution
- It is the to data management what a blueprint is to construction



# What is a DMP?

- A DMP should detail policies and methods pertaining to data:
  - Creation / collection
  - Documentation
  - Access
  - Preservation
  - Dissemination
- And ensure an adequate allocation of resources:
  - Human
  - Computational
  - Financial



# Why Do We Need DMPs?

- The stick:
  - Many funding agencies now require that grant proposals be accompanied by a DMP
  - In particular, they require DMPs that demonstrate intent to comply with the FAIR data principles.
  - Monitoring of the quality and execution of these DMPs is still light, but expected to tighten



# Why Do We Need DMPs?

- The carrot:
  - DMPs are valuable tools in the planning of research activities to ensure the necessary resources are devoted to data management
  - Adequate planning can facilitate the task of ensuring compliance of research outputs with the FAIR principles





# What Should be in a DMP?

## Learning Outcome 2:

List the main topics that should be covered by a DMP

# What should be in a DMP?



CC-BY Margaret Louise Fotland, UiO

# 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)
  - ...



# Project Description

- A **summary** of the project proposal:

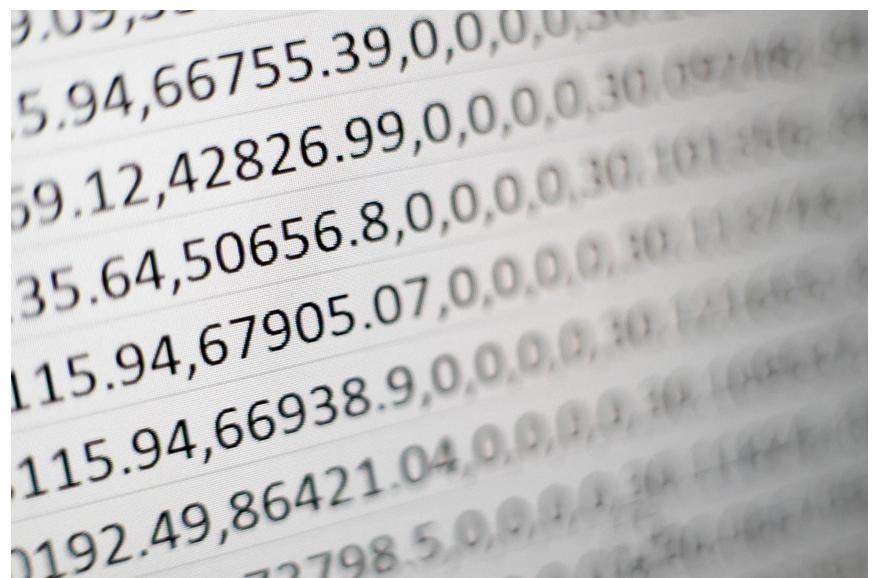
- Goals
- Experiments
- Methods

(Note: not needed if the DMP is part of a project proposal)



# 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?





# DMPs: Present & Future

## Learning Outcome 3:

Describe the current state and future directions of DMPs

# DMPs: Present

- In current practice, DMPs are mainly seen as a bureaucratic hassle
- They are static documents, prepared for grant applications because they are mandatory, but never or rarely updated
- They are generally not validated during the research project, and are never published, which prevents external validation



# DMPs: Present

- The fact that different funding bodies use different DMP templates makes it difficult for researchers to get familiar with them and to recognize the value
- Moreover, most templates are free text questionnaires that look more like surveys than planning documents, and are only human readable
- All this results in poor quality DMPs, of low practical value



# DMPs: Present

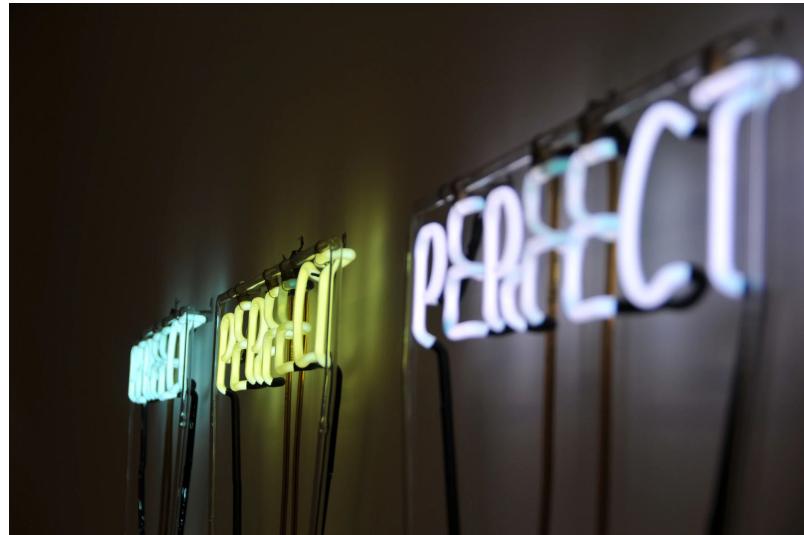
## Example: the H2020 DMP template

A questionnaire covering the following topics:

1. Data Summary
  - Describe the data to be acquired/produced
2. FAIR data
  - **Detail** how you'll comply with the FAIR principles
3. Allocation of resources
  - Who does what and what it costs
4. Data security
5. Ethical aspects
6. Other issues

# DMPs: Future

- To be of practical use, a DMP should be:
  - A living document that is updated as needed
  - Both human and machine-readable
  - Comply with a common standard
  - Be shared



# DMPs: Future

- The Machine-Actionable DMP (maDMP):
  - **Machine and human readable descriptions**
  - **Automated policy enforcement**
  - **Shearable**
  - **Interoperable DMP version**
  - **Extensible**

## Current DMPs

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

## maDMPs

```
"dc:creator": [
    "foaf:name": "John Doe",
    "@id": "orcid.org/000-1111-2222-3333",
    "foaf:mbox": "mailto:jdoe@jdu.edu",
    "madmp:institution": "JDU-John_Doe_University"
],
```

Reuse of existing standards

Use of persistent identifiers

Use of controlled vocabularies

# DMPs: Future

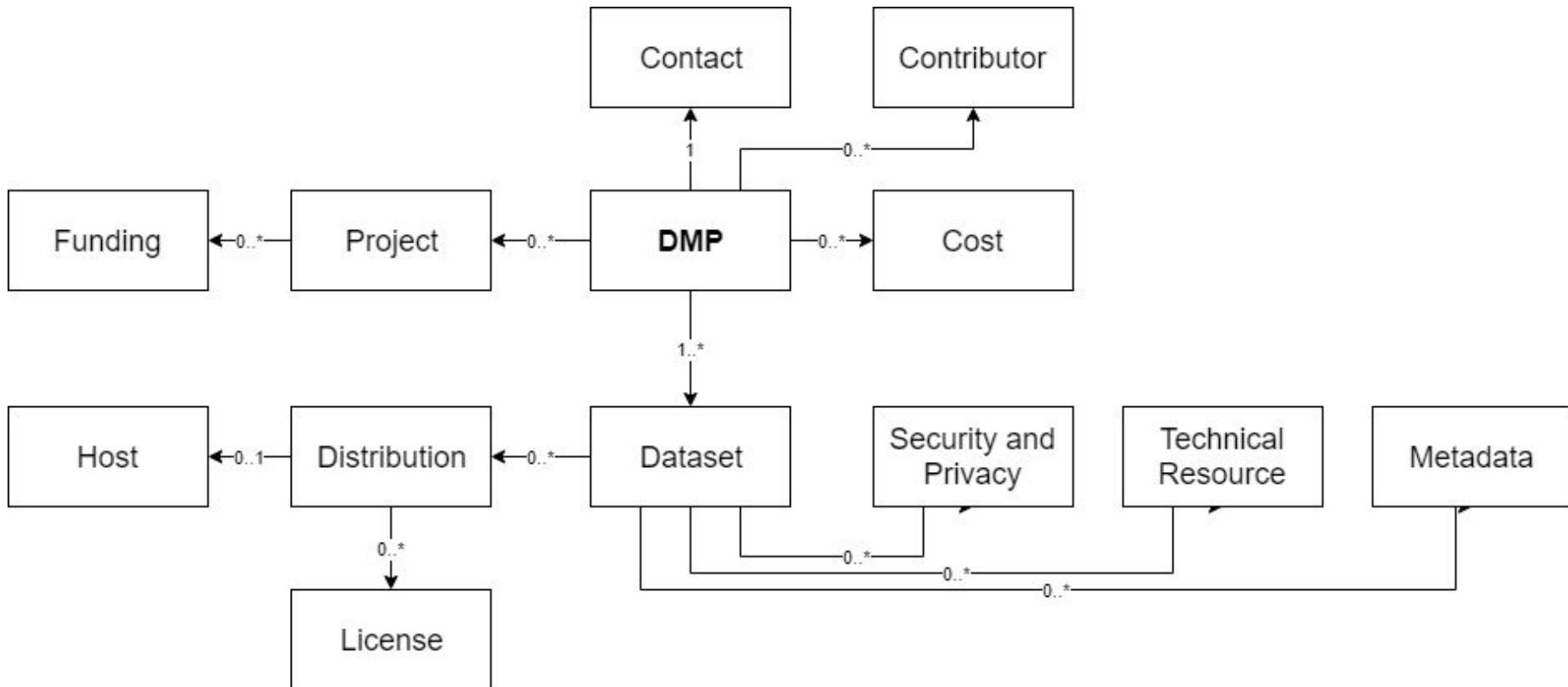
- The RDA DMP Common Standards Working Group was created to focus on the standardization of the knowledge contained in a DMP.
- Its objective was to establish a metadata application standard that defines a core set of elements for a DMP.
- The metadata application standard is modular in design, and allows for extensions.



Scan for more!

# DMPs: Future

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



# DMPs: Future

- 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 good data management practices
  - Assist in compliance with FAIR data principles
  - Ensure adequate allocation of resources to data management activities
  - Enable accountability
- The benefits of maDMPs:
  - Automation (creation, validation, policy enactment)
  - Increase usefulness

