



18th RDA Plenary 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
- Part 2 Adoptions and new use cases
 - Now that machine actionable DMPs are becoming available, what could we do with them?
 - > by Daniel Mietchen
 - Update on maDMPs in a government context: Developing an RDA maDMP extension and logic model
 - > by Claire Austin
 - > Automating Evaluation of Machine-Actionable Data Management Plans with Semantic Web Technologies
 - > by Lea Brugger and Raffael Foidl
 - > maDMP4LS: integration of DMP OPIDoR in the French infrastructure landscape
 - > by Benjamin Faure and Konogan Bourhy
- Part 3 Maintenance of the recommendation
- > Part 3 Wrap up and next steps
 - What are the new topics we should tackle together to ease adoption of maDMPs?
 - > Are there any updates needed in the recommendation?



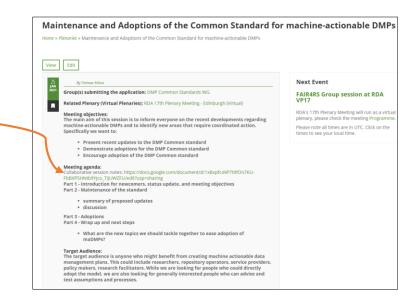


Participants

- Collaborative notes
 - > LINK
 - > Add your name to the list
 - > Co-edit!

> Let's get to know each other

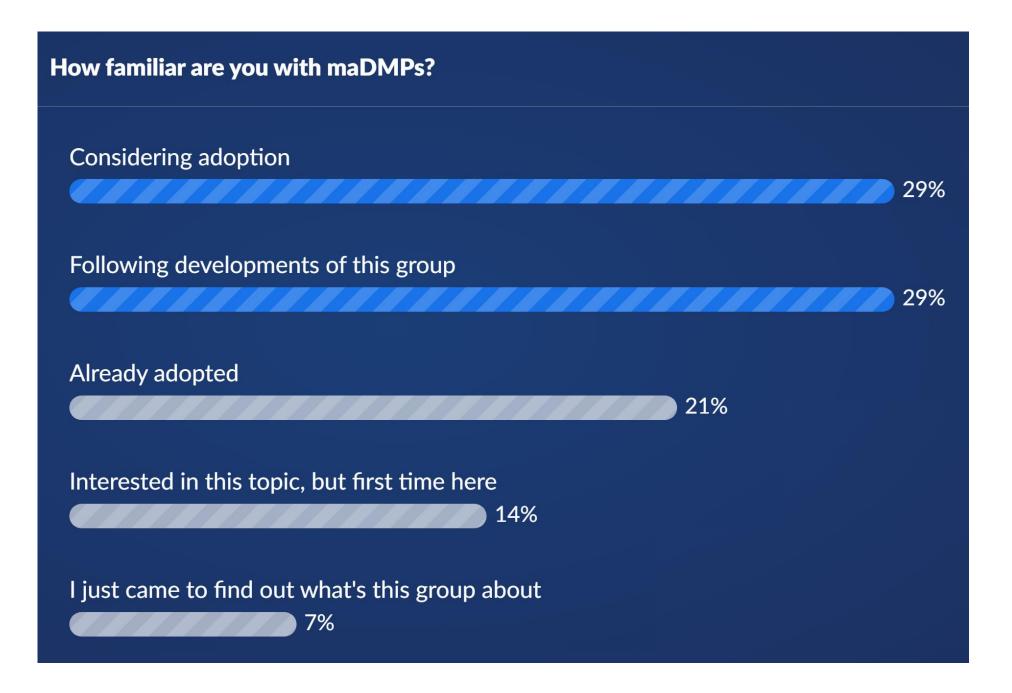




Go to slido.com and enter #926 519







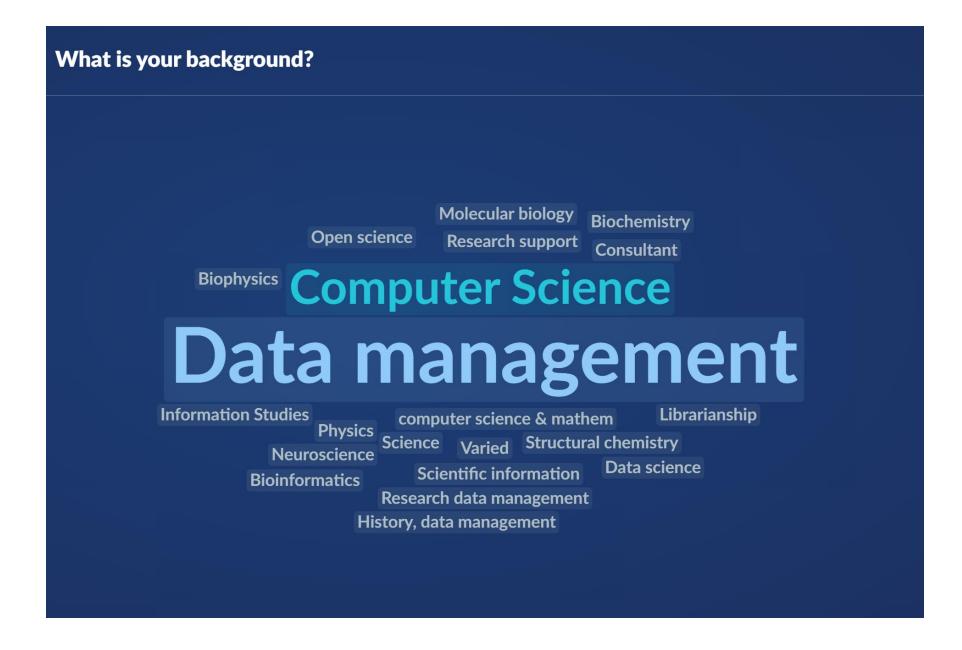
















Introduction for newcomers

Part 1

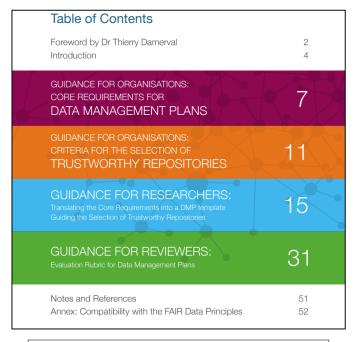




Data Management Plans (DMPs)

	Data Officer Who is responsible for the data management and the DMP of the project (name/e					
	Data Characteristics					
1.1	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 tareat 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 (EAIR) 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.)				
	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?				





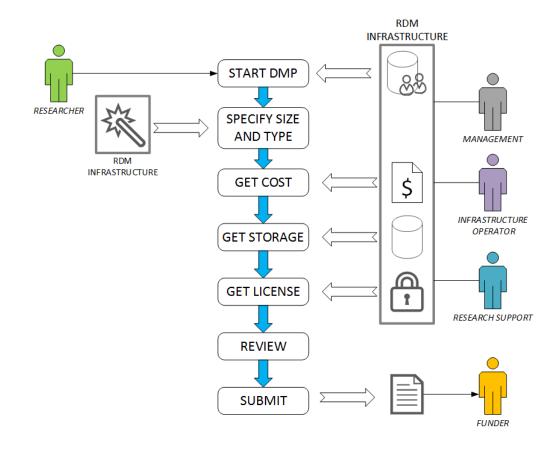
4 For procedural elements of implementing DMPs, see the RDA DMP Common Standards Working Group: https://www.rd-alliance.org/groups/dmp-common-standards-wg





Machine-actionable DMPs (maDMPs)

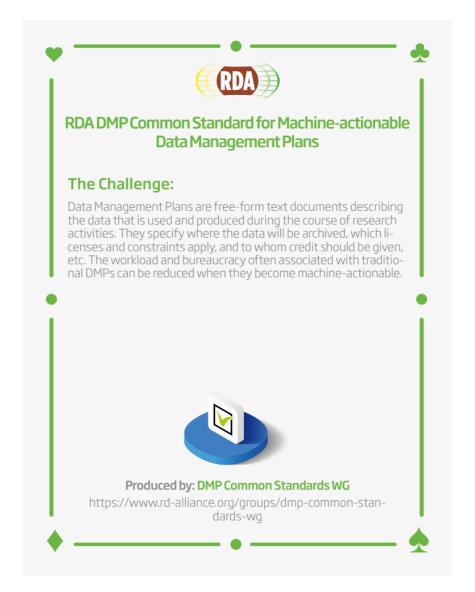
- Machine-actionable DMPs
 - Living documents
 - automate data management
 - > collect information from systems
 - trigger actions in systems
 - facilitate validation
- >This requires
 - > well-defined RDM workflows
 - data management infrastructure
 - common standard to represent information







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.

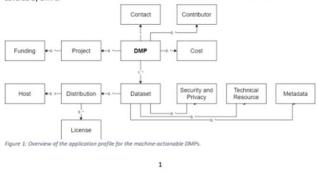
DMP - Provides high level information about the DMP, e.g. its title, modification date, etc. It is the root of this application profile.

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







maDMPs - documentation

Name	Description	Data Type	Cardinality	Example Value	•
contact	Contact person for a DMP	Nested Data Structure	1		NOT a q
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	0n		NOT a t
st	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	0n	Mo	ost field
eated	Date and time of the first version of a DMP. Must not be changed in subsequent DMPs.	DateTime	1	2019-03-13 13:13	
taset	To describe data on a non- technical level.	Nested Data Structure	1n		

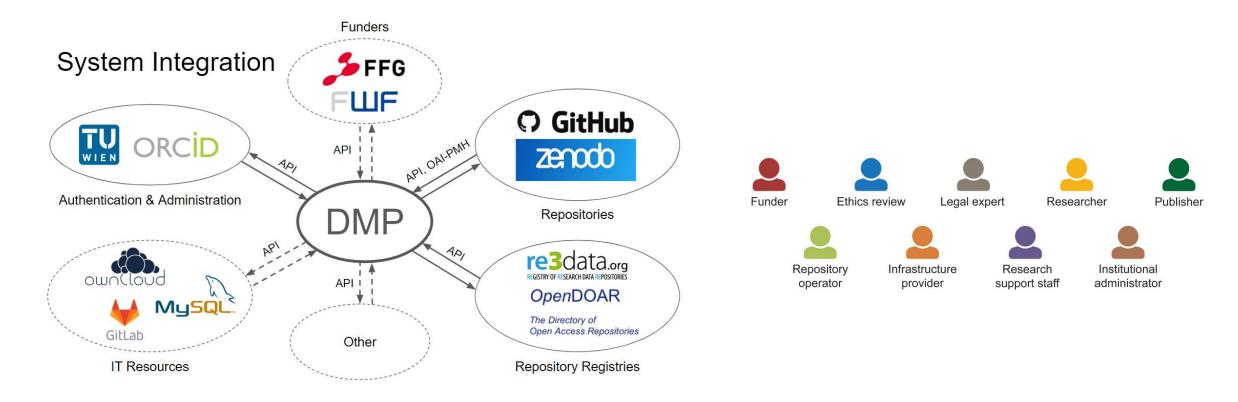
https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard/blob/master/docs/index.md





RDM Infrastructure

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







Adoptions (selected)



















NORWEGIAN CENTRE FOR RESEARCH DATA





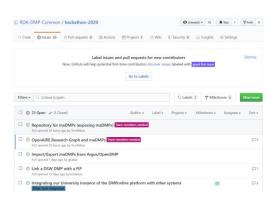
DMP Common Standards WG

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









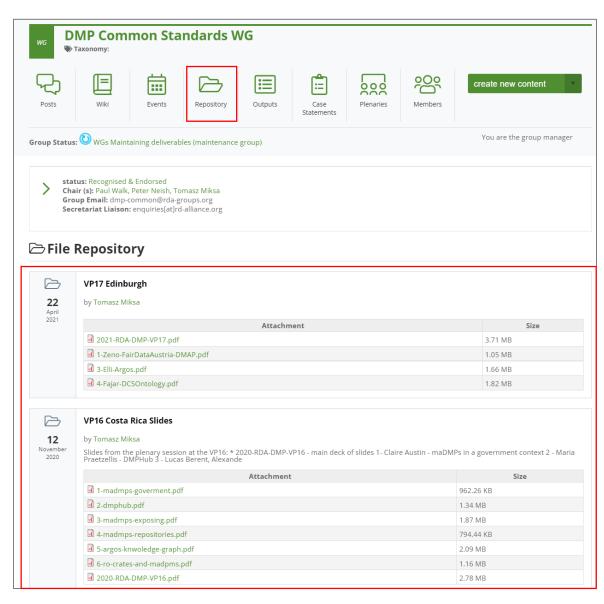




DMP Common Standards WG

- Slides from all our sessions are in the repository
- > Today's presentations will also be there

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







Papers published!

- Describes the full story of developing the recommendation
- Example of a minimal maDMP
- > Presents adoptions
 - > Haplo
 - Open Research Publishing Platforms
 - DMP Tool
 - > DMPonline
 - > DMP OPIDoR
 - Data Stewardship Wizard
 - > NSD DMP
 - > Argos
 - Research infrastructure at TU Wien
 - Easy DMP



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





Papers accepted!

Practice Paper

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

João Cardoso¹, Leyla J. Garcia², Tomasz Miksa³

¹Universidade de Lisboa, Instituto Superior Técnico & INESC-ID, Lisboa, Portugal

²ZB MED Information Centre for Life Sciences, Cologne, Germany

³SBA 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 sapects 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.

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 almost 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, fundine 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)

Contains:

 Summary of results from hackathon

DATA SCIENCE JOURNAL

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 state 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 interoenability of iresearch 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.

 $CCS Concepts: \ \textbf{-Applied computing} \rightarrow \textbf{Enterprise data management}; \ \textbf{Business process management}; \ \textbf{Trarchitectures}; \ \textbf{-Information systems} \rightarrow \textbf{Digital libraries and archives}; \ \textbf{-Social and professional topics} \rightarrow \textbf{Automation}.$

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/10.1145/3409396

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; Vienna, Austria; Andreas Rauber, rauber@ifs.tuwien.ac.at, TU Wien, Vienna, Austria; A

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

 $\,$ $\,$ $\,$ 2020 Copyright held by the owner/author(s).

2158-656X/2021/1-ART1

https://doi.org/10.1145/349039

ACM Trans. Manag. Inform. Syst., Vol. 1, No. 1, Article 1. Publication date: January 2021

Contains:

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

ACM Transactions on Management Information Systems



17



Adoptions and new use cases for maDMPs

Part 2





Adoptions and new use cases for maDMPs

- > Rules
 - > 10 minutes per presentation
 - > 5 minutes for questions directly after each presentation
 - > Ask questions in the zoom chat during the presentation or speak up after the presentation
- > Talks
 - Now that machine actionable DMPs are becoming available, what could we do with them?
 by Daniel Mietchen
 - 2. Update on maDMPs in a government context: Developing an RDA maDMP extension and logic model by Claire Austin
 - 3. Automating Evaluation of Machine-Actionable Data Management Plans with Semantic Web Technologies by Lea Brugger and Raffael Foidl
 - 4. maDMP4LS: integration of DMP OPIDoR in the French infrastructure landscape by Benjamin Faure and Konogan Bourhy





Maintenance of the recommendation

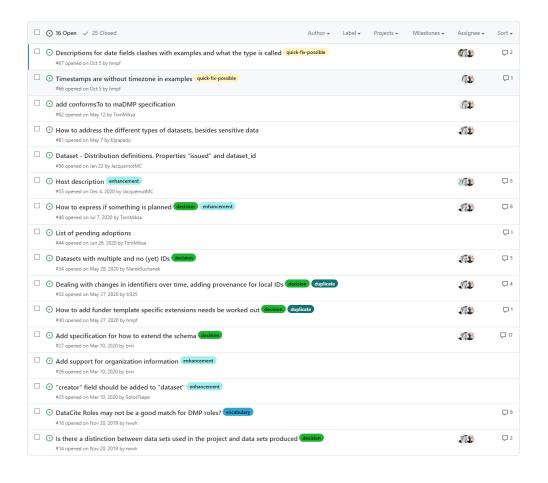
Part 3





Provide feedback on issues

- > Your opinion matters!
 - > https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard/issues

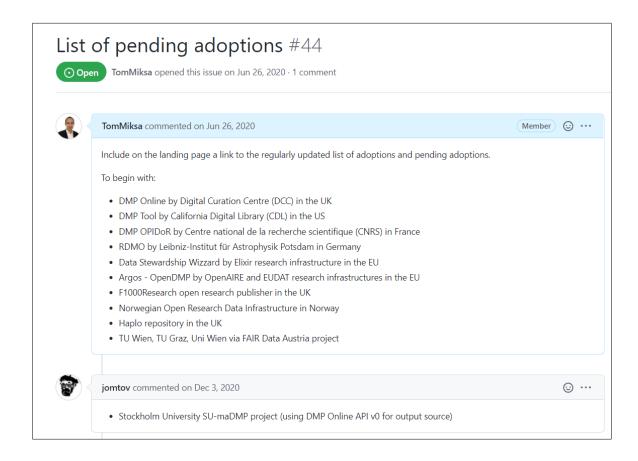






List of pending adoptions

- > Please add yours to the list
 - > https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard/issues/44

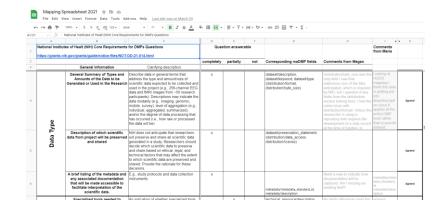






Mapping to funder templates (pending tasks)

- > 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



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

Please review <u>the documents</u> and comment them!





Maintenance of the recommendation



> Anything you would like to report?

> Any improvements needed?





Wrap-up and next steps

Part 4





Community of Practice?

FORMAL REQUIREMENTS

- > Launched by a WG/IG that produced a recommendation
 - Recommendation on maDMPs by the DMP Common Standards WG
- > Members from 10 countries and 3 continents
 - Active DMPs IG: 380+ members...
 - DMP Common Standards WG: 240+ members...
- > Letters of support from institutions
 - Institutions that adopt maDMPs?
- > Review every 18 months
- > At least 3 co-chairs from at least 3 continents
 - No problem, right?

PROPOSAL

- CoP on machine-actionable DMPs
 - Focus on workflows, services, systems that use the recommendation
 - Feedback from adopters and lessons learned
 - Tackling new challenges
 - **E.g.** Extensions, Controlled vocabularies, Serializations, etc.
 - Supports in adoption
 - Promotes success stories / lessons learned
 - To formalize what we do and who we are anyway!
- DMP Common Standards WG continues to maintain the recommendation
 - > Updates to the specification if needed
- Active DMPs IG
 - Place for discussion on all (other) topics related to DMPs

What do you think? Would you be interested? Can you help in establishing it?





- > Join at slido.com #288 176
 - > Different number than before!







Should we establish a community of practice (CoP) on maDMPs?

Yes, I am happy to join it once it's established.

50%

Yes, that's an excellent idea and I can help setting it up.

40%

Well, it's good as it is now. I don't see any need for the CoP.



10%

No, it's a waste of time and pure bureaucracy.



0%





2nd hackathon on maDMPs in 2022?

- Good chance to get together and get some work done?
- Possibility to test new ideas?
- Good for onboarding?

Results of the hackathon

The recording of the Grand Finale session can be found here. Each of the 12 teams participating in the hackathon has summarized the results The recording also includes live demonstrations.

Table below presents a list of teams, topics they dealt with and links to the work completed during the hackathon

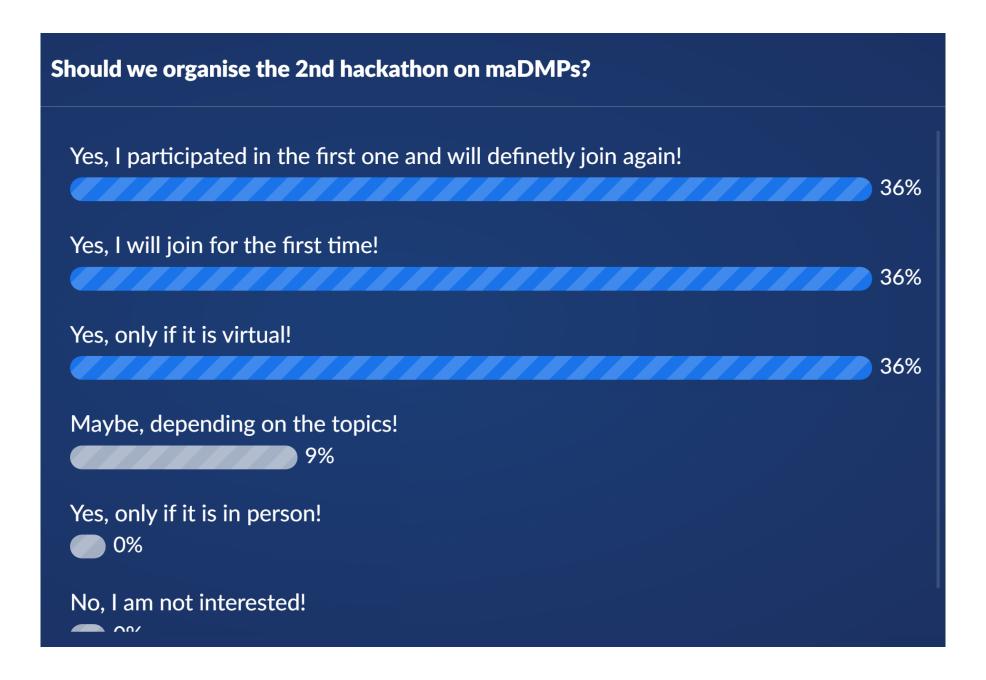
Team	Торіс	Links	
The Unaturals	New version of the DCSO (DMP Common Standard Ontology)	DCSO Repository, Issue Document, slides	
maDMP Link	Integrating a Converis CRIS/RIMS with DMPRoadmap and Export/Import maDMP from Figshare	results	
DSW	Export/import of maDMPs from/to Data Stewardship Wizard	report, slides	
The Datatypists	Add data types and design suppor for datasets to make EasyDMP more compatible with RDA DMP Common Standard	report, json to test imports	
InsTmaDMP	Aligning institution RDM to maDMP Common Standard	results	
TigTag	Mapping of maDMP standard to funder templates	report	
DMP Exchange	Exchange DMPs between DMP Tools using the RDA Common Standard	report	
Something	maDMP integration with project data management workflow	slides	
DMP InvenioRDM	maDMP integration into InvenioRDM	results	
RDMO	maDMP export from RDMO	slides	
DMP Ninjas	Import/Export maDMPs from Argos/OpenDMP	results	
Fancycatmeme	maDMP integration with Research Data Connectome Data Pipeline	article, slides	



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











What would you like to work on during the hackathon? Extend what is considered as minimal yes, extensions Modeling more DMP information as extensions





What can we do to support adoption of maDMPs? What are the interesting use cases?

Tools and use cases in practice that make life better for researchers.

Reuse by the funders

Better embed them in tools

use case: creating a DMP from several several DMPs; ie, modularity

Integration in the research workflow





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

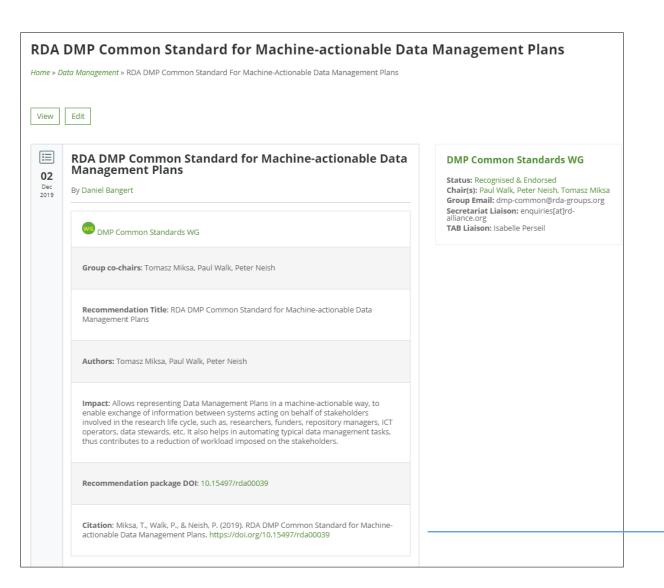


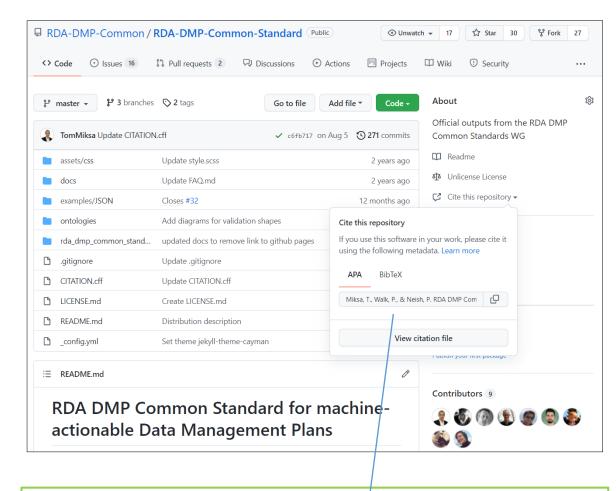
Peter Neish





Please cite the recommendation





Miksa, T., Walk, P., & Neish, P. (2019). RDA DMP Common Standard for Machine-actionable Data Management Plans.

https://doi.org/10.15497/rda00039





Publications about maDMPs

- > 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

