



Machine-actionable DMPs: what? why? how?

Tomasz Miksa

TU Wien & SBA Research, Austria RDA DMP Common Standards WG







Introduction

- > Data Management Plans (DMPs)
- > Machine-actionable DMPs (maDMPs) and the difference towards DMPs
- > RDA DMP Common Standard for maDMPs
 - > What it is and what it is not!
- > Adoptions of the recommendation
- > Getting more information
- Questions





Where are you from?

(i) Start presenting to display the poll results on this slide.



What is your background/role?

(i) Start presenting to display the poll results on this slide.





(i) Start presenting to display the poll results on this slide.



Introduction





>DMP is a formal document

It outlines what you will do with your data during and after you complete your research

- > It ensures your data is safe for the **present** and the **future**
- Required by funders and universities worldwide







DMP is an <u>awareness tool</u>!

>DMP makes you think

> what data you will use and where you get it from
> what infrastructure, software, licenses are needed
> what will be the output of your research
> how you will share your research outputs

>DMP helps you organise yourself better

>DMP can reveal how solid your methodology is







>Shortcomings of existing DMPs

> manually completed, vague, not updated, considered bureaucracy, completed last minute, ...

		· · · · · · · · · · · · · · · · · · ·
	Data Officer	Who is responsible for the data management and the DMP of the project (name/email address)?
I.	Data Characteristics	
l.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?
Ш	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 (<u>FAIR</u>) 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.)
ш	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?
111.2	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?



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
- <u>common standard</u> <u>to represent information</u>





RDA DMP Common Standards Working Group

- > Launched in 2017 by Research Data Alliance (RDA)
- > 240+ members of the WG
- > Open consultations
 - > To identify the scope and existing standards
- > Prototype software
 - > To demonstrate how maDMPs can be used
- Hackathon
 - > to facilitate adoption
 - > 71 participants, 12 teams, 21 countries
 - > Integrations, mappings, etc.



-	The second secon
	DESIMPORT PRODUCTOR
	Correlating Alcohol Consumption and UFI
	Authors Nacionation - State (SSE-SSE-SSE) - SSE - SSEND(SSES) - SSE
	Document Version and Date
	Gathered Data IF US answerth and the Second
	Software and American Street S
	Documentation
	and interpret of the second se
	Ethical Overstions













RDA DMP Common Standard for maDMPs





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.



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 <u>online</u>. 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

Properties in 'dmp'

Nama	Description	Data Tura	Cardinality	Evample	/alua
contact	Contact person for a DMP	Nested Data Structure	1	Example v	
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		NO
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	0n		Mos
created	Date and time of the first version of a DMP. Must not be changed in subsequent DMPs.	DateTime	1	2019-03-13 1	3:13
dataset	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



Machine-actionable DMP

> Example: https://doi.org/10.5281/zenodo.6467730

```
"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.



```
"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"
   } 1,
   "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.



Adoptions (selected)





Adoptions

Some examples





> once-only principle

> do not ask researchers same questions in different places

> maDMPs are the 'glue' between different systems

> Automate getting information in and out







www.damap.org



FAIR Data Austria DAMAP Tool for Machine-actionable Data Management Plans







DMP OPIDoR in France

2. Get the project informations





Getting more information





Read more in...

- 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

ading: Application Profile for Machine-Actionable Data magement Plans

Special Collection: Research Data Alliance Results

Research Papers

Application Profile for Machine-Actionable Data Management Plans

Authors: 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, Sarah Jones

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 recommendation.

Keywords: application profile, maDMPs, common standard, machine actionable, RDA

How to Cite: Miksa, T., Walk, P., Neish, P., Oblasser, S., Murray, H., Renner, T., Jacquemot-Perbal, M.-C., Cardoso, J., Kvamme, T., Praetzellis, M., Suchánek, M., Hooft, R., Faure, B., Moa, H., Hasan, A. and Jones, S., 2021. Application Profile for Machine-Actionable Data Management Plans. *Data Science Journal*, 20(1), p.32. DOI: http://doi.org/10.5334/dsj-2021-032



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



Read more in...

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 imparticula 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 interpenability 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 machineactionability of DMPs. We propose a reference architecture of a machine-actionable DMP support system that can be embedded into an institutional research that 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.

 $\label{eq:CCSConcepts} CCSConcepts \bullet Applied computing \to Enterprise data management; Business process management; IT architectures; \bullet Information systems \to Digital libraries and archives; \bullet Social and professional topics \to Automation.$

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/10.1145/490396

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@sha-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.

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).

© 2020 Copyright held by the owner/aut 2158-656X/2021/1-ART1

https://doi.org/10.1145/3490396

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

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 & NISSC-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 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 on bety 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 resarchers, 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)

DATA SCIENCE JOURNAL

http://doi.org/10.5334/dsj-2021-035

Contains: - Summary of results from hackathon

ACM Transactions on Management Information Systems

https://doi.org/10.1145/3490396



DMP Common Standards WG

Slides from all our sessions are in the repository

> Presentations froom tool providers!

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

WG	MP Common Standards WG	
Posts	Wiki Events Repository Outputs Case Statements Plenaries Members	create new content
Group Statu	: 🕑 WGs Maintaining deliverables (maintenance group)	You are the group manager
File	ir (s): Paul Walk, Peter Neish, Tomasz Miksa up Email: dmp-common@rda-groups.org retariat Liaison: enquiries[at]rd-alliance.org Repository VP17 Edinburgh by Tomasz Miksa	
2021	Attachment	Size
	2021-RDA-DMP-VP17.pdf	3.71 MB
	D	
	d 1-Zeno-FairDataAustria-DMAP.pdf	1.05 MB
	d 1-Zeno-FairDataAustria-DMAP.pdf d 3-Elli-Argos.pdf	1.05 MB 1.66 MB
	d 1-Zeno-FairDataAustria-DMAP.pdf d 3-Elli-Argos.pdf d 4-Fajar-DCSOntology.pdf	1.05 MB 1.66 MB 1.82 MB
12 November 2020	I -Zeno-FairDataAustria-DMAP.pdf I -Zeno-FairDataAustria-DMAP.pdf I -Fajar-DCSOntology.pdf VP16 Costa Rica Slides by Tomasz Miksa Slides from the plenary session at the VP16; * 2020-RDA-DMP-VP16 - main deck of slides 1- Claire Austin - maDM Praetzellis - DMPHub 3 - Lucas Berent, Alexande	1.05 MB 1.66 MB 1.82 MB
12 November 2020	I -Zeno-FairDataAustria-DMAP.pdf I - Zeno-FairDataAustria-DMAP.pdf I - Zeno-FairDataAustria-DMAP.pdf <t< th=""><th>1.05 MB 1.66 MB 1.82 MB IPs in a government context 2 - Maria Size</th></t<>	1.05 MB 1.66 MB 1.82 MB IPs in a government context 2 - Maria Size
12 November 2020	▲ 1-Zeno-FairDataAustria-DMAP.pdf ▲ 3-Elli-Argos.pdf ▲ 4-Fajar-DCSOntology.pdf ✓ <	1.05 MB 1.66 MB 1.82 MB PS in a government context 2 - Maria Size 962.26 KB
12 November 2020	▲ 1-Zeno-FairDataAustria-DMAP.pdf ▲ 3-Elli-Argos.pdf ▲ 4-Fajar-DCSOntology.pdf ✓ <	1.05 MB 1.66 MB 1.82 MB
12 November 2020	I -Zeno-FairDataAustria-DMAP.pdf I -Zeno-FairDataAustria-DMAP.pdf I -Fajar-DCSOntology.pdf VP16 Costa Rica Slides by Tomasz Miksa Slides from the plenary session at the VP16: * 2020-RDA-DMP-VP16 - main deck of slides 1- Claire Austin - maDM Praetzellis - DMPHub 3 - Lucas Berent, Alexande Attachment I 1-madmps-goverment.pdf I 2-dmphub.pdf I 3-madmps-exposing.pdf	1.05 MB 1.66 MB 1.82 MB Ps in a government context 2 - Maria 962.26 KB 1.34 MB 1.87 MB
12 November 2020	▲ 1-Zeno-FairDataAustria-DMAP.pdf ▲ 3-Elli-Argos.pdf ▲ 4-Fajar-DCSOntology.pdf ✓ <	1.05 MB 1.66 MB 1.82 MB Ps In a government context 2 - Maria Ps In a government context 2 - Maria 962.26 KB 1.34 MB 1.87 MB 994.44 KB
12 November 2020	I -Zeno-FairDataAustria-DMAP.pdf I -Zeno-FairDataAustria-DMAP.pdf I -Fajar-DCSOntology.pdf VP16 Costa Rica Slides by Tomasz Miksa Slides from the plenary session at the VP16: * 2020-RDA-DMP-VP16 - main deck of slides 1- Claire Austin - maDM Praetzellis - DMPHub 3 - Lucas Berent, Alexande I -madmps-goverment.pdf I -madmps-exposing.pdf I -madmps-repositories.pdf I -madmps-repositories.pdf	1.05 MB 1.66 MB 1.82 MB Ps in a government context 2 - Maria Ps in a government context 2 - Maria 962.26 KB 1.34 MB 1.87 MB 794.44 KB 2.09 MB
12 November 2020	I -Zeno-FairDataAustria-DMAP.pdf I -Zeno-FairDataAustria-DMAP.pdf I -Fajar-DCSOntology.pdf VP16 Costa Rica Slides by Tomasz Miksa Slides from the plenary session at the VP16: * 2020-RDA-DMP-VP16 - main deck of slides 1- Claire Austin - maDM Praetzellis - DMPHub 3 - Lucas Berent, Alexande I -madmps-goverment.pdf I -madmps-exposing.pdf I -madmps-repositories.pdf I -madmps-repositories.pdf I -madmps-repositories.pdf I -madmps-repositories.pdf	1.05 MB 1.66 MB 1.82 MB



- 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



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

