



# An introduction to Data Management Plans

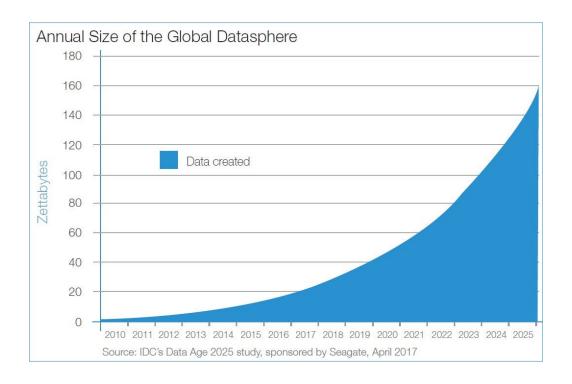
João Cardoso, INESC-ID (joao.m.f.cardoso@tecnico.ulisboa.pt)



### 1. The problem

There's an **exponential growth** in number of **researchers** and their ability to produce **research** data.

How to **make research data available** for the benefic of the scientific community?



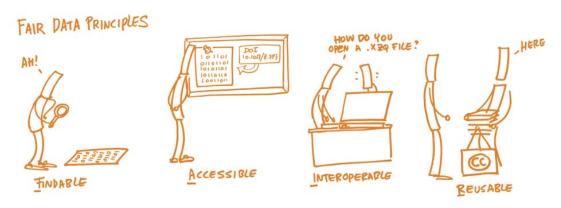
https://www.openaire.eu/how-to-make-your-data-fair



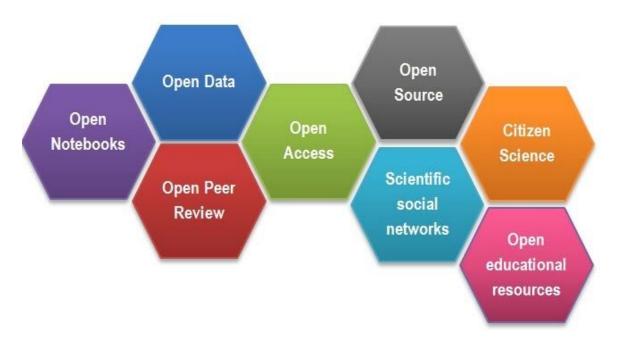


### 1. The problem

This challenge can be mitigated through the application of practices and formal methodologies of research data management, open science and the adoption of the FAIR data principles.







https://www.mysciencework.com/omniscience/open-science-open-access-far-apart





## 2. What is a DMP?

A Data Management Plan (DMP) is a formal document that is used to support the planning and execution of data management activities.

Restaurant technical sheets => How to prepare food

DMP documents => How to manage research data.

#### BLUE FISH GRILL RECIPE BOOK

#### TARTAR SAUCE

Batch
5 gallons
5 days
20-qt. Lexan
Commercial mixer
S/S mixing bowl
Wire whip attachment
Prep cook

B D		
	No.	1
		1

QTY	MEASURE	INGREDIENTS
3	Gallons	Mayonnaise
1 1/2	Gallons	Sour cream
1/2	Cup	Dijon mustard
3	Cups	Shallots (fine-chopped)
3/4	Cup	Dill relish
3	Cups	Fresh dill weed (fine-chopped)
3	Cups	Fresh parsley (fine-chopped)
4	TBSN	Tarragon leaves
4	TBSN	Hot sauce
1/2	Cup	Salt
4	TBSN	Black pepper
	i i	
	× ×	
	i i	
NOTE	S:	

#### PROCEDURE

- Combine all ingredients into mixing bowl for commercial mixer.
- Using the wire whip attachment, mix until fluffy and well-blended.
- 3. Cover and apply date label.
- 4. Refrigerate.





# 2. What is a DMP?

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







The **motivation** for the use of DMP documents is **best understood** in the perspective of the **carrot and stick metaphor**.







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

- Financial
- Computational
- Human









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

- Findable
- Accessible
- Interoperable
- Reusable









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**

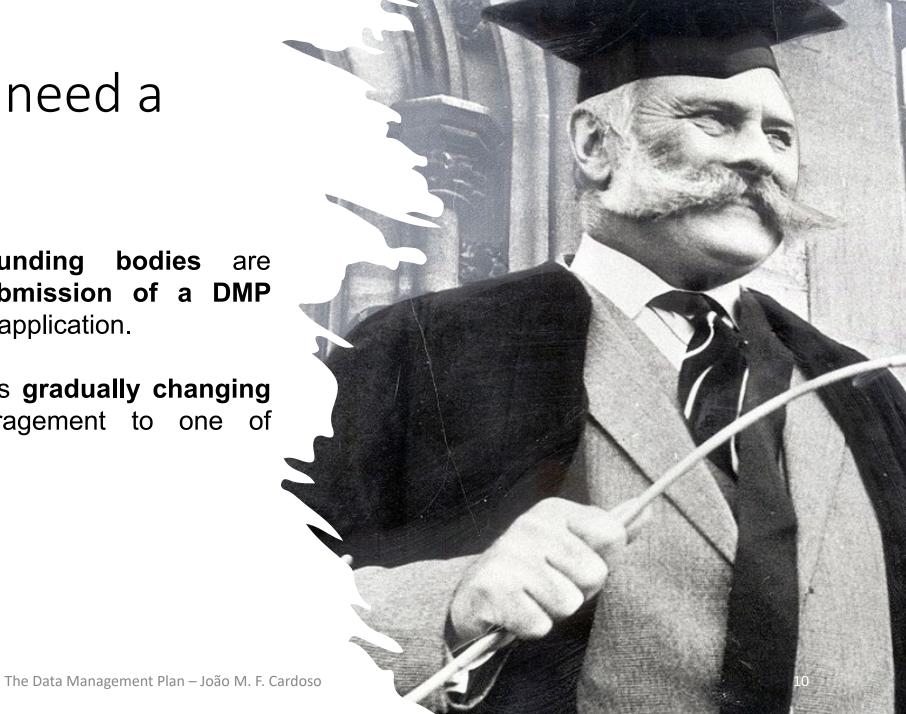






Currently multiple **funding bodies** are **encouraging the submission of a DMP** along with any funding application.

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

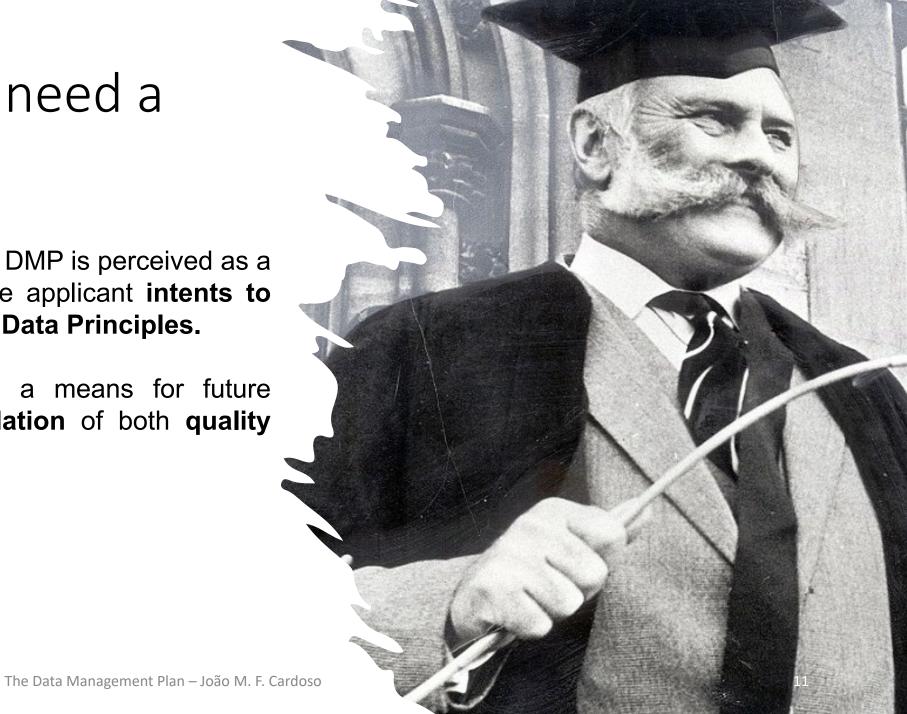






For funding bodies, the DMP is perceived as a declaration of how the applicant intents to comply with the FAIR Data Principles.

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

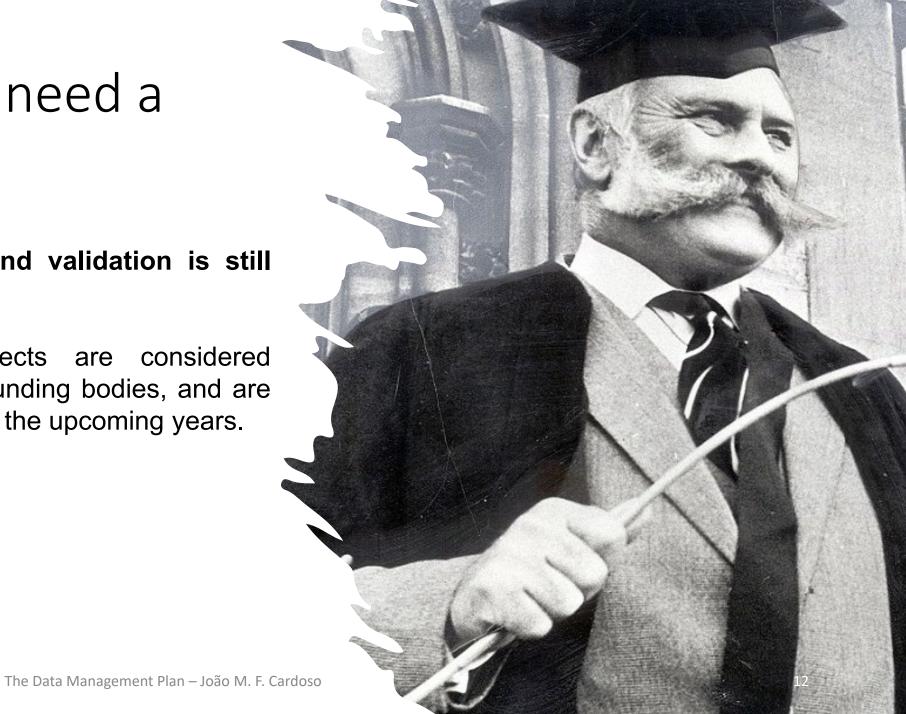






Quality monitoring and validation is still limited to non-existent.

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







**Project Description** 

**Existing Data** 

**Created Data** 

**Data Organisation** 

Preservation and Access Policies

Licences and Ethics

Resources and Responsibilities





**Project Description** 

**Existing Data** 

**Created Data** 

**Data Organisation** 

Preservation and Access Policies

Licences and Ethics

Resources and Responsibilities





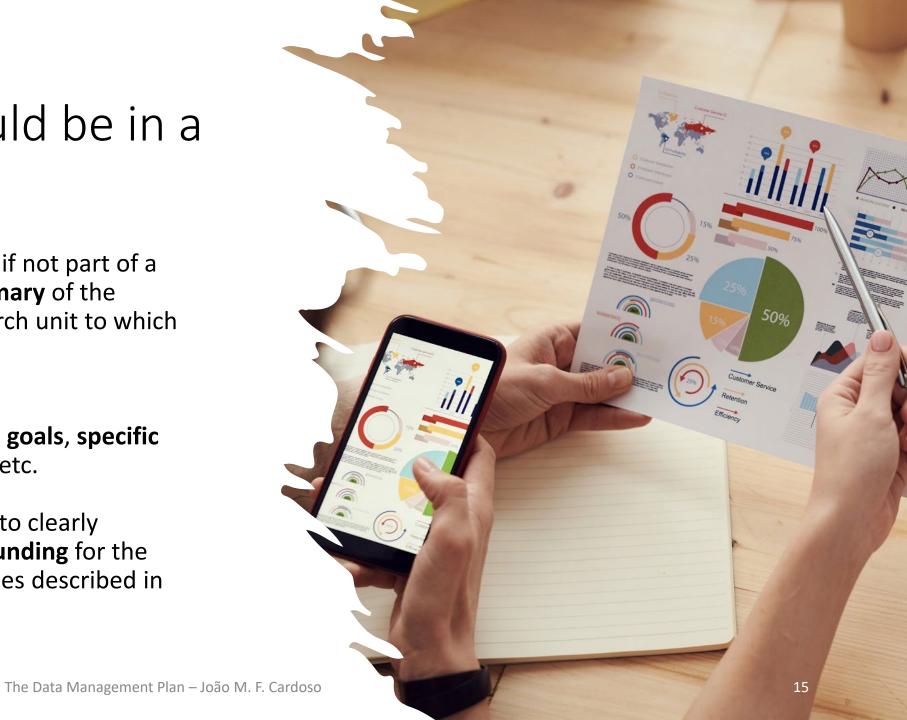
The DMP should include (if not part of a project proposal), a **summary** of the research project or research unit to which it pertains.

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.







**Project Description** 

**Existing Data** 

**Created Data** 

**Data Organisation** 

Preservation and Access Policies

Licences and Ethics

Resources and Responsibilities





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.







**Project Description** 

**Existing Data** 

**Created Data** 

**Data Organisation** 

Preservation and Access Policies

Licences and Ethics

Resources and Responsibilities





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**.







**Project Description** 

**Existing Data** 

**Created Data** 

Data Organisation

Preservation and Access Policies

Licences and Ethics

Resources and Responsibilities





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.







**Project Description** 

**Existing Data** 

**Created Data** 

**Data Organisation** 

Preservation and Access Policies

Licences and Ethics

Resources and Responsibilities





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.







**Project Description** 

**Existing Data** 

**Created Data** 

**Data Organisation** 

Preservation and Access Policies

Licences and Ethics

Resources and Responsibilities





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.







**Project Description** 

**Existing Data** 

**Created Data** 

**Data Organisation** 

Preservation and Access Policies

Licences and Ethics

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.







**Project Description** 

**Existing Data** 

**Created Data** 

**Data Organisation** 

Preservation and Access Policies

Licences and Ethics

Resources and Responsibilities





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







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

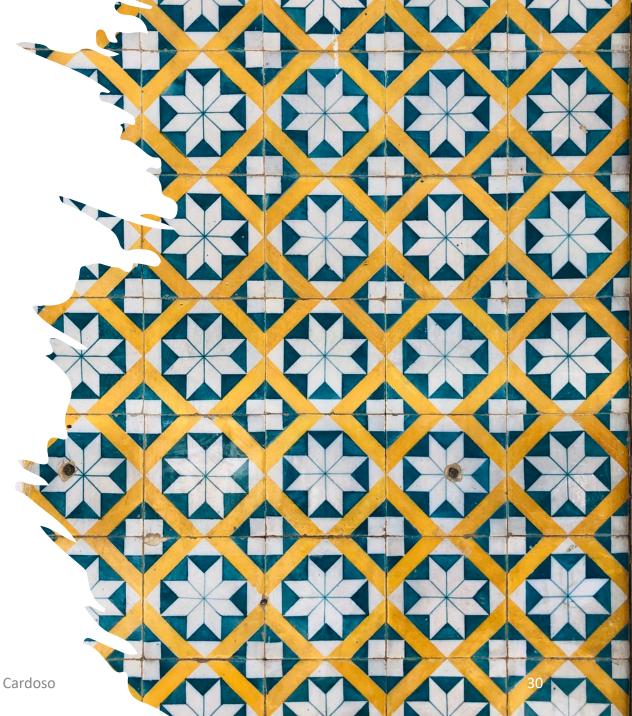
There was little to **no consistency** between DMPs

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

They were not accessible.







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.









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.

H2020 templates: Data management plan v1.0 - 13.10.2016

#### TEMPLATE HORIZON 2020 DATA MANAGEMENT PLAN (DMP)

- Instructions and footnotes in blue must not appear in the text.
- For options [in square brackets]: the option that applies must be chosen.
- For fields in [grey in square brackets] (even if they are part of an option as specified in the previous item): enter the appropriate data.

#### Introduction

This Horizon 2020 DMP template has been designed to be applicable to any Horizon 2020 project that produces, collects or processes research data. You should develop a single DMP for your project to cover its overall approach. However, where there are specific issues for individual datasets (e.g. regarding openness), you should clearly spell this out.

Guidelines on FAIR Data Management in Horizon 2020 are available in the Online Manual.

#### FAIR data manageme

In general terms, your research data should be 'FAIR', that is findable, accessible, interoperable and re-usable These principles precede implementation choices and do not necessarily suggest any specific technolog standard, or implementation-solution.

This template is not intended as a strict technical implementation of the FAIR principles, it is rather inspired be FAIR as a general concept.

More information about FAIR

FAIR data principles (FORCE11 discussion forum)

FAIR principles (article in Nature)

#### Structure of the template

The template is a set of questions that you should answer with a level of detail appropriate to the project.

It is not required to provide detailed answers to all the questions in the first version of the DMP that needs to be submitted by month 6 of the project. Rather, the DMP is intended to be a living document in which information can be made available on a finer level of granularity through updates as the implementation of the project progresses and when significant changes occur. Therefore, DMPs should have a clear version number and include a timetable for updates. As a minimum, the DMP should be updated in the context of the periodic evaluation/assessment of the project. If there are no other periodic reviews envisaged within the grant agreement, an update needs to be made in time for the final review at the latest.

In the following the main sections to be covered by the DMP are outlined. At the end of the document, Table 1 contains a summary of these elements in bullet form.

This template itself may be updated as the policy evolves.



Commission

1





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**.

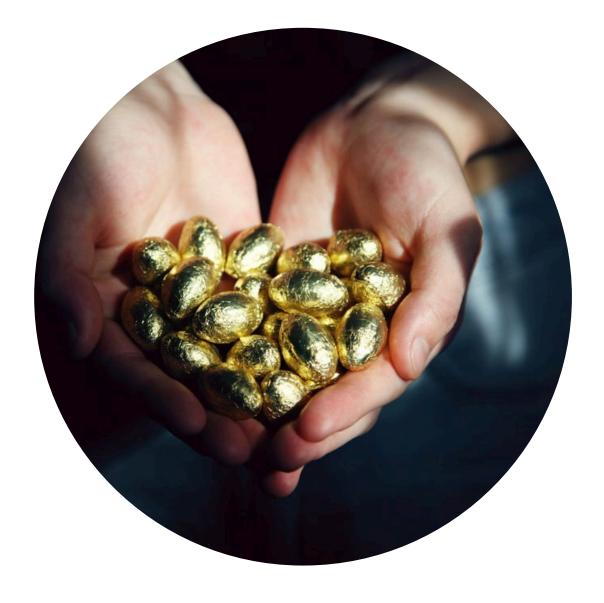






The combination of these issues leads to the overall majority of created DMP documents having poor quality and low practical value.

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







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

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

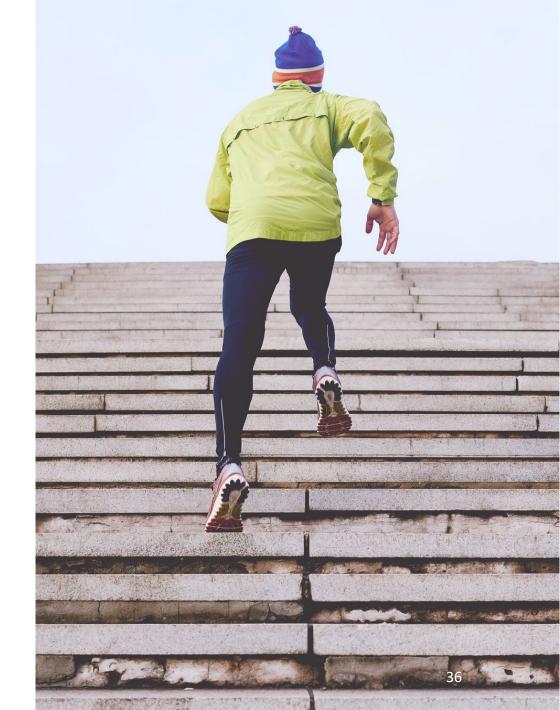






So what is being done to counter this trend?

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







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.

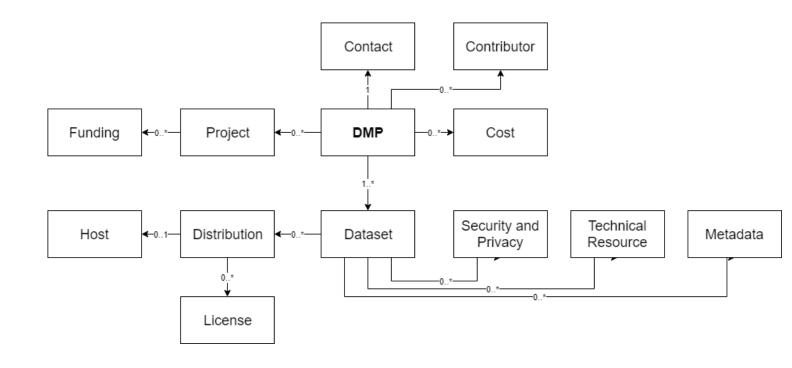






The standard defined a DMP through 13 core terms.

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









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.











The final step was to have machineactionable 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.







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









### 7. **Ready for BioData Management?**

Capacity Building for the Life Sciences













#### Introduction to DMP

An introductory course, where participants are presented to the basics of data management and data management plans.

Collaborative creation of a DMP, based on a fictitious project.

#### **Advanced DMP Creation**

An advanced course, where participants are guided through the creation of a DMP in a selected DMP creation tool.

Participants are encouraged to bring their own data.

http://ready4biodatamanagement.biodata.pt/



#### 8. In conclusion

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.









