



RESEARCH DATA ALLIANCE

Machine-actionable DMPs - tell us your story!

DMP Common Standards WG

Tomasz Miksa, Paul Walk, Peter Neish

Agenda

- Part 1 - Introduction for newcomers, status update, and meeting objectives
- Part 2 – Existing, pending and planned adoptions
 - **Claire Austin** - *maDMPs in a government context*
 - **Maria Praetzellis** - *DMPHub*
 - **Lucas Berent, Alexander Selzer** - *maDMPs repository*
 - **Max Moser** - *Framework for integrating maDMPs with data repositories*
 - **Georgios Kakaletris** - *Argos adoption of RDA ma-DMP standard and integrations with the OpenAIRE ecosystem*
 - **Maroua Jaoua** - *Research Object Crates and Machine-actionable Data Management Plans*

1-2 questions after each presentation, discussion at the end of this part
- Part 3 - Maintenance of the recommendation
 - summary of proposed updates
 - discussion
- Part 4 - Wrap up and next steps
 - 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!

Machine-actionable DMPs - tell us your story!
Home » Plenaries » Machine-actionable DMPs - tell us your story!

[View](#) [Edit](#)

14th April 2020
By Tomaz Miličič

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

Related Plenary (Virtual Plenaries): RDA 16th Plenary Meeting - Costa Rica (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:

- Demonstrate adoptions for the DMP Common standard
- Encourage adoption of the DMP Common standard
- Promote discussion of the DMP Common standard

Meeting agenda:

Collaborative Notes Link: <https://docs.google.com/document/d/1prj10bsEeehp3dv2msCNWkPg/Cuz8dyVuM...>

Part 1 - Introduction for newcomers, status update, and meeting objectives

Part 2 - Adoptions

- Claire Austin - maDMPs in a government context
- Maria Praetzelis - DMPHub
- Lucie Berent, Alexander Selzer - maDMPs repository
- Max Moser - Framework for integrating maDMPs with data repositories
- Georgios Kakaletis - Argos adoption of RDA ma-DMP standard and integrations with the OpenAIRE ecosystem
- Maroua Jacoua - Research Object Crates and Machine-actionable Data Management Plans
- <your name here>
- (discussion)

Part 3 - Maintenance of the standard

- summary of proposed updates


› Let's get to know each other

Join at
slido.com
#55862




slido

How familiar are you with maDMPs?

 Start presenting to display the poll results on this slide.

slido

Where are you located?

 Start presenting to display the poll results on this slide.

slido

Earth is...

 Start presenting to display the poll results on this slide.

Introduction for newcomers

Part 1

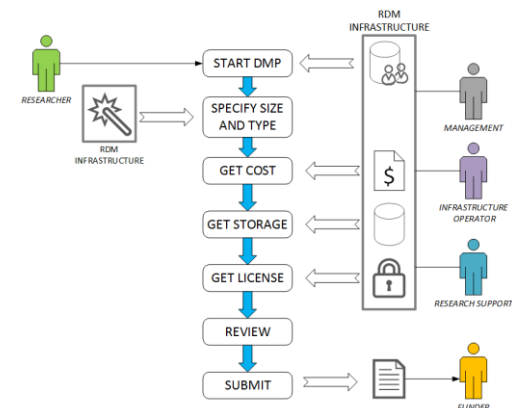
It all started with a big bang! ☺



Why do we need this WG?

- DMPs currently
 - manually completed,
 - vague, not updated, ...
- Machine-actionable DMPs
 - 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

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?



Motivation - example

- Current DMPs – **model questionnaires**

<administrative_data>

<question>Who is the contact person for the DMP?</question>

<answer>The PI will be John Smith from our university.</answer>

</administrative_data>

- Machine-actionable DMPs – **model information**

"contact": {

 "mbox": "JSmith@tuwien.ac.at",

 "name": "John Smith",

 "contact_id": {


 "identifier": "https://orcid.org/0000-0000-0000-0000",

 "type": "orcid"

 }

},


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 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

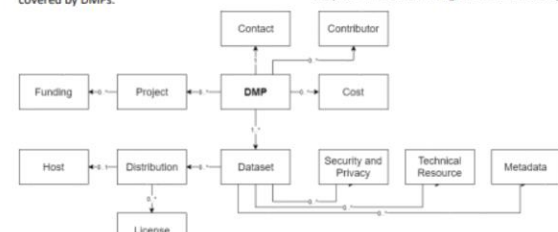


Figure 1: Overview of the application profile for the machine-actionable DMPs.

1

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'

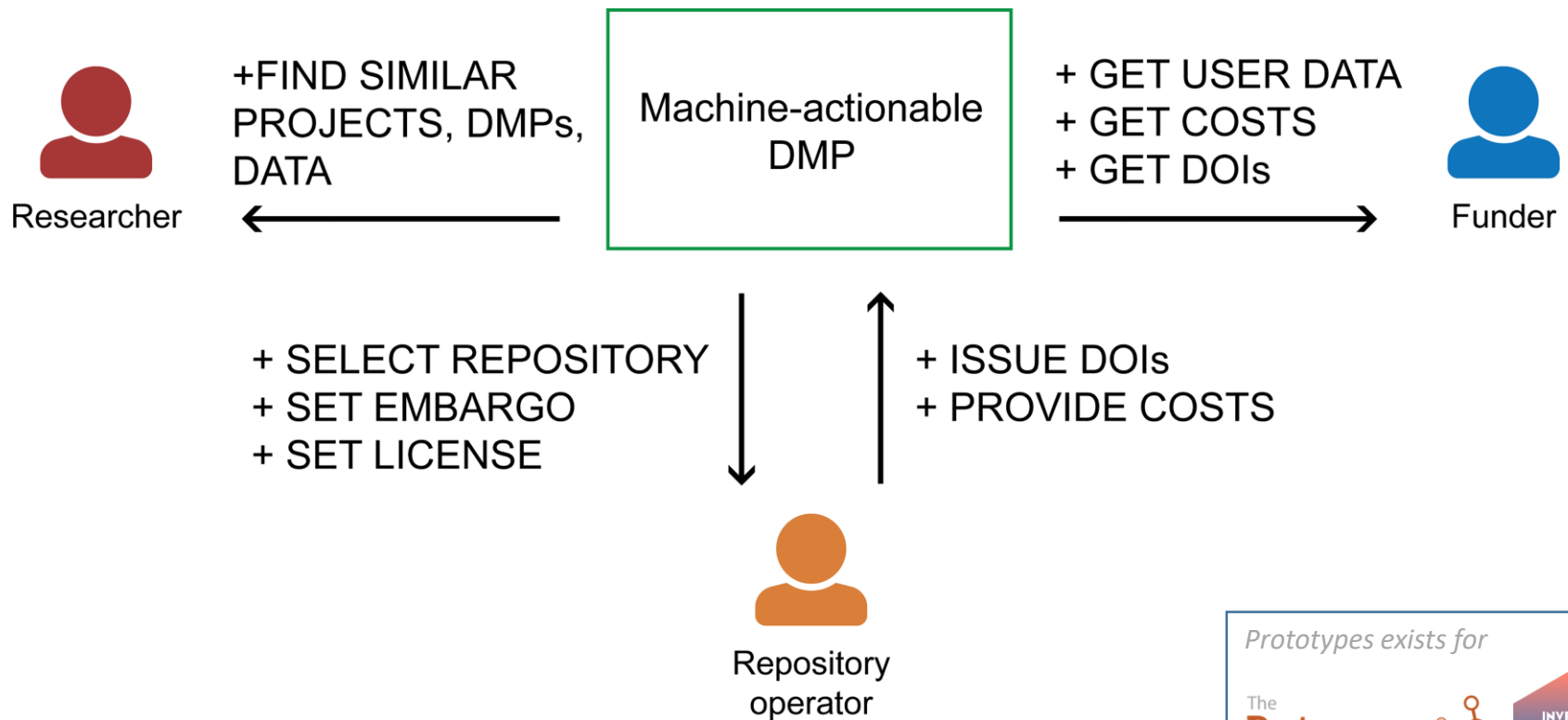
Name	Description	Data Type	Cardinality	Example Value
contact	Contact person for a DMP	Nested Data Structure	1	
contributor	To list people that play role in data management related to this DMP, e.g. responsible for performing actions described in this DMP.	Nested Data Structure	0..n	
cost	To list costs 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>

New opportunities - example



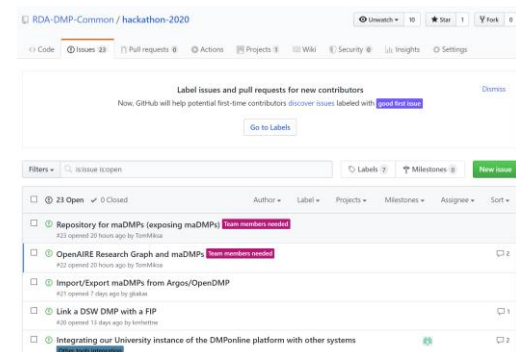
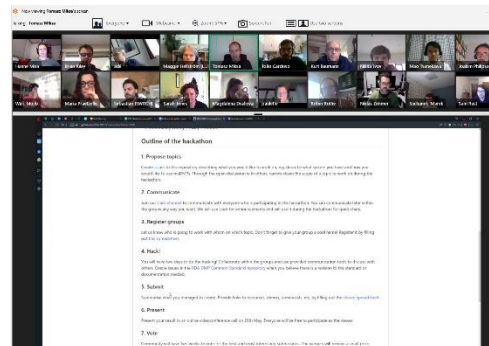
Miksa T, Simms S, Mietchen D, Jones S (2019) Ten principles for machine-actionable data management plans. PLoS Comput Biol 15(3): e1006750. <https://doi.org/10.1371/journal.pcbi.1006750>

Pending adoptions (selected)



RDA Hackathon on maDMPs

- 71 participants, 12 teams, 21 countries
- Integrations, mappings, etc.
- Results
 - Reports, slides
 - Grand finale recording



<https://github.com/RDA-DMP-Common/hackathon-2020>

Existing, pending and planned adoptions

Part 2

Part 2 – Existing, pending and planned adoptions

- **Claire Austin** - *maDMPs in a government context*
- **Maria Praetzellis** - *DMPHub*
- **Lucas Berent, Alexander Selzer** - *maDMPs repository*
- **Max Moser** - *Framework for integrating maDMPs with data repositories*
- **Georgios Kakaletris** - *Argos adoption of RDA ma-DMP standard and integrations with the OpenAIRE ecosystem*
- **Maroua Jaoua** - *Research Object Crates and Machine-actionable Data Management Plans*

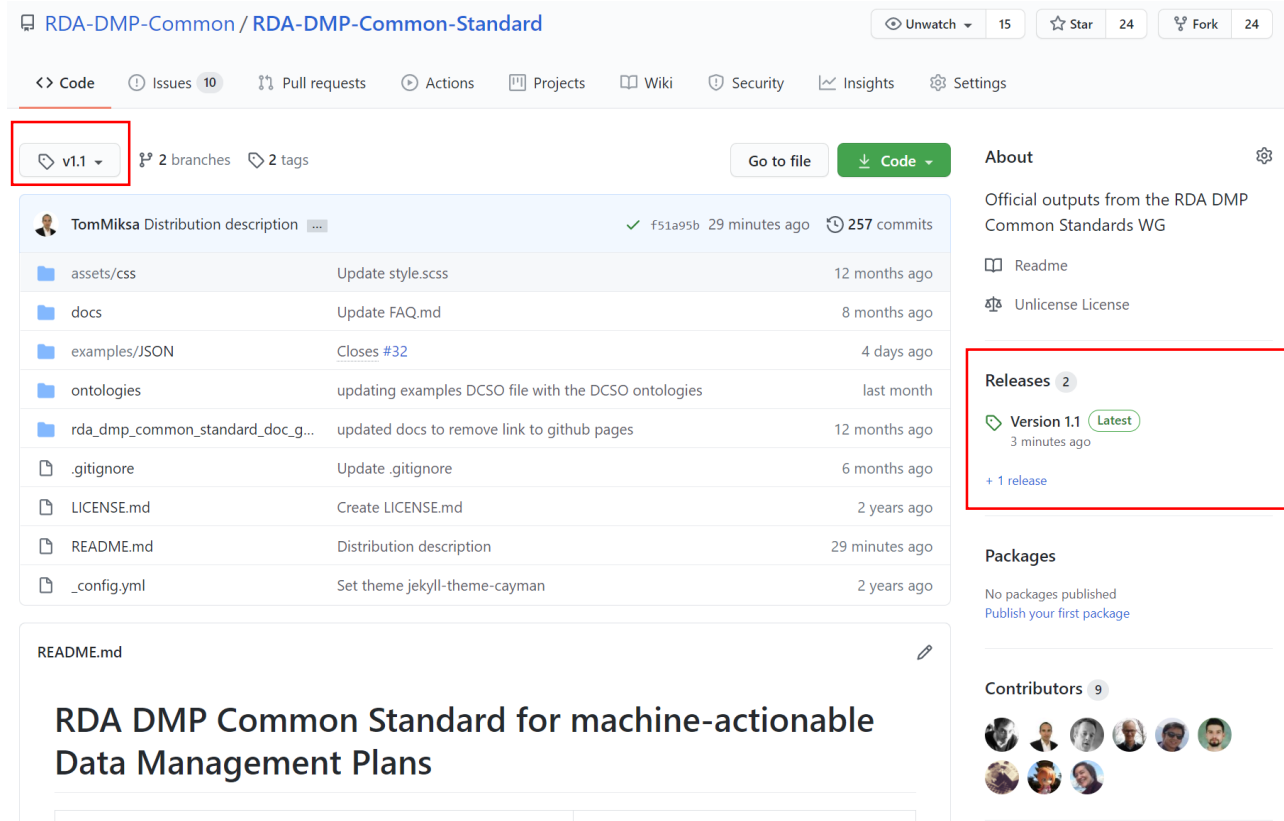
1-2 questions after each presentation, discussion at the end of this part

Maintenance of the recommendation

Part 3

Part 3 - Maintenance

- Release 1.1
- Maintenance process
- Open issues
- Next releases



RDA-DMP-Common / RDA-DMP-Common-Standard

Unwatch 15 Star 24 Fork 24

Code Issues 10 Pull requests Actions Projects Wiki Security Insights Settings

v1.1 2 branches 2 tags

Go to file Code

About

Official outputs from the RDA DMP Common Standards WG

Readme

Unlicense License

Releases 2

Version 1.1 Latest 3 minutes ago + 1 release

Packages

No packages published Publish your first package

Contributors 9

README.md

RDA DMP Common Standard for machine-actionable Data Management Plans

Version 1.1 – summary of changes

1. Relaxed cardinality of following properties:

- `end` (*dmp/project/end*) from required to optional
- `grant_id` (*dmp/project/funding/grant_id*) from required to optional
- `start` (*dmp/project/start*) from required to optional

2. Updated description of following properties by adding: "Encoded using the relevant ISO 8601 Date and Time **compliant string**":

- `available_until` (*dmp/dataset/distribution/available_until*)
- `created` (*dmp/created*)
- `end` (*dmp/project/end*)
- `issued` (*dmp/dataset/issued*)
- `modified` (*dmp//modified*)
- `start` (*dmp/project/start*)
- `start_date` (*dmp/dataset/distribution/license/start_date*)

<https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard/releases/tag/v1.1>

Version 1.1 – JSON Schema

master ▾ RDA-DMP-Common-Standard / examples / JSON / JSON-schema / Go to file Add file ▾

 SotosTsepe Closes #32 ... 4 days ago History

..

 1.0	Add missing required property	3 months ago
 1.1	Closes #32	4 days ago

1. All of the above to match specification.
2. Removed all `minItems` restrictions to arrays.






















Example:

This will now pass validation:

```
"license": []
```

<https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard/tree/master/examples/JSON/JSON-schema/1.1>

Version 1.1 – resolved issues

<input type="checkbox"/>	Author ▾	Label ▾	Projects ▾	Milestones ▾	Assignee ▾	Sort ▾			
<input type="checkbox"/>		Project dates should be optional	quick-fix-possible				 1		 1
		#49 by mariapraetzelis was closed 43 minutes ago							
<input type="checkbox"/>		Funding cardinality mismatch	quick-fix-possible				 1		 2
		#42 by jochenklar was closed 43 minutes ago							
<input type="checkbox"/>		Optional lists with mandatory items (license, security_and_privacy)	quick-fix-possible				 1		 6
		#32 by MarekSuchanek was closed 43 minutes ago							
<input type="checkbox"/>		Number of funding per project inconsistency	quick-fix-possible				 1		 1
		#31 by MarekSuchanek was closed 43 minutes ago							
<input type="checkbox"/>		grant_id should not be required	quick-fix-possible				 1		 3
		#29 by brii was closed 43 minutes ago							
<input type="checkbox"/>		Definition of time and dates to be refined	quick-fix-possible				 1		 5
		#6 by cmzwolf was closed 43 minutes ago							

Thank you for the feedback!

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?

<input type="checkbox"/>	10 Open ✓ 24 Closed	Author ▾	Label ▾	Projects ▾	Milestones ▾	Assignee ▾	Sort ▾
<input type="checkbox"/>	How to express if something is planned decision enhancement #46 opened on Jul 7 by TomMiksa						
<input type="checkbox"/>	List of pending adoptions #44 opened on Jun 26 by TomMiksa						
<input type="checkbox"/>	Datasets with multiple and no (yet) IDs decision #34 opened on May 28 by MarekSuchanek						3
<input type="checkbox"/>	Dealing with changes in identifiers over time, adding provenance for local IDs decision duplicate #33 opened on May 27 by tr325						2
<input type="checkbox"/>	How to add funder template specific extensions needs be worked out decision duplicate #30 opened on May 27 by hmpf						1
<input type="checkbox"/>	Add specification for how to extend the schema decision #27 opened on Mar 10 by briri						4
<input type="checkbox"/>	Add support for organization information enhancement #26 opened on Mar 10 by briri						
<input type="checkbox"/>	"creator" field should be added to "dataset" enhancement #25 opened on Mar 10 by SotosTsepe						
<input type="checkbox"/>	DataCite Roles may not be a good match for DMP roles? vocabulary #16 opened on Nov 20, 2019 by rwwh						6
<input type="checkbox"/>	Is there a distinction between data sets used in the project and data sets produced decision #14 opened on Nov 20, 2019 by rwwh						2

<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 will **organise a call in November** to resolve open issues
 - Doodle will be sent out to the DMP Common Standard mailing list



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.

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

3. [Making things explicit \(planned, performed\)](#)

- Currently we use dates to indicate this -> some arithmetic needed
- Should this be made explicit, e.g. by adding special fields?

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?
 - Continue work from the [hackathon on maDMPs](#)
 - [Mapping to funder templates](#)

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

maDMPs – tell us your story!

➤ 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

- › [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.** TPD L 2018: 69-80](#)