



# Integrating RDM Services using maDMPs

**Tomasz Miksa**

DMP Common Standards WG

2022 Webinar Series Highlighting RDA Outputs

## ➤ Introduction

### ➤ Data Management Plans

## ➤ Machine-actionable DMPs

### ➤ Context

### ➤ RDA Recommendation

#### ➤ What it is

#### ➤ Example

## ➤ Integrating RDM Services

## ➤ RDA working group and further reading

# Data Management Plans (DMPs)

	Data Officer	Who is responsible for the data management and the DMP of the project (name/email address)?
I	Data Characteristics	
I.1	Description of the data	What kinds of data/source code will be generated or reused (type, format, volume)? How will the research data be generated and which methods will be used? How will you structure the data and handle versioning? Who is the target audience?
II	Documentation and Metadata	
II.1	Metadata standards	What metadata standards (if any) will be in use and why? (see <a href="#">Digital Curation Centre</a> )
II.2	Documentation of data	What information is needed for the data to be findable, accessible, interoperable and re-usable ( <a href="#">FAIR</a> ) in the future? Is the data machine-readable? How are you planning to document this information?
II.3	Data quality control	What quality assurance processes will you adopt? How will the consistency and quality of data collection be controlled and documented? (This may include processes such as repeat samples or measurements, standardised data capture, peer review of data or representation with controlled vocabularies.)
III	Data Availability and Storage	
III.1	Data sharing strategy	How and when will the data be shared and made accessible? What repository will you be using? What persistent identifier will be used?
III.2	Data storage strategy	What data are to be preserved for the long-term, and what data will not be stored? How and where will the data be stored and backed up during the research? How and where will the data be stored after the project ends? For how long will the data be stored? Are there any costs that need to be covered for storage? At what point during or after the project will the data be stored? Are there any technical barriers to making the research data fully or partially accessible?

## Directorate for Engineering Data Management Plans Guidance for Principal Investigators *updated: November 2018*

The Directorate for Engineering (ENG) supports research covering a broad spectrum of communities of investigators, and each community has its own best practices. ENG is aware of the need to provide flexibility to programs, principal investigators (PIs), and reviewers in assessing the quality of individual Data Management Plans (DMPs) from various communities. Therefore, guidance has evolved to accommodate changing community standards and expectations. ENG relies on the merit review process to determine which DMPs best serve each community.

The following guidance is to assist ENG investigators, reviewers and Program Officers in developing and evaluating effective, complete, and competitive DMPs. It is important to recognize that while all DMPs should address the five categories of information as specified in the PAPPG, they should not be generic. Each DMP should appropriately identify the data, metadata, samples, software, algorithms, curricula, documentation, publications, and other materials generated in the course of the proposed research. Moreover, the DMPs should describe how these materials will be disseminated, made accessible, and archived while incorporating the best practices and standards for the proposed research. DMPs are subject to peer review. Please contact your specific Program Officer if you have any questions related to DMPs in the program context.

### PAPPG and NSF-WIDE REQUIREMENTS

All proposals must include a supplementary document of no more than two pages labeled "Data Management Plan," as described in [PAPPG Chapter II.C.2.j](#). The DMP is NOT part of the 15-page Project Description. *Proposals that do not include a Data Management Plan will be returned without review.*

You may request funds to cover costs of publication, page charges, or preparation of data as a direct cost in your budget proposal, which is evaluated as part of the merit review process. Any costs associated with implementing the DMP should be explained in the Budget Justification.

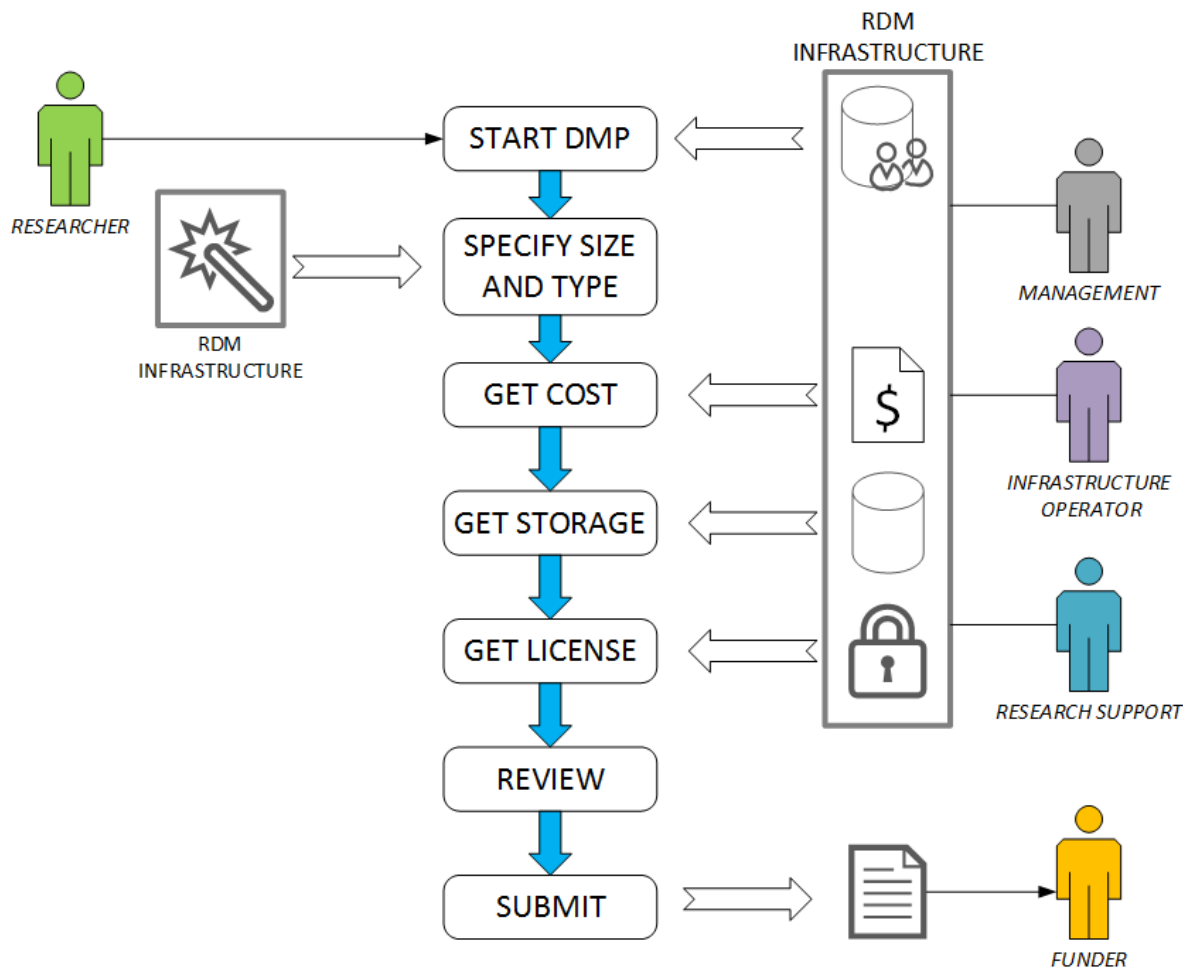
Some NSF Program Solicitations may contain specific and/or additional instructions that deviate from this guidance and/or provide exceptions to the two-page limit. Instructions in the solicitation take precedence over this guidance. Please check solicitations carefully for this information.

### DATA MANAGEMENT PLAN (DMP) CONTENT

The DMP should clearly articulate how the investigators plan to manage and disseminate both the physical and digital data generated by the project, taking advantage of emerging information

# Machine-actionable DMPs (maDMPs)

- Machine-actionable DMPs
  - Living documents
  - automate data management
    - collect information from systems
    - trigger actions in systems
  - facilitate validation
- This requires
  - well-defined RDM workflows
  - data management infrastructure
  - common standard to represent information



# DMPs vs maDMPs

## Traditional DMP

<administrative\_data>

<question>Who is responsible for the DMP?</question>

<answer>Moritz from our university.</answer>


</administrative\_data>

## Machine-actionable DMP

```
"contributor" : [ {  
  "contributor_id" : {  
    "identifier" : "0000-0002-5164-2690",  
    "type" : "orcid"  
  },  
  "mbox" : "moritz.staudinger@tuwien.ac.at",  
  "name" : "Moritz Staudinger",  
  "role" : [ "Data Manager" ]  
},
```

maDMPs use PIDs and controlled vocabularies.


Example shows that Moritz is the one responsible for data management.



## RDA DMP Common Standard for Machine-actionable Data Management Plans

**The Challenge:**

Data Management Plans are free-form text documents describing the data that is used and produced during the course of research activities. They specify where the data will be archived, which licenses and constraints apply, and to whom credit should be given, etc. The workload and bureaucracy often associated with traditional DMPs can be reduced when they become machine-actionable.



Produced by: **DMP Common Standards WG**  
<https://www.rd-alliance.org/groups/dmp-common-standards-wg>

## RDA DMP Common Standard for Machine-actionable Data Management Plans

Recommendations of the RDA DMP Common Standards WG  
*Tomasz Miksa, Paul Walk, Peter Neish*

**Purpose**

This application profile is meant for exchange of machine-actionable DMPs between systems. It is independent of any internal data organisation used by these systems. The application profile does not prescribe how information must be presented to the end user and does not enforce any specific logic on how this information must be collected or used. The application profile is an information carrier and the full machine-actionability can only be achieved when systems using the application profile implement appropriate logic.

This application profile is intended to cover a wide range of use cases and does not set any business (e.g. funder specific) requirements. It represents information over the whole DMP lifecycle, that is, it can express planned actions, as well as actions already performed.

The application profile is NOT intended to be a prescriptive template or a questionnaire, but to provide a re-usable way of representing machine-actionable information on themes covered by DMPs.

**Overview**

Figure 1 presents concepts used within the application profile. Each concept is further broken down into specific fields (not depicted). The full application profile specification can be found [online](#). Below we outline main concepts used within the application profile that are depicted in Figure 1.

**DMP** - Provides high level information about the DMP, e.g. its title, modification date, etc. It is the root of this application profile.

**Project** - Describes the project associated with the DMP, if applicable. It can be used to describe any type of project: that is, not only funded projects, but also internal projects, PhD theses, etc.

**Funding** - For specifying details on funded projects, e.g. NSF or EC funded projects.

**Contact** - Specifies the party which can provide information on the DMP.

**Contributor** - For listing all parties involved in the process of data management described by

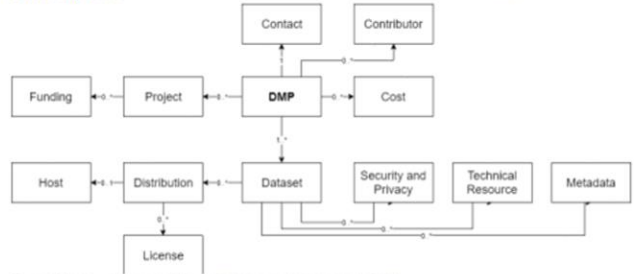


Figure 1: Overview of the application profile for the machine-actionable DMPs.

1

# Adoptions (selected)



# maDMPs - documentation

## Properties in 'dmp'

Name	Description	Data Type	Cardinality	Example Value
<a href="#">contact</a>	Contact person for a DMP	Nested Data Structure	1	
<a href="#">contributor</a>	To list people that play role in data management related to this DMP, e.g. responsible for performing actions described in this DMP.	Nested Data Structure	0..n	
<a href="#">cost</a>	To list costs related to data management. Providing multiple instances of a 'Cost' allows to break down costs into details. Providing one 'Cost' instance allows to provide one aggregated sum.	Nested Data Structure	0..n	
<a href="#">created</a>	Date and time of the first version of a DMP. Must not be changed in subsequent DMPs.	DateTime	1	2019-03-13 13:13
<a href="#">dataset</a>	To describe data on a non-technical level.	Nested Data Structure	1..n	

NOT a questionnaire!  
NOT a template!

Most fields are optional!



# Machine-actionable DMP

```
"dataset" : [ {  
  "description" : "For each dataset (fish and employee) the original dataset will be split into two subsets, one for training and one for testing the  
  performance.",  
  "distribution" : [ {  
    "access_url" : "https://zenodo.org/record/6467615",  
    "byte_size" : 2999302,  
    "data_access" : "open",  
    "description" : "For each dataset (fish and employee) the original dataset will be split into two subsets, one for training and one for testing the  
    performance.",  
    "format" : [ "STRUCTURED_TEXT" ],  
    "host" : {  
      "description" : "ZENODO builds and operates a simple and innovative service that enables researchers, scientists, EU projects and institutions to share  
      and showcase multidisciplinary research results (data and publications) that are not part of the existing institutional or subject-based repositories of  
      the research communities.\nZENODO enables researchers, scientists, EU projects and institutions to:\neasily share the long tail of small research results  
in a wide variety of formats including text, spreadsheets, audio, video, and images across all fields of science.\ndisplay their research results and get  
credited by making the research results citable and integrate them into existing reporting lines to funding agencies like the European  
Commission.\neasily access and reuse shared research results.",  
      "pid_system" : [ "doi" ],  
      "storage_type" : "other",  
      "support_versioning" : "unknown",  
      "title" : "Zenodo",  
      "url" : "https://zenodo.org/"  
    },  
  },  
  "license" : [ {  
    "license_ref" : "https://creativecommons.org/licenses/by/4.0/",  
    "start_date" : "2022-05-01 22:00:00.0"  
  } ],  
  "title" : "Training and Test Subsets for Performance Comparison of kNN and GD"
```

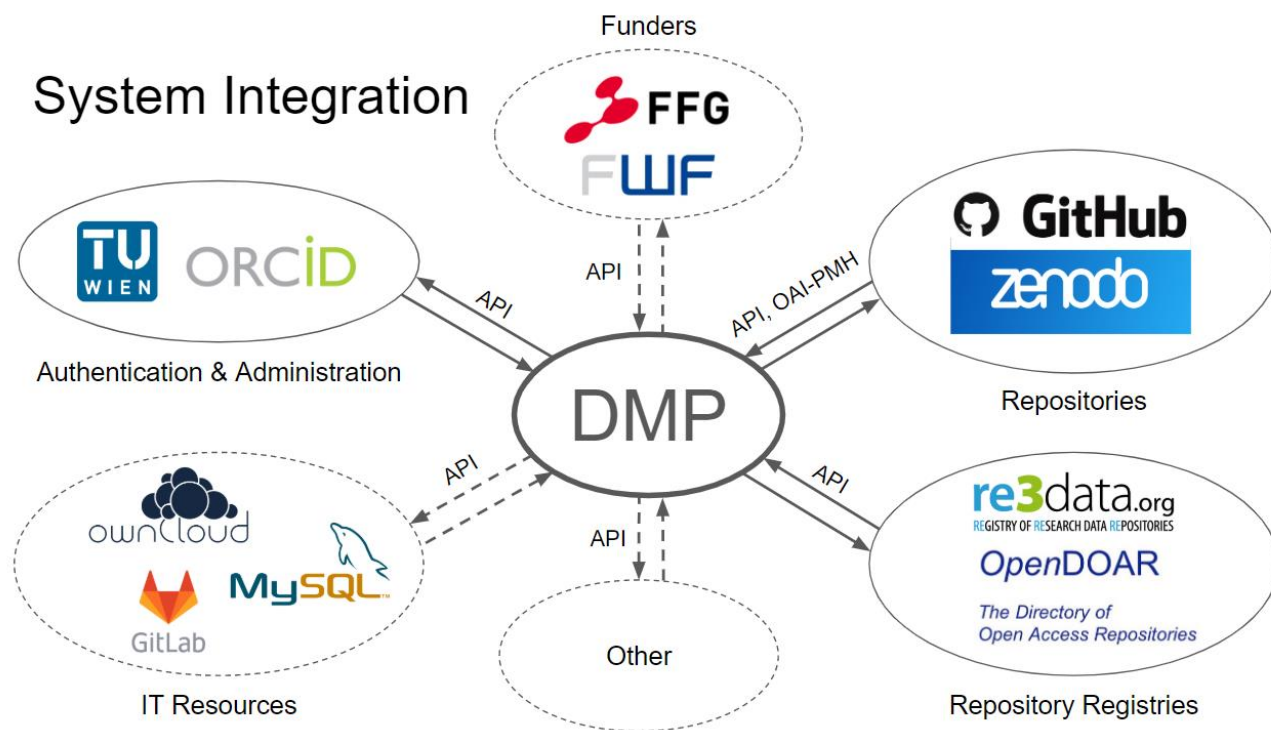
Each **dataset** has a **title** and a human readable **description**.

It is also clear what the **format**, **size** and the **location** of the dataset are.

**License** and mode of **access**, including any exact **embargo** periods, are specified as well.

# RDM Infrastructure

- maDMPs are the 'glue' between different systems
  - Automate getting information **in** and **out**

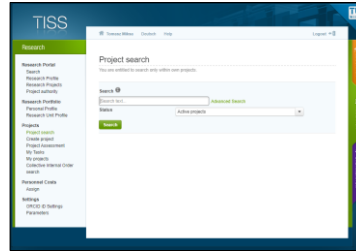


# Example: TU Wien in Austria

Knowledge Graph



CRIS



IT Storage



API

API

list of services

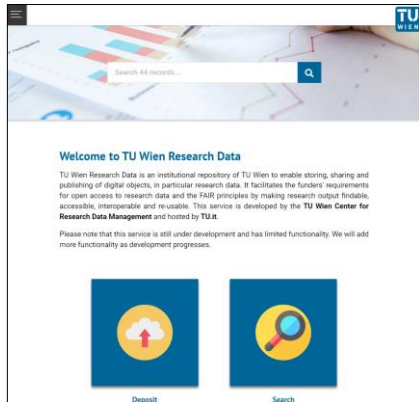
maDMP

maDMP

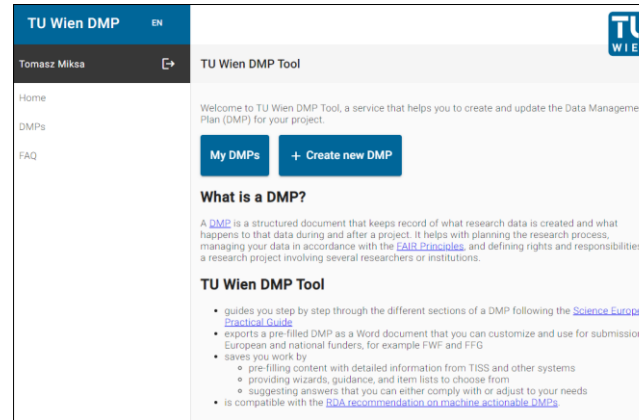
DMP

Funder integration?

Automated assessment and feedback?



Data Repository



[www.damap.org](http://www.damap.org)

› Slides from all our sessions are in the repository

<https://www.rd-alliance.org/node/56938/file-repository>

› Join the group!

We have **250+** members!



› Contact

› [tomasz.miksa@tuwien.ac.at](mailto:tomasz.miksa@tuwien.ac.at)

WG

DMP Common Standards WG

Taxonomy:

Posts

Wiki

Events

Repository

Outputs

Case Statements

Plenaries

Members

create new content

Group Status:

WGs Maintaining deliverables (maintenance group)

You are the group manager

›

status: Recognised & Endorsed

Chair (s): Paul Walk, Peter Neish, Tomasz Miksa

Group Email: [dmp-common@rda-groups.org](mailto:dmp-common@rda-groups.org)

Secretariat Liaison: [enquiries\[at\]rd-alliance.org](mailto:enquiries[at]rd-alliance.org)

File Repository

VP17 Edinburgh

by Tomasz Miksa

22 April 2021

Attachment	Size
2021-RDA-DMP-VP17.pdf	3.71 MB
1-Zeno-FairDataAustria-DMAP.pdf	1.05 MB
3-Elli-Argos.pdf	1.66 MB
4-Fajar-DCSOntology.pdf	1.82 MB

VP16 Costa Rica Slides

by Tomasz Miksa

12 November 2020

Slides from the plenary session at the VP16: \* 2020-RDA-DMP-VP16 - main deck of slides 1 - Claire Austin - maDMPs in a government context 2 - Maria Praetzelis - DMPHub 3 - Lucas Berent, Alexandre

Attachment	Size
1-madmps-goverment.pdf	962.26 KB
2-dmphub.pdf	1.34 MB
3-madmps-exposing.pdf	1.87 MB
4-madmps-repositories.pdf	794.44 KB
5-argos-knowledge-graph.pdf	2.09 MB
6-ro-crates-and-madmps.pdf	1.16 MB
2020-RDA-DMP-VP16.pdf	2.78 MB

- Key elements of the whole ecosystem needed to make DMPs machine-actionable

- Enterprise Architecture that uses maDMPs
- Examples of tasks automation at institutions using maDMPs

***ACM Transactions  
on Management Information Systems***

<https://doi.org/10.1371/journal.pcbi.1006750>

<https://doi.org/10.1145/3490396>

# Publications about maDMPs

- [Tomasz Miksa, Simon Oblasser, and Andreas Rauber. \*\*Automating research data management using machine-actionable data management plans\*\*. ACM Transactions on Management Information Systems, 13\(2\), dec 2021.](#)
- [Tomasz Miksa, Paul Walk, Peter Neish, Simon Oblasser, Hollydawn Murray, Tom Renner, Marie-Christine Jacquemot-Perbal, João Cardoso, Trond Kvamme, Maria Praetzellis, Marek Suchánek, Rob Hooft, Benjamin Faure, Hanne Moa, Adil Hasan, and Sarah Jones. \*\*Application profile for machine-actionable data management plans\*\*. CODATA Data Science Journal, 20\(1\):32, October 2021](#)
- [Raffael Foidl, Lea Salome Brugger, and Tomasz Miksa. \*\*Automating Evaluation of Machine-Actionable Data Management Plans with Semantic Web Technologies\*\*. In DaMaLOS - 2nd Workshop on Data and Research Objects Management for Linked Open Science : Co-located at the International Semantic Web Conference ISWC 2021. PUBLISSO, November 2021.](#)
- [Tomasz Miksa, Maroua Jaoua, and Ghaith Arfaoui. \*\*Research Object Crates and Machine-actionable Data Management Plans\*\*. In DaMaLOS - First Workshop on Data and Research Objects Management for Linked Open Science : Co-located at the International Semantic Web Conference ISWC 2020. PUBLISSO, November 2020.](#)
- [João Cardoso, Leyla Jael Garcia Castro, Fajar Ekaputra, Marie-Christine Jacquemot-Perbal, Tomasz Miksa, and José Borbinha. \*\*Towards semantic representation of machine-actionable Data Management Plans\*\*. In DaMaLOS - First Workshop on Data and Research Objects Management for Linked Open Science : Co-located at the International Semantic Web Conference ISWC 2020. PUBLISSO, 2020.](#)
- [Simon Oblasser, Tomasz Miksa, Asanobu Kitamoto: \*\*Finding a Repository with the Help of Machine-Actionable DMPs: Opportunities and Challenges\*\*. IDCC 2020](#)
- [Tomasz Miksa, Stephanie Simms, Daniel Mietchen, Sarah Jones \(2019\) \*\*Ten principles for machine-actionable data management plans\*\*. PLOS Computational Biology 15\(3\): e1006750.](#)
- [Tomasz Miksa, Peter Neish, Paul Walk, Andreas Rauber: \*\*Defining requirements for machine-actionable Data Management Plans\*\*. iPres 2018](#)
- [Tomasz Miksa, João Cardoso, José Luis Borbinha: \*\*Framing the scope of the common data model for machine-actionable Data Management Plans\*\*. BigData 2018: 2733-2742](#)
- [Asztrik Bakos, Tomasz Miksa, Andreas Rauber: \*\*Research Data Preservation Using Process Engines and Machine-Actionable Data Management Plans\*\*. TPDL 2018: 69-80](#)