



RDA's Approach to Machine Actionable Data Management Plans

João Cardoso, joao.m.f.cardoso@tecnico.ulisboa.pt RDA's DMP Common Standards WG member



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**.



Workshop machine-actionable Software Management Plans - Metadata C



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.



European Commission





H2020 templates: Data management plan v1.0 13.10.2016

TEMPLATE HORIZON 2020 DATA MANAGEMENT PLAN (DMP)

1. The reality of DMPs

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.

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: collector or processors 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 spall this out.

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

FAIR data management

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 technology, standard, or implementation-solution.

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

More information about FAIR:

FAIR data principles (FORCE11 discussion forum)

FAIR principles (article in Nature)

Structure of the template

European Commission 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 all vining document in which information can be made available on a finar 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 to 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.



1

The existence of **multiple** 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**.





Workshop machine-actionable Software Management Plans

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.

So **what is being done** to counter this trend? How is the community **making DMPs better**?





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

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





2. Defining the DMP

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

The **RDA DMP Common Standards Working Group (DCS WG)** was tasked with addressing that challenge.

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

This would serve as the **starting point** for future extensions.





Workshop machine-actionable Software Management Plans - Metadata Croswalk

2. Defining the DMP

The resulting **DCS application profile**, defines the DMP through 13 core terms.

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

https://doi.org/10.5334/dsj-2021-032







2. Defining the DMP

The **DCS application profile** is documented in full in the DCS WG **GitHub repository**.

The DCS application profile is **not to be interpreted** as either as a **questionnaire** or a **DMP template**.

It is intended as a **universal definition of a DMP**, so as to act as an **interchange** format.



https://github.com/RDA-DMP-Common/



3. The maDMP

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

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



3. The maDMP

- Enforcing persistent identification of artefacts
- Facilitating the updating process
- Facilitating the creation process



3. The maDMP

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 DCS application profile had **interoperability challenges**.

- Lack of explicit linking with existing ontologies
- Lack of mechanisms to describe controlled vocabularies
- Lack of formal mechanisms for its extension





The community created a **semantic-based maDMP representation** that would take advantage of its distinct characteristics and **address the challenges**.

This effort resulted in the DMP Common Standard Ontology (DCSO).

Its functional requirements were:

- Reuse of terms of referenced domain ontologies
- Allow and enforce the usage of controlled vocabularies
- Be extendable



The DCSO is a representation of the DCS application profile.

It is defined through 24 **classes**, 20 **object properties** and 42 **data properties**.

https://doi.org/10.1186/s13326-022-00274-4





The DCSO is a representation of the DCS application profile.

It is defined through 24 **classes**, 20 **object properties** and 42 **data properties**.

dcso:Cost dcso:DMP dcso:Funding dcso:Host dcso:Licence dcso:Metadata dcso:Project dcso:SecurityPrivacy dcso:TechnicalResource dcso:Id dcso:ContactId dcso:ContributorId dcso:DatasetId dcso:DMPId dcso:FunderId dcso:GrantId dcso:MetadataStandardId dcat:Dataset dcso:Dataset dcat:Distribution dcso:Distribution foaf:Agent dcso:Contact dcso:Contributor



13 of its classes can be **directly mapped** to fields in the DCS application profile.

8 of the classes are 'identifier' classes.

3 classes are directly imported from **external domain ontologies.** Namely:

- W3C DCAT Specification
- DCMI Metadata Terms
- Friend of a Friend





13 of its classes can be **directly mapped** to fields in the DCS application profile.

8 of the classes are 'identifier' classes.

3 classes are directly imported from **external domain ontologies.** Namely:

- W3C DCAT Specification
- DCMI Metadata Terms
- Friend of a Friend





13 of its classes can be **directly mapped** to fields in the DCS application profile.

8 of the classes are 'identifier' classes.

3 classes are directly imported from **external domain ontologies.** Namely:

- W3C DCAT Specification
- DCMI Metadata Terms
- Friend of a Friend





The DCSO also comes with a **constraints validation layer**, that aims to **validate** and **enforce the compliance** with the DCS application profile.

The constraints validation layer is expressed in ShEx.

The created ShEx schemas follow directives established in the DCS application profile, focusing on defining conditions on:

- Element relations;
- Cardinality;
- Existence.



Workshop machine-actionable Software Management Plans - Metadata Croswalk

E610 M



5. maDMP adoption

There have been multiple entities striving to have easier, and consistent means to create and use maDMP documents.

The following are a **selection of the systems and services** that have **adopted** the DCS application profile.





Workshop machine-actionable Software Management Plans - Metadata Croswalk

6. Next steps

What's next for the DCS WG?

- Maintenance of the DCS WG documentation;
 - Needs to be reorganized, to make it easier to contribute and adopt.
- Further the adoption of the DCS application profile;
 - We need more (and distinct) use cases.
- Extend the DCS application profile.

RESEARCH DATA ALLIANCE

• Maybe with SMP integration?



7. In conclusion

Having an active community focusing on **how to best integrate maDMPs in the research data management** workflow is paramount.

The DCS WG is here to **aid in the coordination** of such an effort.

All of the **documentation** is **available online**

https://github.com/RDA-DMP-Common/

Don't forget to attend the IDW2023!

https://internationaldataweek.org/











Workshop machine-actionable Software Management Plans - Metadata Croswalk