



Introduction to Data Management Plans

João Cardoso, INESC-ID & IST, Universidade de Lisboa
(joao.m.f.cardoso@ulisboa.pt)



Before we start...

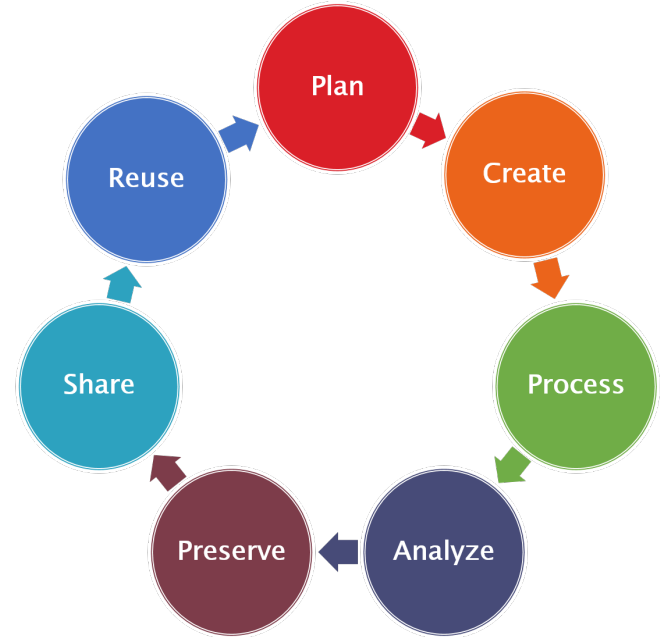
Join the Poll at **slido.com** with the code **#853548**

slido



What is a DMP?

The DMP should detail **practices** and **methods** applied to data throughout each of the stages of the **research data lifecycle**.



Why do I need a DMP?

The Carrot

When planning RDM activities, the DMP can be a helpful resource in order to **adequately allocate resources** to data management activities.

- Financial
- Computational
- Human

MOTIVATION



Why do I need a DMP?

The Carrot

With **adequate planning**, it's easier to have the product of research **comply** with the **FAIR Data Principles**.

- Findable
- Accessible
- Interoperable
- Reusable

MOTIVATION



Why do I need a DMP?

The Carrot

There is high potential for **automation** of data management activities to be enacted, based on a **frequently updated**, and **machine-readable** DMP.

This in turn can both **accelerate** and **simplify** the **research process**.

MOTIVATION



Why do I need a DMP?

The Stick

Currently multiple **funding bodies** are **encouraging the submission of a DMP** along with any funding application.

However, this stance is **gradually changing** from one of encouragement to one of **requirement**.



Why do I need a DMP?

The Stick

For funding bodies, the DMP is perceived as a **declaration** of how the applicant **intends to comply with the FAIR Data Principles**.

It can also serve as a means for future **monitoring and validation** of both **quality** and **FAIR compliance**.



Why do I need a DMP?

The Stick

Quality monitoring and validation is still limited
to non-existent.

However, these aspects are considered **priorities**
by several funding bodies, and are **expected to**
tighten in the upcoming years.



How's it going so far?

Join the Poll at **slido.com** with the code **#853548**

slido



What should be in a DMP?

Project
Description

Existing Data

Created Data

Data
Organisation

Preservation
and Access
Policies

Licences and
Ethics

Resources and
Responsibilities

Data
Management
Costs



What should be in a DMP?

Project
Description

Existing Data

Created Data

Data
Organisation

Preservation
and Access
Policies

Licences and
Ethics

Resources and
Responsibilities

Data
Management
Costs



What should be in a DMP?

Project Description

The DMP should include (if not part of a project proposal), a **summary** of the research project or research unit to which it pertains.

This implies describing its **goals, specific methodologies, context**, etc.

One of the key aspects is to clearly describe the **sources of funding** for the data management activities described in the DMP.



What should be in a DMP?

Project
Description

Existing Data

Created Data

Data
Organisation

Preservation
and Access
Policies

Licences and
Ethics

Resources and
Responsibilities

Data
Management
Costs



What should be in a DMP?

Existing Data

If applicable, a description of any **existing data** should be provided.

This implies describing **sources of data**, and its **volume**, any **licenses** that apply or any **costs** associated with its usage.



What should be in a DMP?

Project
Description

Existing Data

Created Data

Data
Organisation

Preservation
and Access
Policies

Licences and
Ethics

Resources and
Responsibilities

Data
Management
Costs

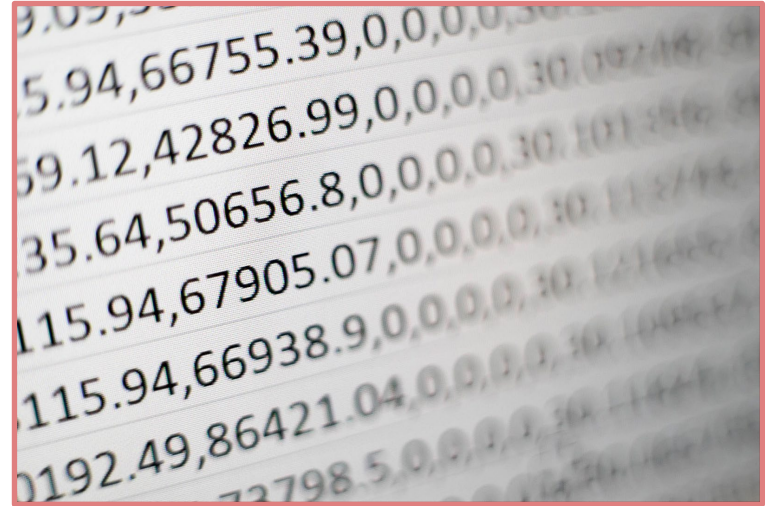


What should be in a DMP?

Created Data

All **data created** in the context of the research project or unit, **should be described in the DMP**.

This implies describing the **methodology of how data is created**, what **type of data** is to be created, and in what **volume**.



What should be in a DMP?

Project
Description

Existing Data

Created Data

Data
Organisation

Preservation
and Access
Policies

Licences and
Ethics

Resources and
Responsibilities

Data
Management
Costs



What should be in a DMP?

Data Organisation

Metadata schemas that are applied must be **identified**, and how these metadata schemas are applied should also be characterized within the context of the research.

The **representation** of the data must also be addressed, this implies describing the **data format**.

Any **data structures** that apply should also be defined.



What should be in a DMP?

Project
Description

Existing Data

Created Data

Data
Organisation

Preservation
and Access
Policies

Licences and
Ethics

Resources and
Responsibilities

Data
Management
Costs



What should be in a DMP?

Preservation and Access Policies

The DMP should **detail preservation and access policies**. These should be applied to one or multiple of the previously described datasets.

A **preservation and access policy** should define **where** the data will be hosted, **who** can access it, and **how** that access is to be performed.



What should be in a DMP?

Project
Description

Existing Data

Created Data

Data
Organisation

Preservation
and Access
Policies

Licences and
Ethics

Resources and
Responsibilities

Data
Management
Costs



What should be in a DMP?

Licenses and Ethics

In sequence with the definition of preservation and access policies for the datasets, it is essential to consider any existing **ethics issues**.

Selecting the **right licence** for the desired policy, is also fundamental.



What should be in a DMP?

Project
Description

Existing Data

Created Data

Data
Organisation

Preservation
and Access
Policies

Licences and
Ethics

Resources and
Responsibilities

Data
Management
Costs



What should be in a DMP?

Resources and Responsibilities

Having a clear **identification** of the **existing resources**, and how they will be allocated, is key to a good DMP.

The **assets responsible** for determined **data management activities** should also be clearly identified.

These two factors combined, improve **traceability**.



What should be in a DMP?

Project
Description

Existing Data

Created Data

Data
Organisation

Preservation
and Access
Policies

Licences and
Ethics

Resources and
Responsibilities

Data
Management
Costs



What should be in a DMP?

Data Management Costs

The DMP should have a **detailed description of all costs** that are related with **data management activities**.



The reality of DMPs

The early DMPs

As DMPs began to be adopted by funding bodies as viable data management tools, the community quickly identified several problems.

There was little to **no consistency** between DMPs

Their **quality and granularity** was tightly linked to the expertise of the creator.

They were not **accessible**.



The reality of DMPs

The DMP Template

In order to counter the identified issues several funding bodies introduced **Data Management Guidelines**, which often come with their own **DMP templates**.

DMP Templates gather information from researchers in a questionnaire like format with open answers.

However these templates **serve the interests of each funding body** and are not designed with assistance to data management activities in mind.

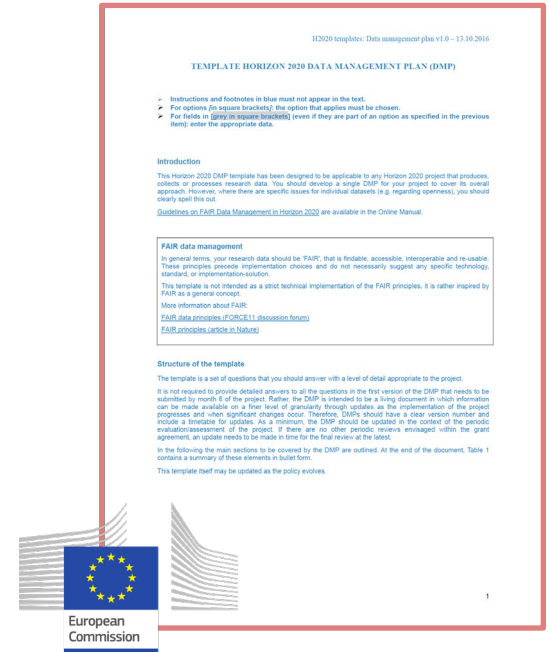


The reality of DMPs

The Horizon 2020 DMP Template

For example, the **Horizon 2020 DMP Template** reflects the EC's focus on having researchers describe how their project will be **complying with the FAIR Data Principles**.

However, there are no clear definition of preservation and access policies. With existing **policies having to be inferred** from knowledge expressed in multiple questions.



The reality of DMPs

Issues with current DMPs

Multiple existing and **conflicting DMP Templates** lead to **confusion** amongst the research community.

Submitted DMPs rarely go through a **verification and evaluation** process, being perceived as a **bureaucratic hassle**, and not as a key tool for data management.

This leads to many DMPs being solely **static documents**, that once created are **rarely if ever updated or published**.



The reality of DMPs

Issues with current DMPs

The combination of these issues leads to the overall **majority of created DMP** documents having **poor quality and low practical value**.

As such, the research community **fails to recognise the potential value** of having a DMP.



The reality of DMPs

Issues with current DMPs

Ideally, to be of practical use, the DMP should have the following features:

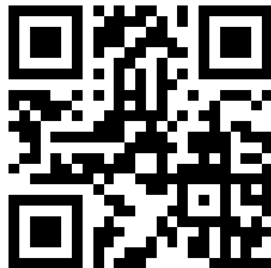
- Easily created
- Frequently updated
- Standardised
- Human and machine-readable representation
- Accessible



Crew check

Join the Poll at **slido.com** with the code **#853548**

slido

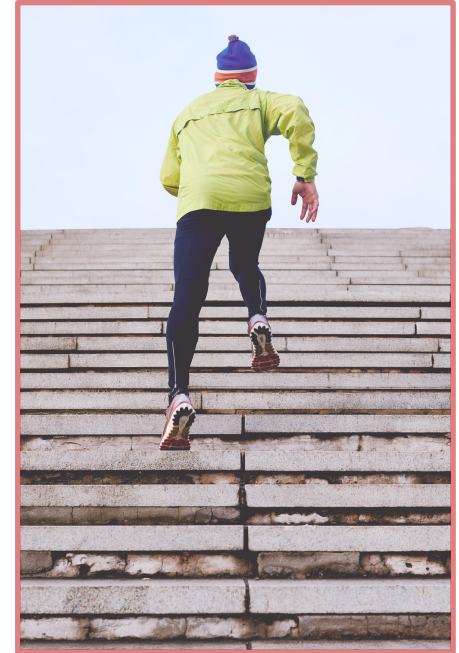


A better DMP

What is being done?

So **what is being done** to counter this trend?

How is the community **making DMPs better**?



A better DMP

DMP Common Standard

The first step was to **standardise the knowledge** contained in a DMP.

The **RDA DMP Common Standards Working Group** was tasked with addressing that challenge.

It's objective was to establish a standard to define a **core set of elements for a DMP**.

The standard should serve as the **starting point** for future extensions.

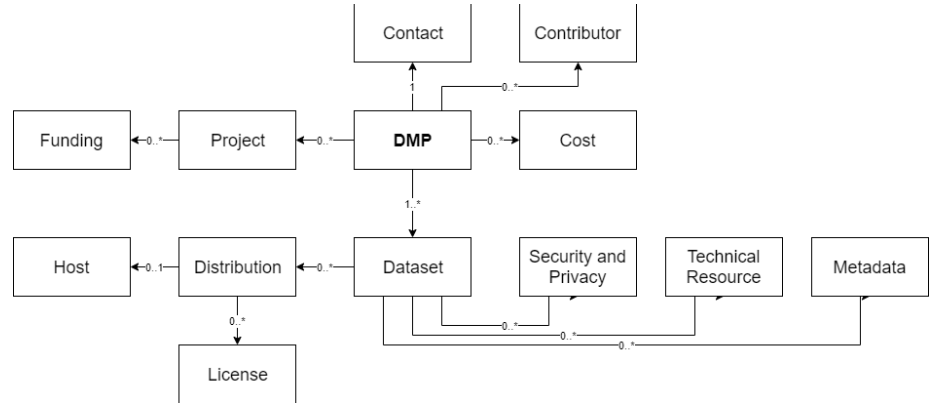


A better DMP

DMP Common Standard

The standard defined a DMP through 13 core terms.

These were to be the **minimal set of universal terms** to ensure **interoperability between systems** using DMPs.



A better DMP

DMP Creation

Secondly there needed to be an **easier, and consistent means to create DMP documents.**

Multiple DMP creation tools were developed and introduced to the community.

Some are **associated with funding bodies** and focus exclusively on following their DMP Templates.

Others are independent of any funding body, and **offer flexibility.**



A better DMP

The Machine-Actionable DMP

The final step was to have **machine-actionable representations of the DMPs**.

These maDMPs are represented using both **human as well as machine-readable data interchange formats**.

Examples of popular representation languages are JSON, and OWL.

These representations follow the RDA's **DCS application profile**.



A better DMP

The Machine-Actionable DMP

The maDMP concept addresses some of the issues of traditional DMPs by:

- Enabling the exchange of information between systems
- Allowing the integration of DMPs in RDM workflows
- Enforcing persistent identification of artefacts
- Facilitating the updating process
- Facilitating the creation process



Summary

DMPs are formal documents that describe **methods** and **practices** applied to data throughout the **research data lifecycle**.

They facilitate the adoption of **good RDM practices**, and in particular the **FAIR Data Principles**.

An important tool for the planning and adequate allocation of resources to RDM activities.

They have the potential to enable the **automation** of RDM activities.



Thank you

Join the Poll at **slido.com** with the code **#853548**

slido

