Maintenance and Adoptions of the Common Standard for machine-actionable DMPs

DMP Common Standards WG

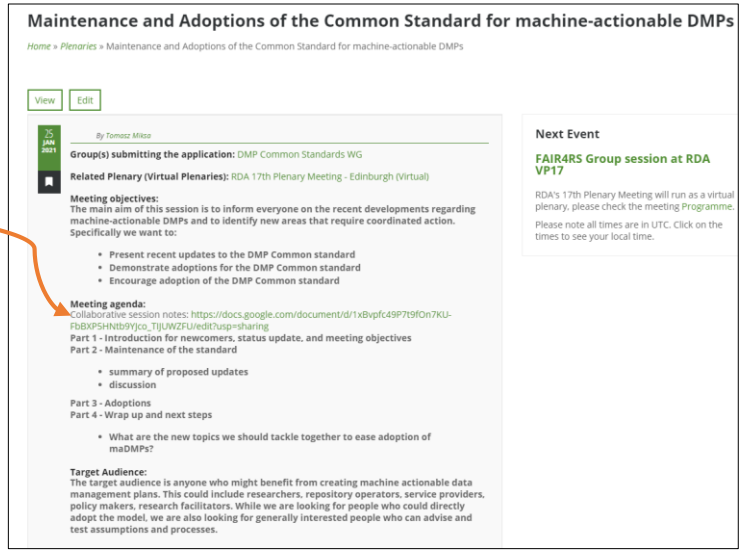
Tomasz Miksa, Paul Walk, Peter Neish

Agenda

- Part 1 - Introduction for newcomers, status update, and meeting objectives (max. 10 min)
- Part 2 – Community updates and adoptions (max 30 min.)
  - Quick News
  - Zeno Casellato – New tool for machine-actionable DMPs implemented by FAIR Data Austria
  - Björn Brötz - Automated checks and metadata annotation based on maDMPs for the early data life cycle
  - Elli Papadopoulu – Argos DMP outputs and integrations
  - Fajar Ekaputra – Recent developments in ontological representation of machine-actionable DMPs
- Part 3 – Maintenance of the standard (max 40 min.)
- Part 4 - Wrap up and next steps (max 10 min.)
  - What are the new topics we should tackle together to ease adoption of maDMPs?

Participants

- Collaborative notes
  - LINK
  - Add your name to the list
  - Co-edit!
- Let's get to know each other



Maintenance and Adoptions of the Common Standard for machine-actionable DMPs

Home » Plenaries » Maintenance and Adoptions of the Common Standard for machine-actionable DMPs

View | Edit

By Tomasz Miksa

Group(s) submitting the application: DMP Common Standards WG

Related Plenary (Virtual Plenaries): RDA 17th Plenary Meeting - Edinburgh (Virtual)

Meeting objectives:
The main aim of this session is to inform everyone on the recent developments regarding machine-actionable DMPs and to identify new areas that require coordinated action.
Specifically we want to:

- Present recent updates to the DMP Common standard
- Demonstrate adoptions for the DMP Common standard
- Encourage adoption of the DMP Common standard

Meeting agenda:
Collaborative session notes: https://docs.google.com/document/d/1xBvpfc49P7t9fOn7KU-FbBXPSHNtb9YJco_TJUWZFU/edit?usp=sharing
Part 1 - Introduction for newcomers, status update, and meeting objectives
Part 2 - Maintenance of the standard

- summary of proposed updates
- discussion

Part 3 - Adoptions
Part 4 - Wrap up and next steps

- What are the new topics we should tackle together to ease adoption of maDMPs?

Target Audience:
The target audience is anyone who might benefit from creating machine actionable data management plans. This could include researchers, repository operators, service providers, policy makers, research facilitators. While we are looking for people who could directly adopt the model, we are also looking for generally interested people who can advise and test assumptions and processes.

Next Event

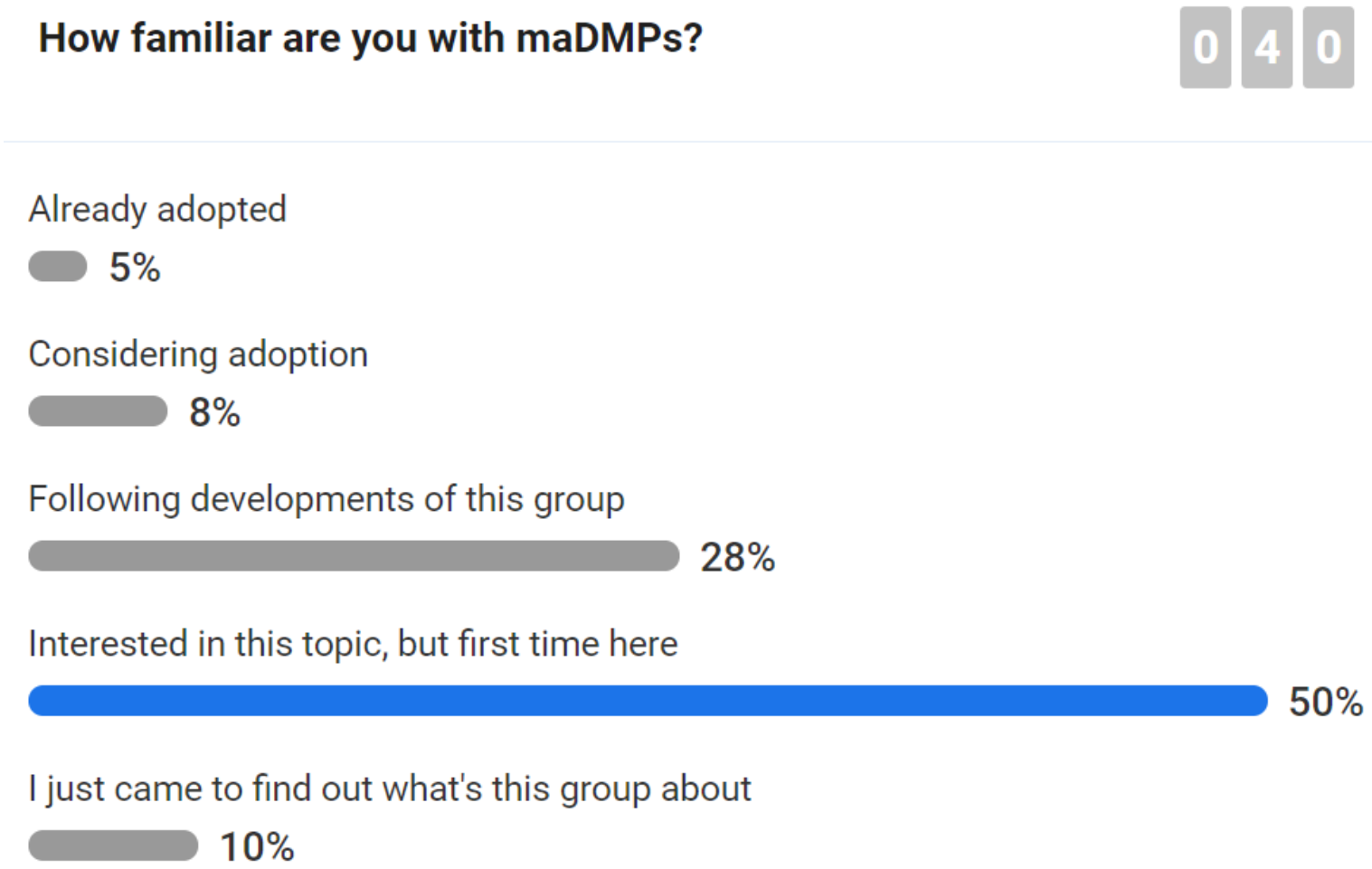
FAIR4RS Group session at RDA VP17

RDA's 17th Plenary Meeting will run as a virtual plenary, please check the meeting Programme.

Please note all times are in UTC. Click on the times to see your local time.

Join at
slido.com
#926 519

How familiar are you with maDMPs?



What is your background?

Where are you located?

Introduction for newcomers

Part 1

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?*

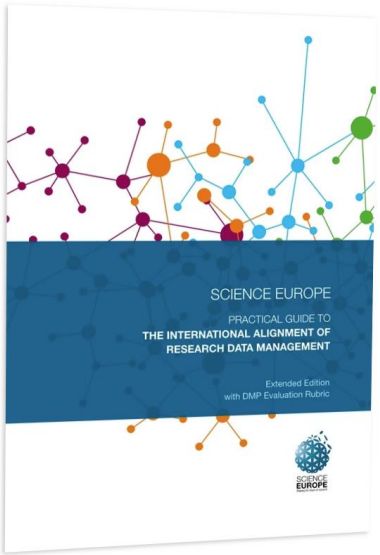
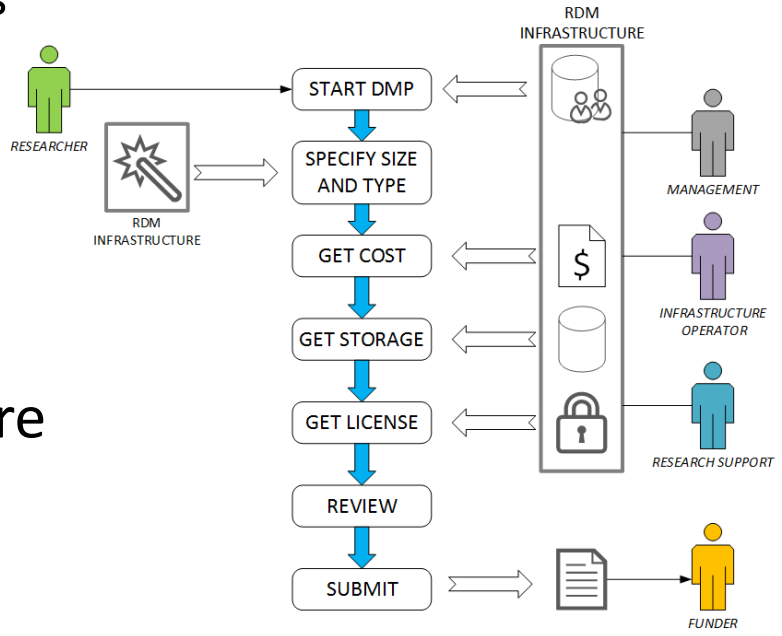


Table of Contents

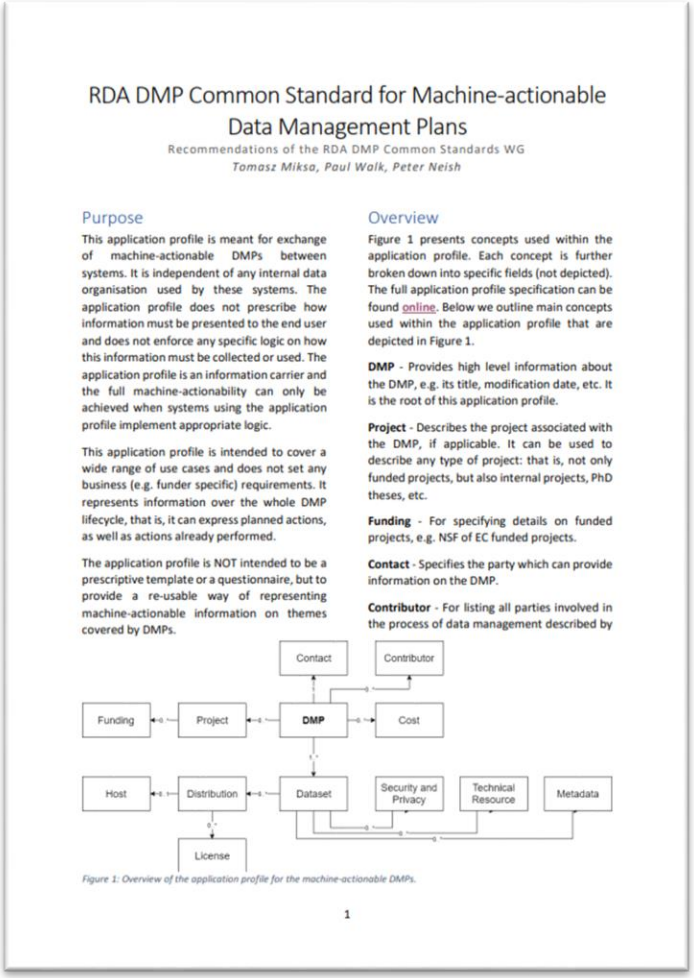
4 For procedural elements of implementing DMPs, see the RDA DMP Common Standards Working Group: https://www.rd-alliance.org/groups/dmp-common-standards-wg

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



Official RDA Recommendation on maDMPs



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

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	

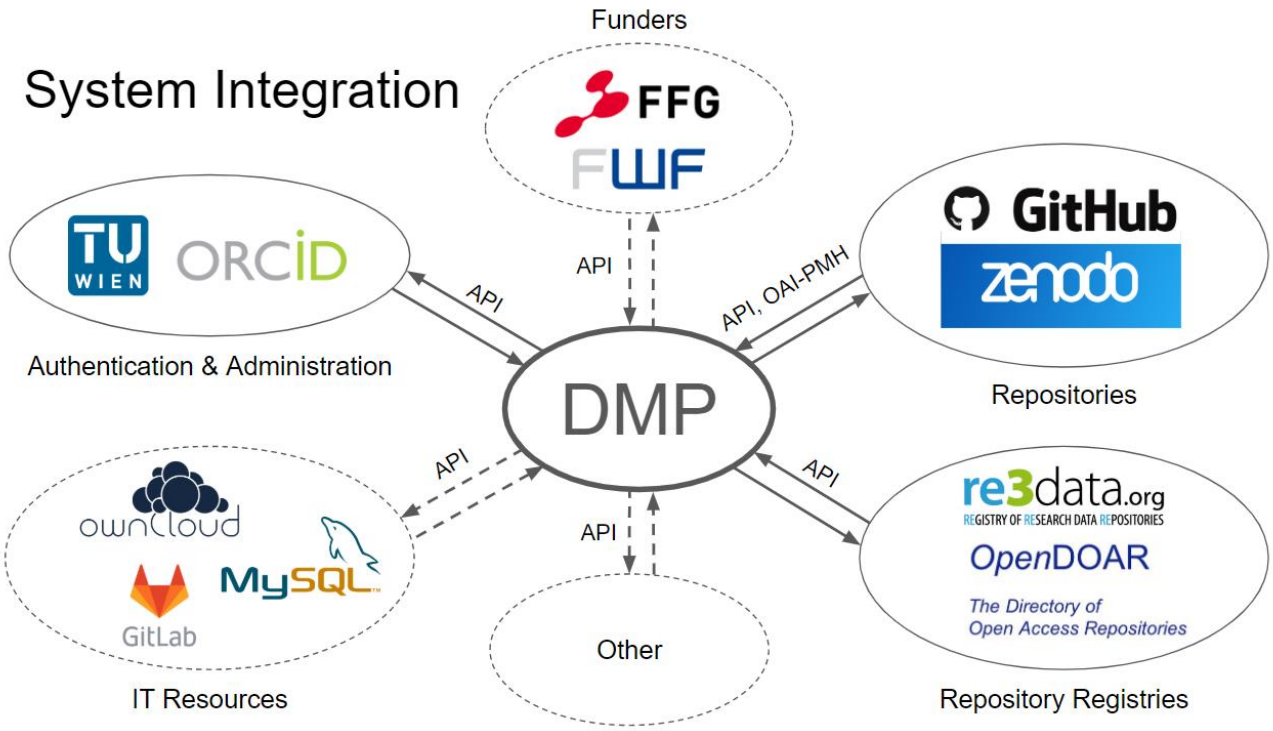
NOT a questionnaire!
NOT a template!

Most fields are optional!

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

RDM Infrastructure

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

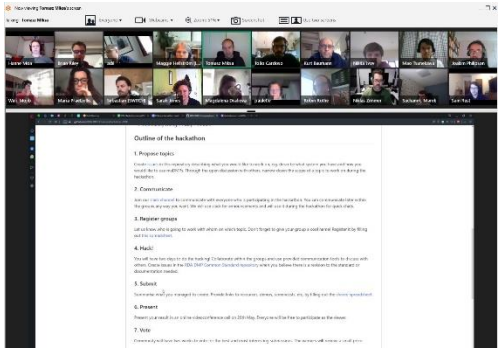


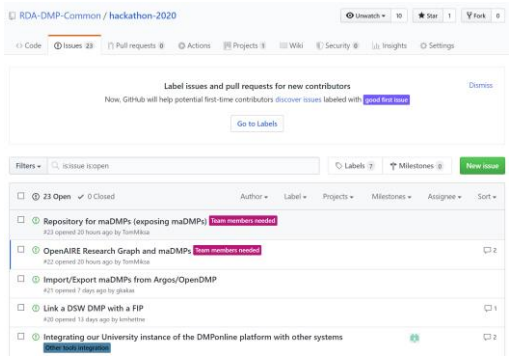
Adoptions (selected)

DMP Common Standards WG

- Activities
  - Calls announced via mailing list
  - GitHub issues
- RDA Hackathon on maDMPs
  - 71 participants, 12 teams, 21 countries
  - Integrations, mappings, etc.
  - Results

We have **230+** members!



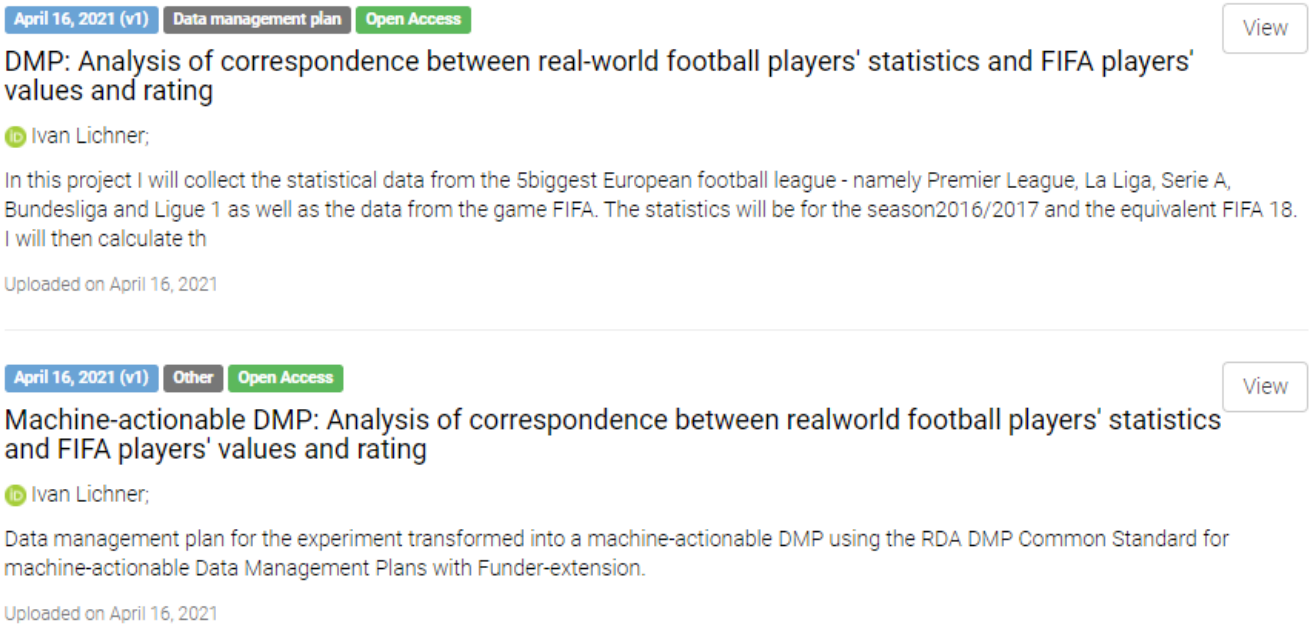


Community updates and adoptions

Part 2

News

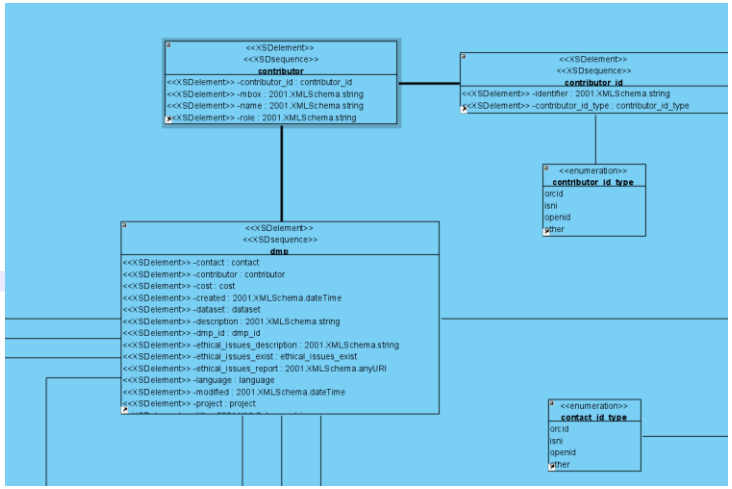
- New examples of maDMPs and DMPs
  - Created by Data Stewardship students at TU Wien
  - https://zenodo.org/communities/dast-2021/



News

- maDMPs in XML
  - CNRS in France
  - XML and XSD Schema
  - Visual paradigm visualization
- Will be released when ready

```
<element name="dataset">
  <complexType>
    <sequence>
      <element minOccurs="0" name="data_quality_assurance">
        <annotation>
          <documentation>Data Quality Assurance</documentation>
        </annotation>
        <complexType>
          <sequence>
            <element maxOccurs="unbounded" minOccurs="0" name="item" type="string"/>
          </sequence>
        </complexType>
      </element>
      <element name="dataset_id">
        <annotation>
          <documentation>Dataset ID</documentation>
        </annotation>
        <complexType>
          <sequence>
            <element name="identifier" type="string">
              <annotation>
                <documentation>Identifier for a dataset, examples:["https://hdl.handle.net/11353/10.923628"]</documentation>
              </annotation>
            </element>
```

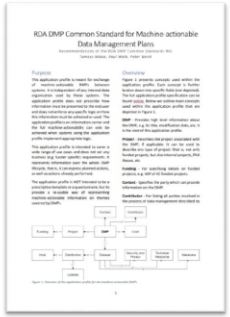


News

- maDMPs at Science Europe
  - Invited talk to present maDMPs
  - Positive feedback
    - Questions on specific tools/software

Conclusion

- Common Standard for machine-actionable DMPs
  - Allows to exchange information between systems
  - Includes all stakeholders involved in RDM lifecycle
  - Can increase quality of DMPs
    - e.g. automate DMP validation
  - Can reduce workload
    - e.g. get estimations and help from other stakeholders
  - Is not a competing template
- Science Europe template can be made machine-actionable
- Your feedback and potential endorsement are the key drivers for change!



12/01/2021 www.rd-alliance.org - @resdatall CC BY-SA 4.0

SCIENCE EUROPE

News

- maDMPs and data repositories
  - Webinar: https://www.youtube.com/watch?v=kY0LCmR6FJQ

Texas Digital Library

search here ... Go

HOME ABOUT EVENTS COMMUNITY SERVICES SUPPORT

Home › Events › Webinars @ TDL › Texas Data Repository Webinars

Texas Data Repository Webinars

TOPICS IN
RESEARCH DATA MANAGEMENT
FREE WEBINAR SERIES HOSTED BY THE TEXAS DATA REPOSITORY STEERING COMMITTEE

Texas Digital Library and the Texas Data Repository Steering Committee present a webinar series showcasing topics in research data management and hosted by research data experts in academic libraries throughout North America.

All the webinars in this series are free and open to anyone who is interested.

Featuring presentations from:

The Odum Institute | University of Virginia Libraries | MIT Libraries | University of Toronto Libraries | Carnegie Mellon University Libraries | Harvard Dataverse Project | University of British Columbia Library | Data Curation Network| California Digital Library| Qualitative Data Repository

The series, rebooted Fall 2019, will feature the following webinars with live Q&A following each presentation

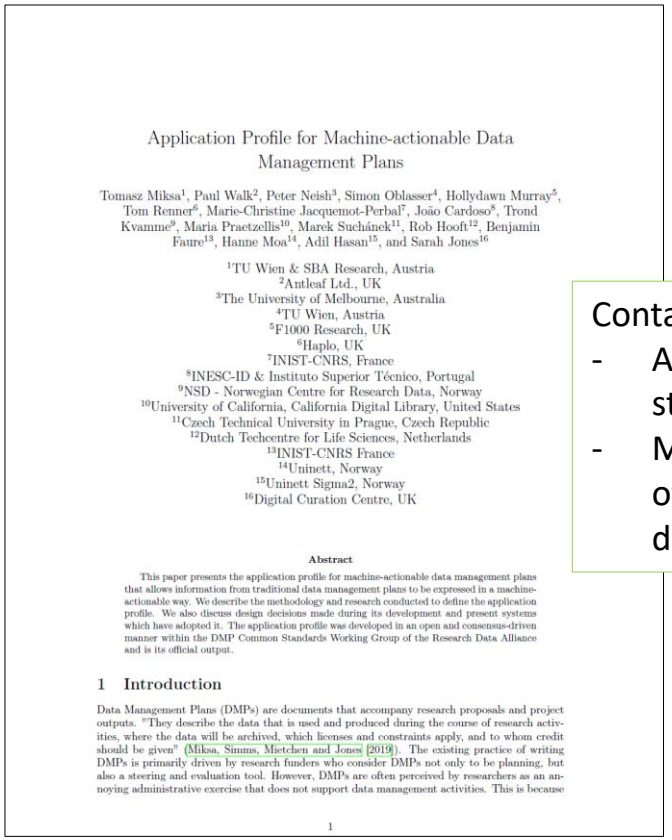
January 21, 2021. 1-2PM, *External Tools for Dataverse* with Tomasz Miksa, Senior Researcher at SBA Research.

Register here.

Past presentations:

January 21, 2021. 1-2PM, *External Tools for Dataverse* with Tomasz Miksa, Senior Researcher at SBA Research. **View slides** | recording

Papers in review…



Application Profile for Machine-actionable Data Management Plans

Tomasz Miksa[1], Paul Walk[2], Peter Neish[3], Simon Oblasser[4], Hollydawn Murray[5], Tom Renner[6], Marie-Christine Jacquemot-Perbal[7], João Cardoso[8], Trond Kvamme[9], Maria Praetzellis[10], Marek Suchánek[11], Rob Hooft[12], Benjamin Faure[13], Hanne Moa[14], Adil Hasan[15], and Sarah Jones[16]

[1]TU Wien & SBA Research, Austria
[2]Antleaf Ltd., UK
[3]The University of Melbourne, Australia
[4]TU Wien, Austria
[5]F1000 Research, UK
[6]Haplo, UK
[7]INIST-CNRS, France
[8]INESC-ID & Instituto Superior Técnico, Portugal
[9]NSD - Norwegian Centre for Research Data, Norway
[10]University of California, California Digital Library, United States
[11]Czech Technical University in Prague, Czech Republic
[12]Dutch Techcentre for Life Sciences, Netherlands
[13]INIST-CNRS France
[14]Uninett, Norway
[15]Uninett Sigma2, Norway
[16]Digital Curation Centre, UK

Abstract

This paper presents the application profile for machine-actionable data management plans that allows information from traditional data management plans to be expressed in a machine-actionable way. We describe the methodology and research conducted to define the application profile. We also discuss design decisions made during its development and present systems which have adopted it. The application profile was developed in an open and consensus-driven manner within the DMP Common Standards Working Group of the Research Data Alliance and is its official output.

1 Introduction

Data Management Plans (DMPs) are documents that accompany research proposals and project outputs. "They describe the data that is used and produced during the course of research activities, where the data will be archived, which licenses and constraints apply, and to whom credit should be given" (Miksa, Simms, Mietchen and Jones, 2019). The existing practice of writing DMPs is primarily driven by research funders who consider DMPs not only to be planning, but also a steering and evaluation tool. However, DMPs are often perceived by researchers as an annoying administrative exercise that does not support data management activities. This is because

1

Contains:
- Adoption stories
- Methodology of standard development

Practice Paper

Interconnecting systems using machine-actionable Data Management Plans - hackathon report

João Cardoso[1], Leyla J. García[2], Tomasz Miksa[3]
[1]Universidade de Lisboa, Instituto Superior Técnico & INESC-ID, Lisboa, Portugal
[2]ZB MED Information Centre for Life Sciences, Cologne, Germany
[3]SBA Research & TU Wien, Vienna, Austria

Corresponding author, joao.m.f.cardoso@tecnico.ulisboa.pt

This paper presents outputs of the Research Data Alliance Hackathon on Machine-actionable Data Management Plans, where participants proposed a series of topics looking forward to using and improving aspects related to this subject. The hackathon served three main purposes: broadening the community, improving the core supporting machine-actionable plans and exposing a growing endorsement on the adoption of the RDA DMP Common Standard application profile in a wide range of settings to enable exchange of DMP specific information in a machine-actionable way.

Keywords: Data management plans, machine-actionable data management plans, semantic web, community practice, open science.

1 Introduction

The Data Management Plan (DMP) was introduced to document and publish both data management practices and policies that are applied to data throughout its lifecycle. This implies describing the techniques, methods and policies on how data is to be created, collected, documented, processed, accessed, preserved, disseminated as well as the roles and responsibilities of associated actors (Michener, 2015).

The premise behind the concept of a machine-actionable DMP (maDMP) is that information contained within a DMP can be enacted both by humans and automated systems, thus addressing some of the limitations associated with traditional DMP documents. To that effect, data management workflows should integrate maDMPs and data management policies should take into account not only human agents but also machines. maDMPs should support both human and machine-processable representations so they act as an interchange format for dissemination and public access of the maDMP (Simms et al., 2017). In order to provide a machine-actionable representation of a maDMP, it becomes necessary to establish a standardised representation of the maDMP. The Research Data Alliance (RDA) (RDA, 2020) DMP Common Standards (DCS) working group (Miksa, Cardoso, and Borbinha, 2018; Miksa, Neish, et al., 2018; Miksa, Walk, and Neish, 2019) developed an application profile making it easier to express information from traditional DMP documents in a machine-actionable way. The DCS maDMP application profile allows for automatic exchange, integration, and validation of information provided in DMP documents. Thus, facilitating the exchange of information between systems acting on behalf of stakeholders involved in the research life cycle, such as researchers, funding bodies, repository managers, ICT providers, librarians, etc.

This paper reports on a hackathon organised by the DCS working group, which had as main motivation to promote the adoption of the maDMP concept by the research community, and, in particular, the usage of the DCS application profile for interchange of maDMPs. To that effect four main areas were identified: (1) serialisation, to encourage community development of serialisations of the DCS application profile; (2)

1

Contains:
- Summary of results from hackathon

DATA SCIENCE JOURNAL

ID	Days Since Submission	Sec	Authors	Title	Status
1242	280	RES	Miksa, Walk, Neish, Oblasser, Jones,...	Application Profile for Machine-actionable Data...	In Review

Part 2 – Community updates and adoptions

- **Zeno Casellato** – New tool for machine-actionable DMPs implemented by FAIR Data Austria
- **Björn Brötz** - Automated checks and metadata annotation based on maDMPs for the early data life cycle
- **Elli Papadopoulu** – Argos DMP outputs and integrations
- **Fajar Ekaputra** – Recent developments in ontological representation of machine-actionable DMPs

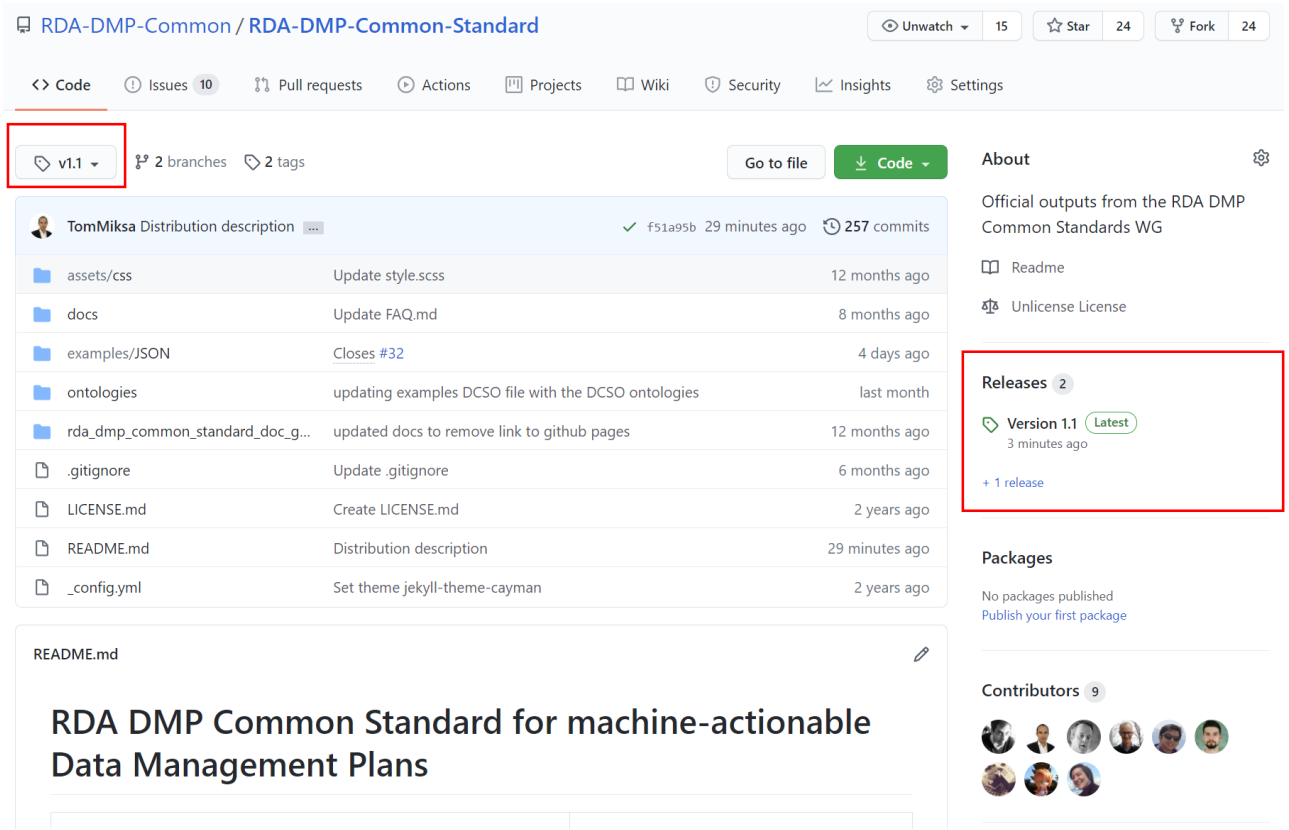
1-2 questions after each presentation

Max 30 minutes

Maintenance of the recommendation

Part 3

Part 3 - Maintenance



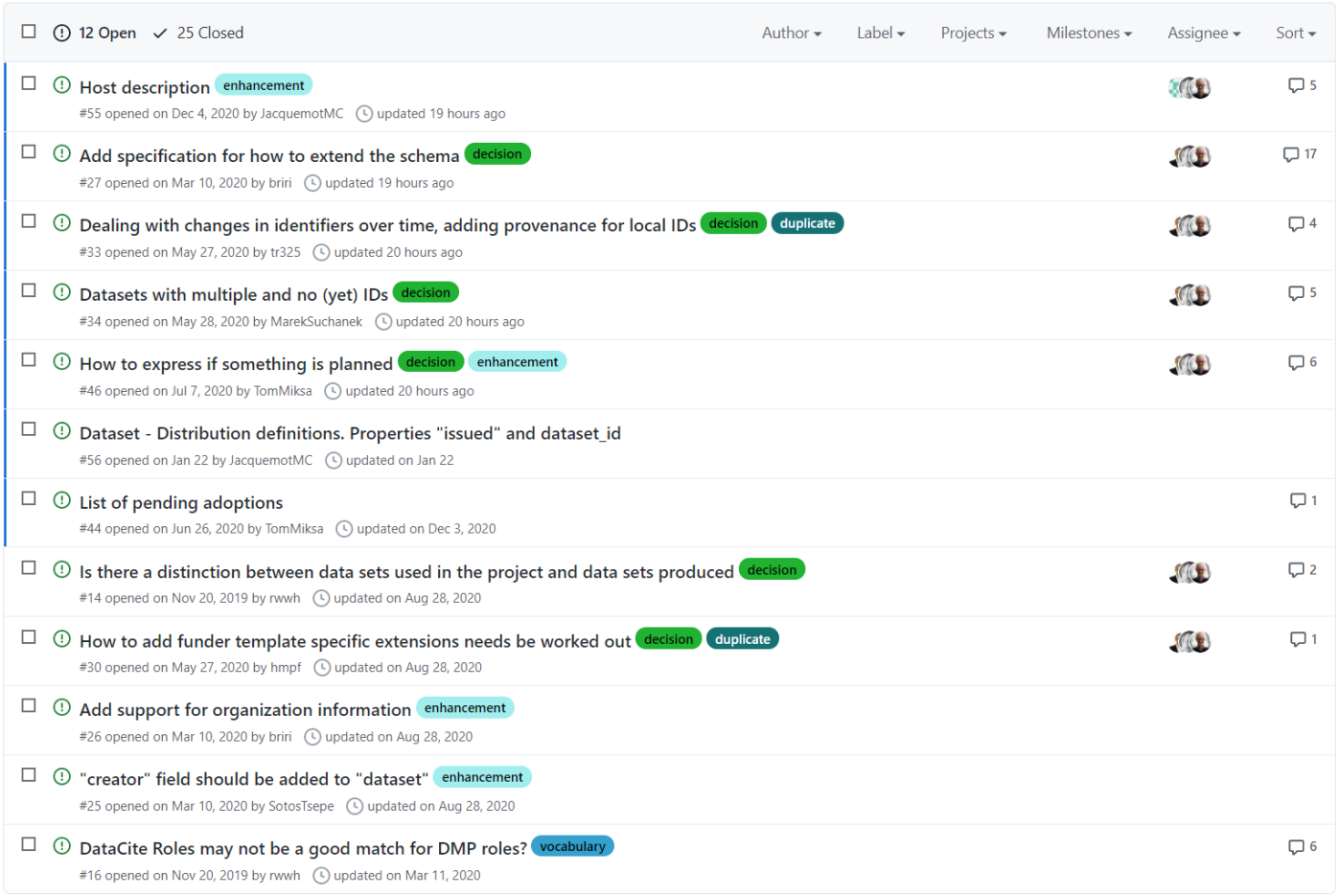
Maintenance process

1. Everyone can create issues
   - Bugs
   - Lack of clarity
   - Changes needed
   - Extensions needed
   - Etc.
2. Community discusses each issue openly
   - In GitHub under specific issue
   - In video calls if needed
3. Chairs of the DMP Common WG
   - Review and label issues
   - Identify needs for new releases

Maintenance – release types

- PATCH version to make **backwards compatible bug fixes**
  - E.g. spelling mistakes, etc.
  - Immediate effect
  - 1.1 can change to 1.1.1
- MINOR version to **add functionality in a backwards compatible manner**
  - E.g. relaxing constraints
  - Grouped and announced at plenaries
  - 1.1. can change to 1.2
- MAJOR version when we make **incompatible changes**
  - E.g. remove entities
  - Broadly discussed, announced at plenaries
  - 1.1 can change to 2.0

Open issues – possible next release?



https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard/issues?q=is%3Aopen+is%3Aissue

Open issues and new releases

- Likely not enough time today to talk in detail
- Please comment on GH
- We may organise extra calls if needed

We had dedicated calls since the last plenary.
Meeting minutes:

- 01/2021
- 02/2021
- 04/2021

We discussed issues and funder extensions.

Top 3 open issues – discussion needed

1. How to indicate that extensions are used?
   - Needed by some tools that want to transfer more information
   - Can be a mechanism to incorporate funder specific needs.
   - **Recent suggestion: use ontology representation**
2. Changes in identifiers over time and their cardinality
   - Sometimes datasets have a DOI and a handle
   - Before a dataset gets a DOI, it may have internal system ID
   - W3C DCAT however assumes only one identifier
   - **Recent suggestion: use Distributions!**
3. Making things explicit (planned, performed)
   - Currently we use dates to indicate this -> some arithmetic needed
   - **Recent suggestion: mark things that are reused**

How to indicate that extensions are used

- Add *extensions* field to *dmp*
  - To list all extensions
  - Must provide a link to the schema / specification
- Example

```
{
  "dmp": {
     ...
    "extensions": [
      {"uri": "http://university/citations", name: "university-citations"},
      {"uri": "http://school/citations", name:"school-citations"},
    ],
    ...
    "dataset": [
      {
        "title": "My Dataset",
        "school-citations-specific-field": "generated by DMPTool",
        "university-citations-specific-field": "something else",
       ....
      }
    ]
  }
}
```

Changes in identifiers over time and their cardinality

- Sometimes datasets have a DOI and a handle
  - Argument for changing cardinality
- Before a dataset gets a DOI, it may have internal system ID
  - Could be modelled as two different datasets
  - Sharing internal IDs may cause security threats?
- W3C DCAT however assumes only one identifier
  - Breaking compliance with the standard is rather bad practice
- Solution (as original planned): use multiple distributions for one dataset

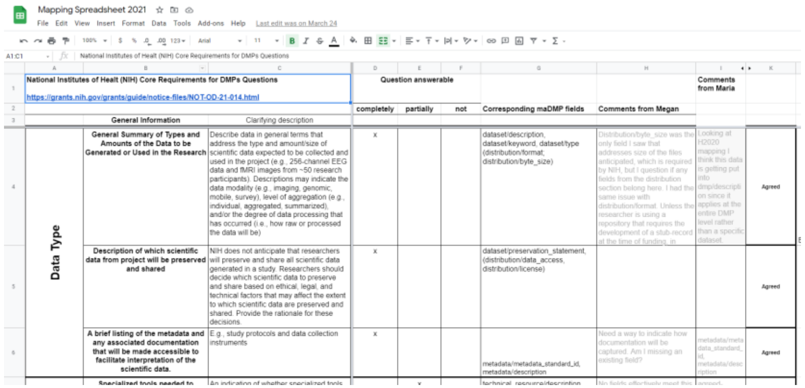
Making things explicit (planned, performed)

- We use dates to indicate planned actions.
- DMP has a **modification timestamp** that contains a timestamp of the last DMP modification. Dataset contains **issue date** that indicates whether the actions are planned or already performed:
  - if the issue date is set in the future (compared to DMP modification date), then the actions are planned,
  - if the issue date is set in the past (compared to DMP modification date), the actions were performed.
- Comment from rwwh:
  - "if a trivial update is performed by a tool, and the modification date is changed, this can put it beyond the planned date for some action. **Unless planned actions are indicated in an explicit way, this can be a problem**! Every tool must now check that they don't inadvertently move the dmp modification date past any other date in the DMP!“

What’s your opinion on this?

Mapping to funder templates

- Based on outputs from the hackathon
- [Methodology]
- Templates analysed
  - EC Horizon 2020
  - Science Europe
  - NIH DMP for 2023
  - NSF Generic Template
  - US Geological Survey
- Based on this analysis we defined
  - Proposal for changes in the standard
  - Proposal for *funder-extension*


Mapping Spreadsheet 2021

National Institutes of Healt (NIH) Core Requirements for DMPs Questions https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-014.html			Question answerable					Comments from Maria	
			completely	partially	not	Corresponding maDMP fields	Comments from Megan		
	General Information	Clarifying description							
Data Type	General Summary of Types and Amounts of the Data to be Generated or Used in the Research	Describe data in general terms that address the type and amount/size of scientific data expected to be collected and used in the project (e.g., 256-channel EEG data and fMRI images from ~50 research participants). Descriptions may indicate the data modality (e.g., imaging, genomic, mobile, survey), level of aggregation (e.g., individual, aggregated, summarized), and/or the degree of data processing that has occurred (i.e., how raw or processed the data will be)	x			dataset/description, dataset/keyword, dataset/type (distribution/format; distribution/byte_size)	Distribution/byte_size was the only field I saw that addresses size of the files anticipated, which is required by NIH, but I question if any fields from the distribution section belong here. I had the same issue with distribution/format. Unless the researcher is using a repository that requires the development of a stub-record at the time of funding, in	Looking at H2020 mapping I think this data is getting put into dmp/description since it applies at the entire DMP level rather than a specific dataset.	Agreed
	Description of which scientific data from project will be preserved and shared	NIH does not anticipate that researchers will preserve and share all scientific data generated in a study. Researchers should decide which scientific data to preserve and share based on ethical, legal, and technical factors that may affect the extent to which scientific data are preserved and shared. Provide the rationale for these decisions.	x			dataset/preservation_statement, (distribution/data_access, distribution/license)			Agreed
	A brief listing of the metadata and any associated documentation that will be made accessible to facilitate interpretation of the scientific data.	E.g., study protocols and data collection instruments	x			metadata/metadata_standard_id, metadata/description	Need a way to indicate how documentation will be captured. Am I missing an existing field?	metadata/metadata_standard_id, metadata/description	Agreed

Many thanks to:
Megan Potterbusch, Maria Praetzellis, Paulette Lieby

Proposal for changes in the documentation of the common standard

Field	Current Description	Proposed Description
Contributor	For listing all parties involved in the process of the data management described by this DMP, and those parties involved in the creation and management of the DMP itself.	For listing all parties involved in the process of the data management described by this DMP, and those parties involved in the creation and management of the DMP itself. Parties may be persons and/or institutions.
Metadata	Provides a pointer to a metadata standard used to describe the data. It does not contain any actual metadata relating to the dataset.	Provides a pointer to a metadata standard or to any additional documentation or other that describe the data. Such a documentation could be a README file detailing the naming convention, a pointer to controlled vocabularies, ...
DataQualityAssurance	Describes in free text how data quality is achieved. It cites "naming convention" as an example.	Description of methods and/or pointers to artifacts used to ensure data quality.
SecurityAndPrivacy	Used to indicate any specific requirements related to security and privacy of a specific dataset, e.g. to indicate that data is not anonymized.	Used to indicate any specific requirements related to security and privacy of a specific dataset, e.g. the description of measures to be taken when dealing with sensitive and/or personal data.
Technical Resource		Include software and code (NIH discussion)

https://docs.google.com/document/d/1MiJZnsVJVyN-C_f5iiv4S-u_hmf8mk2_MBe7eI2pg0E/edit?usp=sharing

Proposal for changes in the structure of the common standard

Replace **dataset/quality_assurance** that is a String with a nested data structure and to change its cardinality to 0..n. "Should be focused on calibration, ensuring accuracy and reliability of data. Naming conventions fit better under metadata."

Direct change of properties in 'dataset'

Name	Description	Data Type	Cardinality	Example Value
data_quality_assurance	Provide any information on the measures taken during the research process to ensure the data quality.	Nested Data Structure	0..n	

Properties in 'data_quality_assurance'

Name	Description	Data Type	Cardinality	Example Value
description	Free text to describe a method used in the data quality process.	String	1	We use certification/calibration/metrology/management of signal to noise ratio, standard operating procedures,...
data_quality_assurance_id	Identifier for a Data Quality Assurance artefact	Nested Data Structure	0..1	

https://docs.google.com/document/d/1MiJZnsVJVyN-C_f5iiv4S-u_hmf8mk2_MBe7eI2pg0E/edit?usp=sharing

Proposal for *funder-extension*

Funder-extension properties in 'dmp'

Name	Description	Data Type	Cardinality	Example Value
related_policy	To link to all documents needed to be compliant to requirements within this DMP (e.g. legal, ethical, contractual, guidelines, procedures, standards,...)	Nested Data Structure	0..n	

Funder-extension properties in 'related_policy'

Name	Description	Data Type	Cardinality	Example Value
description	Description	String	0..1	RDM policy of TU Wien.
related_policy_id	Related policy ID	Nested Data Structure	0..1	

https://docs.google.com/document/d/1rLV-_IucmONMLUvEhkxQqv3qTORRFKQtS-mHpq1l2aQ/edit?usp=sharing

Proposal for *funder-extension*

Should we include them directly in the recommendation?

Funder-extension properties in 'dataset'

Name	Description	Data Type	Cardinality	Example Value
is_reused	To explicitly indicate whether the dataset is reused or was produced in the course of research. Allowed values are: reused, produced.	Term from Controlled Vocabulary	1	reused
target_audience	To state for whom this dataset can be relevant.	String	0..n	This dataset is of special interest to ethnomusicologists working on...
methodology	To describe methodology, procedures, workflows. etc. on how the dataset is created, can be recovered, ...	Nested Data Structure	0..n	This data is a result of simulation made using...

https://docs.google.com/document/d/1rLV-_IucmONMLUvEhkxQqv3qTORRFKQtS-mHpq1l2aQ/edit?usp=sharing

Proposal for *funder-extension*

Funder-extension properties in 'methodology'

Name	Description	Data Type	Cardinality	Example Value
description	Methodology, procedures, workflows, etc.	String	0..1	Diagram X explaining the data collection process.
methodology_id	Identifier for a methodology artefact	Nested Data Structure	0..1	

Funder-extension properties in 'methodology_id'

Name	Description	Data Type	Cardinality	Example Value
identifier	A unique identifier for a methodology artefact	String	1	https//thejmfc.org/.methodologies/xpto1.pdf
type	Identifier type	String	1	URL

Funder-extension to properties in 'distribution'

Name	Description	Data Type	Cardinality	Example Value
restriction_explanation	To describe any reasons why data cannot be shared openly. How to handle IPR? (Science Europe 30)	String	0..n	The data will be kept closed, because I am selfish and also the policy of my institution doesn't allow me to share it.

Restriction explanation
could be part of recommendation?

https://docs.google.com/document/d/1rLV-_IucmONMLUvEhkxQqv3qTORRFKQtS-mHpq1l2aQ/edit?usp=sharing

Wrap-up and next steps

Part 4

Next steps

- DMP Common Standards WG continues to maintain the recommendation
  - Updates to the specification if needed
  - Supports in adoption
  - Promotes success stories / lessons learned

- Active DMPs IG
  - Place for discussion on all topics related to DMPs
    - e.g. What are the new topics we should tackle together to ease adoption of maDMPs?
  - Join the session today!

- Ideas? Comments? Anything you would like to share?

maDMPs – summary

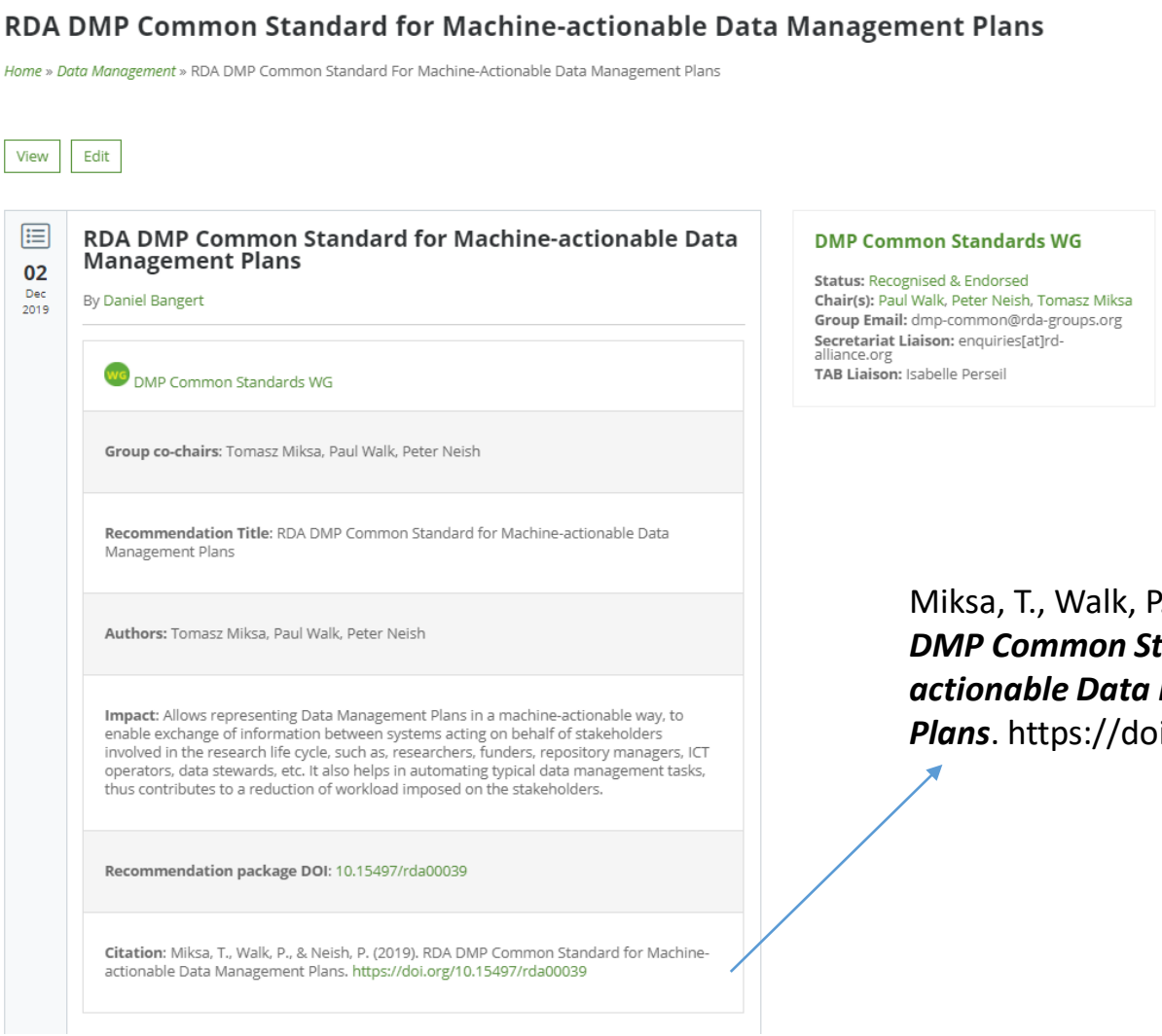
- Recommendation
  - https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard
  - http://doi.org/10.15497/rda00039
- Participate in recommendation adoption!
- Contact group chairs
  - Questions
  - Ideas
  - Success stories

Tomasz Miksa

Paul Walk

Peter Neish

Please cite the recommendation



RDA DMP Common Standard for Machine-actionable Data Management Plans

Home » Data Management » RDA DMP Common Standard For Machine-Actionable Data Management Plans

View Edit

02 Dec 2019

RDA DMP Common Standard for Machine-actionable Data Management Plans

By Daniel Bangert

DMP Common Standards WG

Group co-chairs: Tomasz Miksa, Paul Walk, Peter Neish

Recommendation Title: RDA DMP Common Standard for Machine-actionable Data Management Plans

Authors: Tomasz Miksa, Paul Walk, Peter Neish

Impact: Allows representing Data Management Plans in a machine-actionable way, to enable exchange of information between systems acting on behalf of stakeholders involved in the research life cycle, such as, researchers, funders, repository managers, ICT operators, data stewards, etc. It also helps in automating typical data management tasks, thus contributes to a reduction of workload imposed on the stakeholders.

Recommendation package DOI: 10.15497/rda00039

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

DMP Common Standards WG

Status: Recognised & Endorsed
Chair(s): Paul Walk, Peter Neish, Tomasz Miksa
Group Email: dmp-common@rda-groups.org
Secretariat Liaison: enquiries[at]rd-alliance.org
TAB Liaison: Isabelle Perseil

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

Publications

- 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