RESEARCH DATA ALLIANCE

maDMPs - where are we now?

Tomasz Miksa

DMP Common Standards WG

@miksa_tomasz

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 Digital Curation Centre)*
II.2	**Documentation of data**	*What information is needed for the data to be findable, accessible, interoperable and re-usable (FAIR) 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.

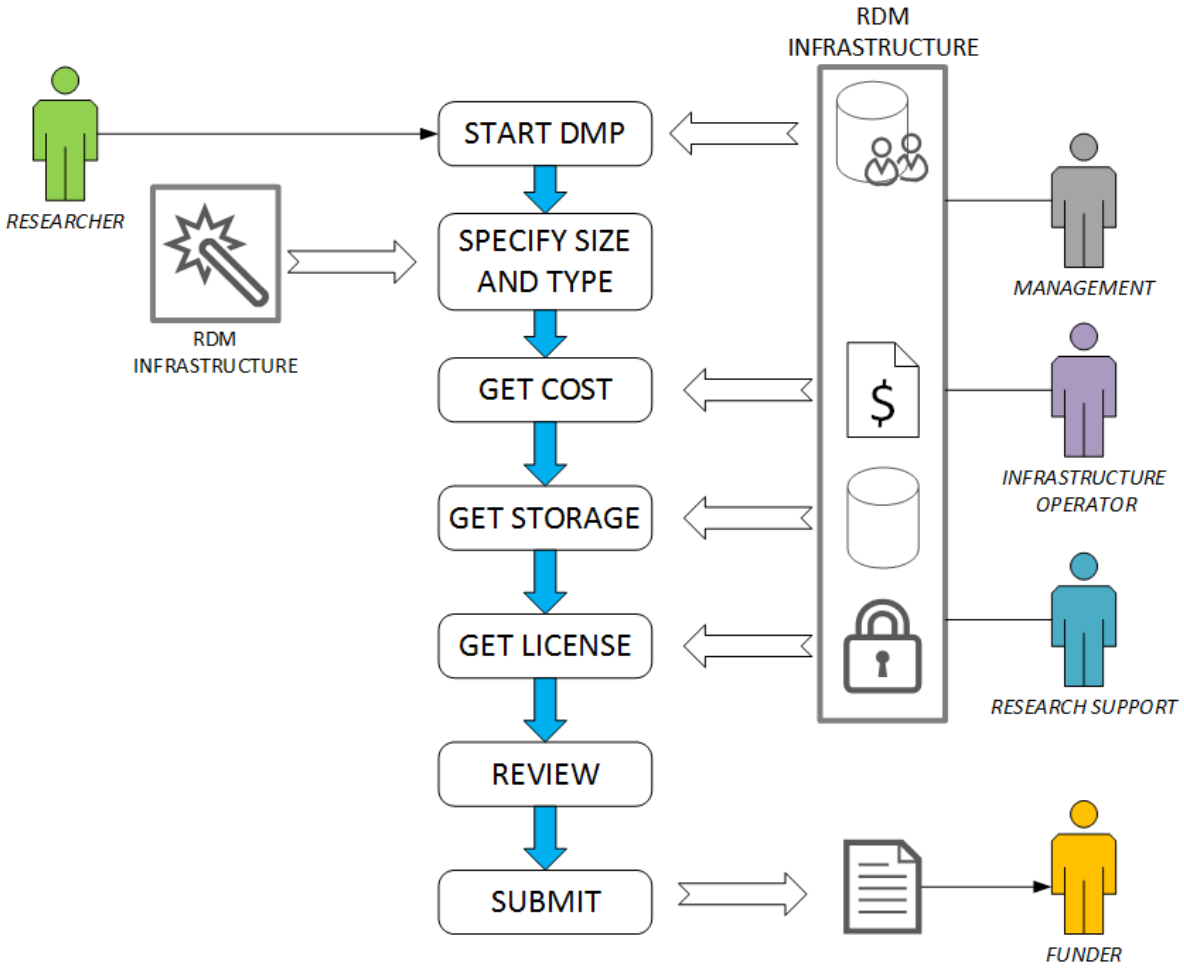
Some NSF Program Solicitations may contain specific and/or additional instructions that deviate from this guidance and/or provide exceptions to the two-page limit. Instructions in the solicitation take precedence over this guidance. Please check solicitations carefully for this information.

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.

Official RDA Recommendation on maDMPs

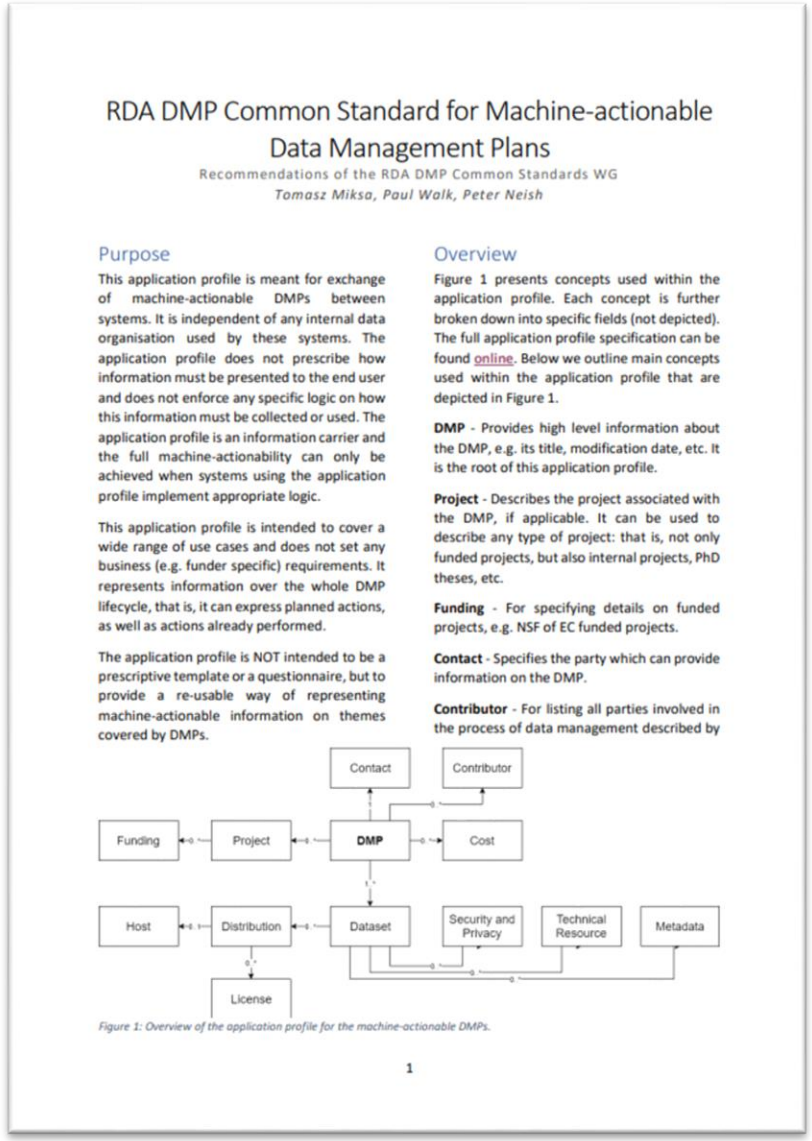
RDA DMP Common Standard for Machine-actionable Data Management Plans

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 of 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

Miksa, T., Walk, P., & Neish, P. (2019). RDA DMP Common Standard for Machine-actionable Data Management Plans. https://doi.org/10.15497/rda00039

Adoptions (selected)

maDMPs - documentation

Properties in 'dmp'

Name	Description	Data Type	Cardinality	Example Value
contact	Contact person for a DMP	Nested Data Structure	1	
contributor	To list people that play role in data management related to this DMP, e.g. resoponsible for performing actions described in this DMP.	Nested Data Structure	0..n	
cost	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	
created	Date and time of the first version of a DMP. Must not be changed in subsequent DMPs.	DateTime	1	2019-03-13 13:13
dataset	To describe data on a non-technical level.	Nested Data Structure	1..n	

NOT a questionnaire!
NOT a template!

Most fields are optional!

https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard/blob/master/docs/index.md

Machine-actionable DMP

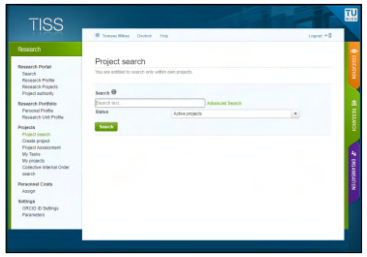

```
"dataset" : [ {
  "description" : "For each dataset (fish and employee) the original dataset will be split into two subsets, one for training and one for testing the
  performance.",
  "distribution" : [ {
    "access_url" : "https://zenodo.org/record/6467615",
    "byte_size" : 2999302,
    "data_access" : "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.

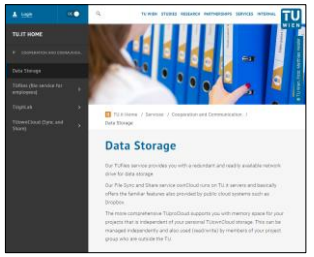
Example: TU Wien in Austria

Knowledge Graph

CRIS



IT Storage

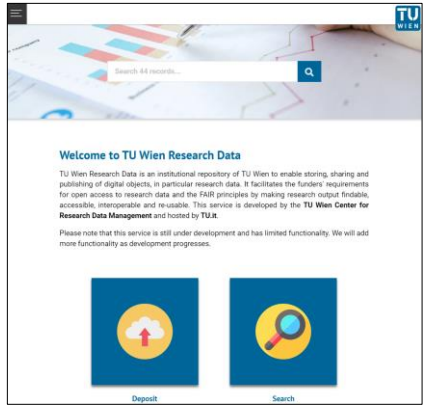


Funder integration?

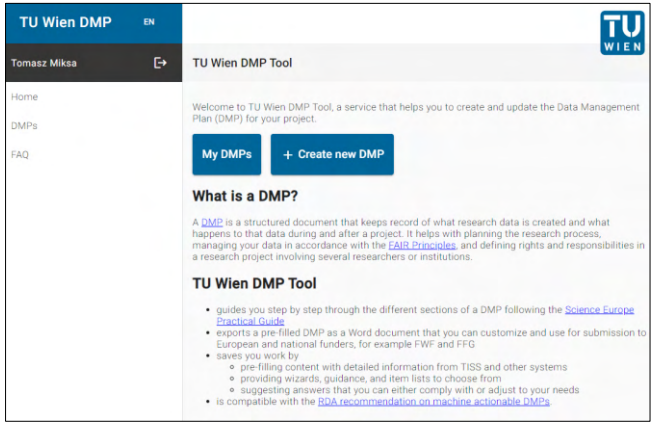
API

API

list of services



maDMP



maDMP

DMP

Automated assessment and feedback?

Data Repository

www.damap.org

DMP Common Standards WG

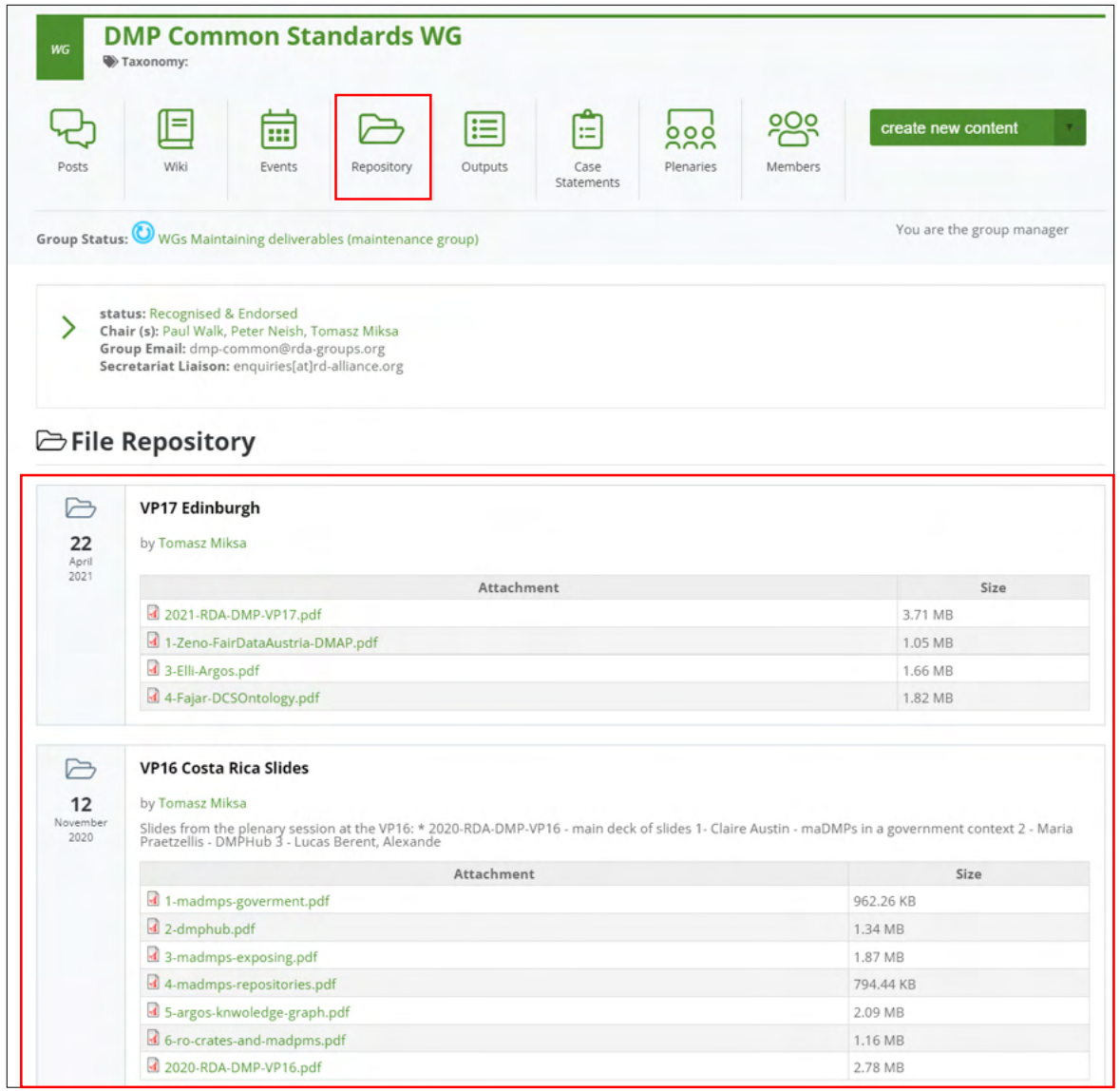
- Slides from all our sessions are in the repository

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

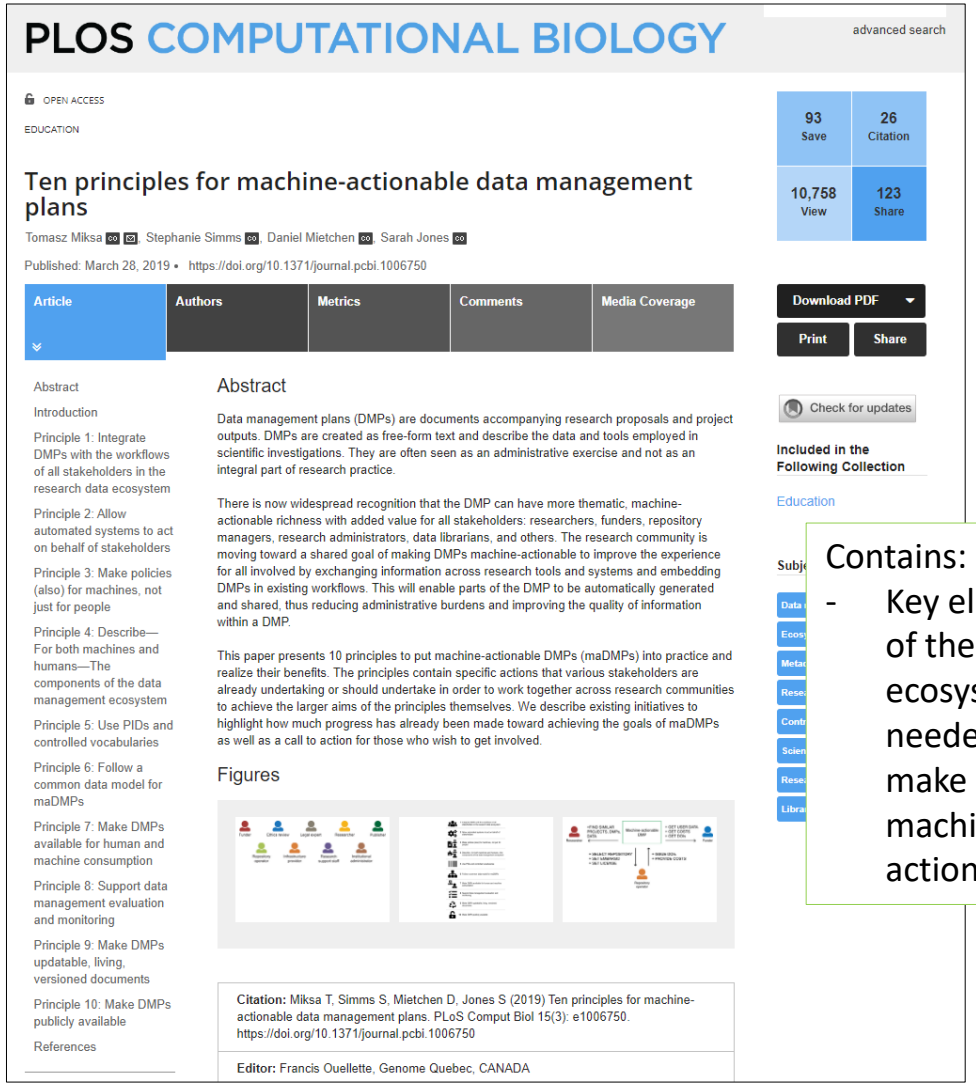
- Join the group!

We have **250+** members! ❤️😮

- Contact
  - tomasz.miksa@tuwien.ac.at



Read more in…



PLOS COMPUTATIONAL BIOLOGY

advanced search

OPEN ACCESS

EDUCATION

Ten principles for machine-actionable data management plans

Tomasz Miksa, Stephanie Simms, Daniel Mietchen, Sarah Jones

Published: March 28, 2019 • https://doi.org/10.1371/journal.pcbi.1006750

93 Save | 26 Citation | 10,758 View | 123 Share

Article | Authors | Metrics | Comments | Media Coverage

Download PDF | Print | Share

Check for updates

Included in the Following Collection

Education

Abstract

Data management plans (DMPs) are documents accompanying research proposals and project outputs. DMPs are created as free-form text and describe the data and tools employed in scientific investigations. They are often seen as an administrative exercise and not as an integral part of research practice.

There is now widespread recognition that the DMP can have more thematic, machine-actionable richness with added value for all stakeholders: researchers, funders, repository managers, research administrators, data librarians, and others. The research community is moving toward a shared goal of making DMPs machine-actionable to improve the experience for all involved by exchanging information across research tools and systems and embedding DMPs in existing workflows. This will enable parts of the DMP to be automatically generated and shared, thus reducing administrative burdens and improving the quality of information within a DMP.

This paper presents 10 principles to put machine-actionable DMPs (maDMPs) into practice and realize their benefits. The principles contain specific actions that various stakeholders are already undertaking or should undertake in order to work together across research communities to achieve the larger aims of the principles themselves. We describe existing initiatives to highlight how much progress has already been made toward achieving the goals of maDMPs as well as a call to action for those who wish to get involved.

Figures

Citation: Miksa T, Simms S, Mietchen D, Jones S (2019) Ten principles for machine-actionable data management plans. PLoS Comput Biol 15(3): e1006750. https://doi.org/10.1371/journal.pcbi.1006750

Editor: Francis Ouellette, Genome Quebec, CANADA

Contains:

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

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

1

Automating Research Data Management Using Machine-actionable Data Management Plans

TOMASZ MIKSA, TU Wien & SBA Research, Austria
SIMON OBLASSER, TU Wien, Austria
ANDREAS RAUBER, TU Wien, Austria

Many research funders mandate researchers to create and maintain Data Management Plans (DMPs) for research projects that describe how research data is managed to ensure its reusability. A DMP being a static textual document is difficult to act upon and can quickly become obsolete and impractical to maintain. A new generation of machine-actionable DMPs was therefore proposed by the Research Data Alliance to enable automated integration of information and updates. Machine-actionable DMPs open up a variety of use cases enabling interoperability of research systems and automation of data management tasks.

In this paper we describe a system for machine-actionable data management planning in an institutional context. We identify common use cases within research that can be automated to benefit from machine-actionability of DMPs. We propose a reference architecture of a machine-actionable DMP support system that can be embedded into an institutional research data management infrastructure. The system semi-automates creation and maintenance of DMPs, and thus eases the burden for the stakeholders responsible for various DMP elements. We evaluate the proposed system in a case study conducted at the largest technical university in Austria and quantify to what extent the DMP templates provided by the European Commission and a national funding body can be pre-filled. The proof-of-concept implementation shows that machine-actionable DMP workflows can be semi-automated, thus workload on involved parties can be reduced and quality of information increased. The results are especially relevant to decision makers and infrastructure operators who want to design information systems in a systematic way that can utilise the full potential of machine-actionable DMPs.

CCS Concepts: • **Applied computing** → **Enterprise data management**; **Business process management**; **IT architectures**; • **Information systems** → **Digital libraries and archives**; • **Social and professional topics** → **Automation**.

Additional Key Words and Phrases: data management plan, machine-actionable, business processes, enterprise architecture, funder template, requirements engineering, automation, RDM, RDA, FAIR

ACM Reference Format:
Tomasz Miksa, Simon Oblasser, and Andreas Rauber. 2021. Automating Research Data Management Using Machine-actionable Data Management Plans . *ACM Trans. Manag. Inform. Syst.* 1, 1, Article 1 (January 2021), 22 pages. https://doi.org/10.1145/3490396

1 INTRODUCTION

The data revolution continues to transform every sector of science, industry, and government [AS19]. The economic and societal benefits and increased effectiveness of research funding by ensuring that data generated and (pre-) processed as part of research remains available for re-use,

Authors' addresses: Tomasz Miksa, tmiksa@sba-research.org, TU Wien & SBA Research, Vienna, Austria; Simon Oblasser, simon.oblasser@student.tuwien.ac.at, TU Wien, Vienna, Austria; Andreas Rauber, rauber@ifs.tuwien.ac.at, TU Wien, Vienna, Austria.

2158-656X/2021/1-ART1
https://doi.org/10.1145/3490396

Contains:

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

ACM Transactions on Management Information Systems

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