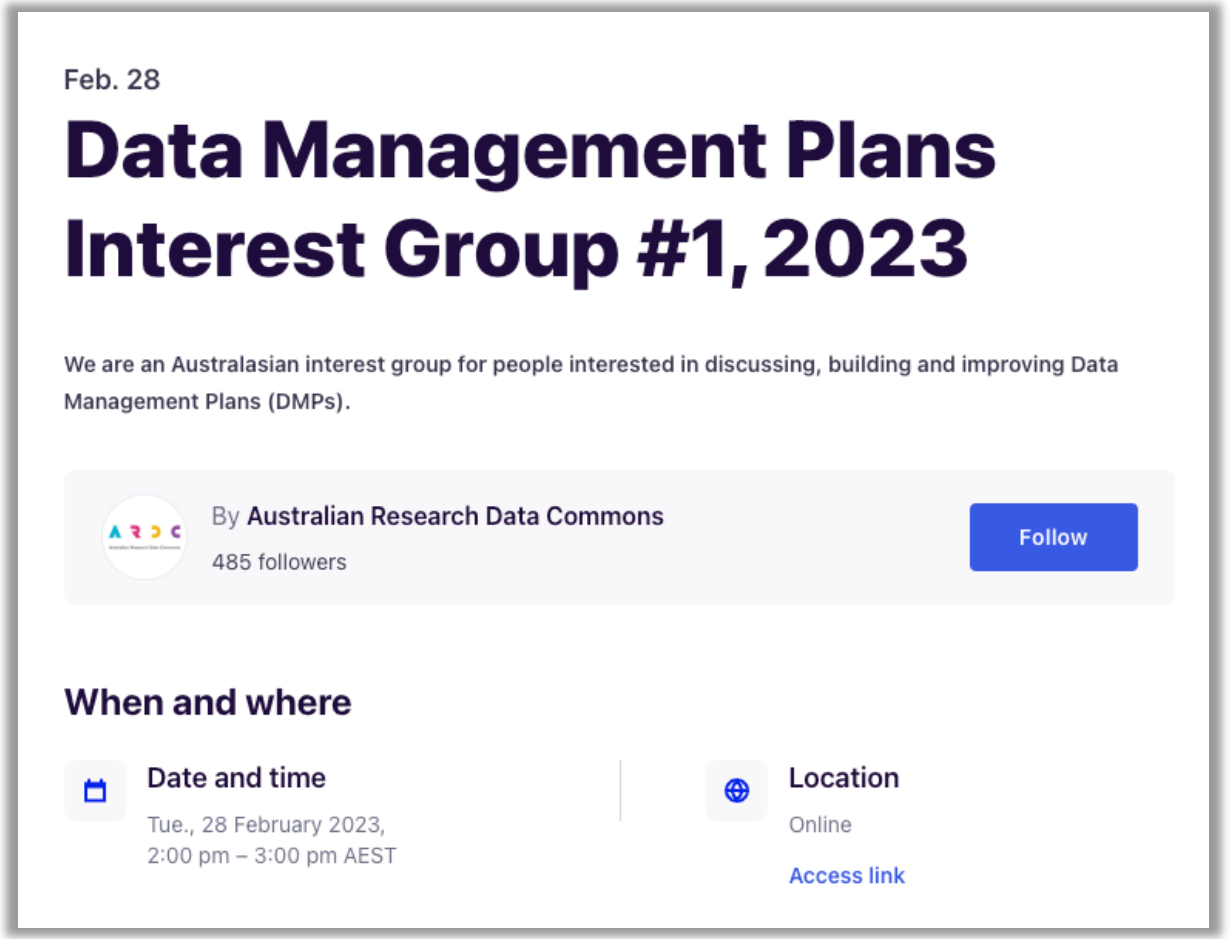
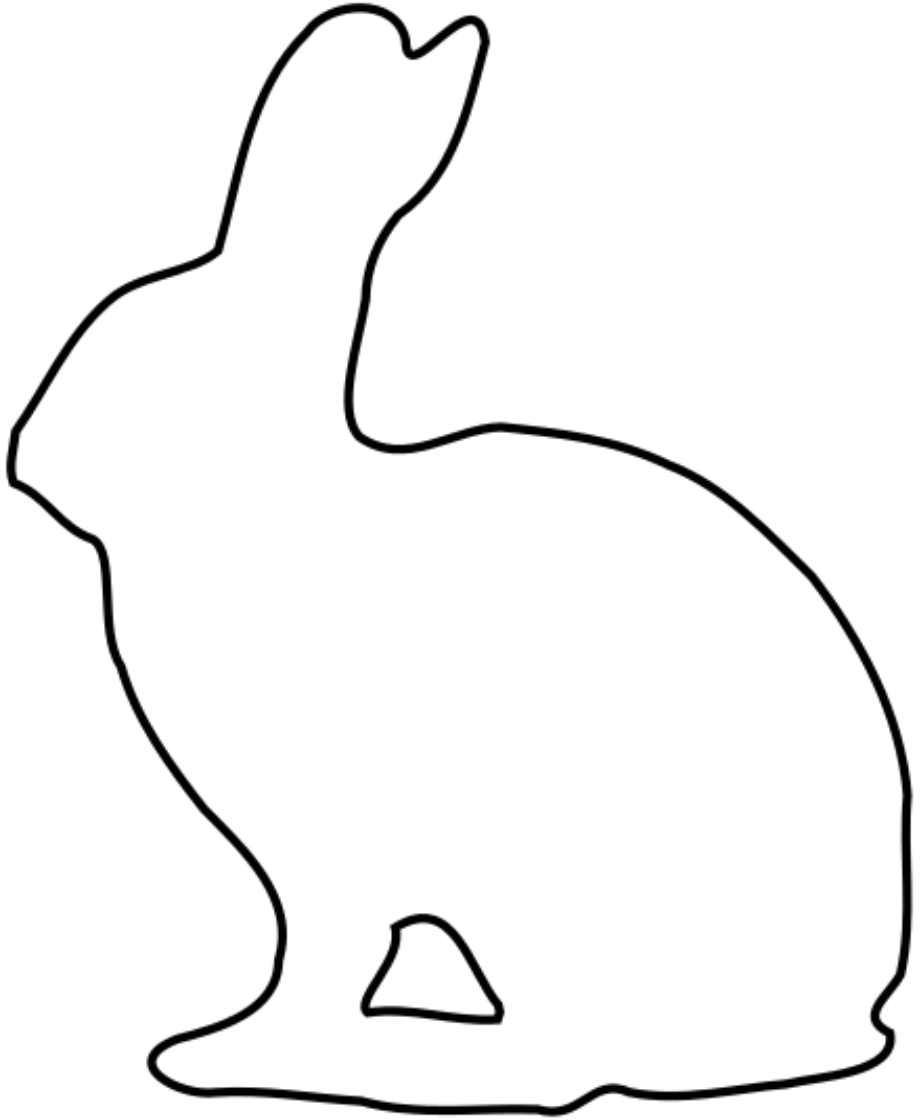
Machine-Actionable Data Management Plans



Peter Neish @peterneish

Outline

- Research Data Alliance DMP groups
- Introductions to machine-actionable DMPs
- Some example adoptions
- DMP ontology
- Where to find out more



RDA Groups

Active DMPs *Interest Group*

400+ members!

- Place for discussion on all topics related to DMPs
- Can trigger new WGs
  - Like it happened in the past with the DMP Common Standards WG

DMP Common Standards *Working Group*

240+ members!

- Maintenance mode
  - Updates the recommendation when necessary
  - Can add new serializations
    - JSON, RDF, …
  - Supports adopters of the recommendation

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?*

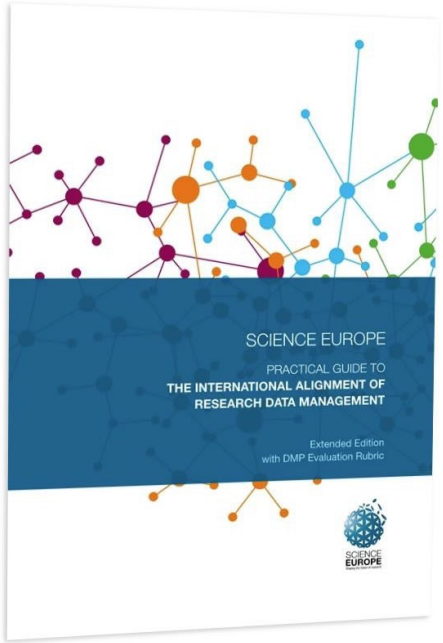


Table of Contents

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

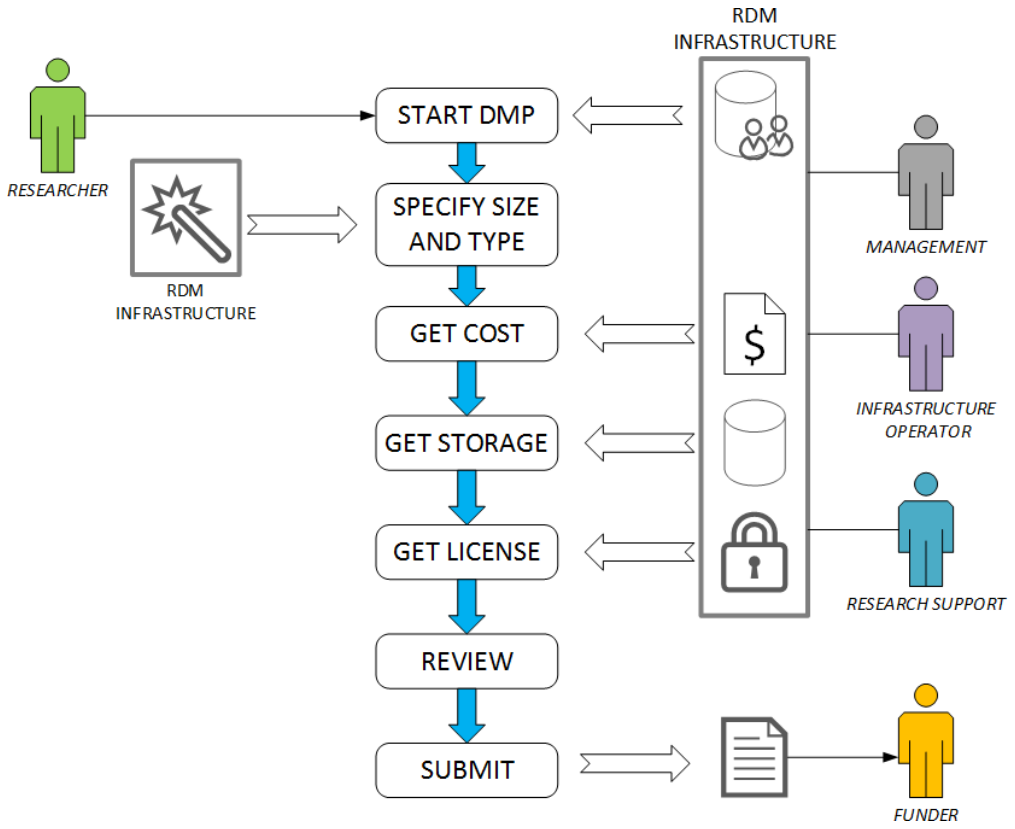
Machine-actionable DMPs (maDMPs)

Machine-actionable DMPs

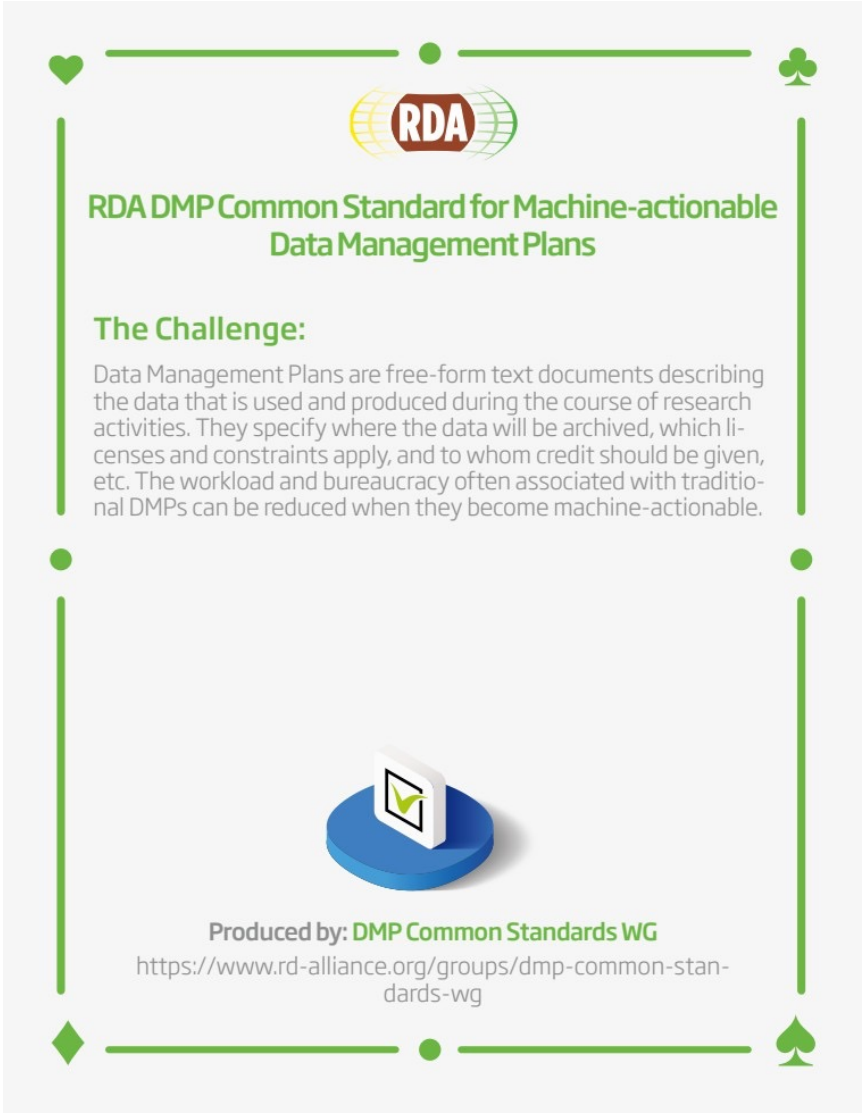
- Living documents
- automate data management
  - collect information from systems
  - trigger actions in systems
- facilitate validation

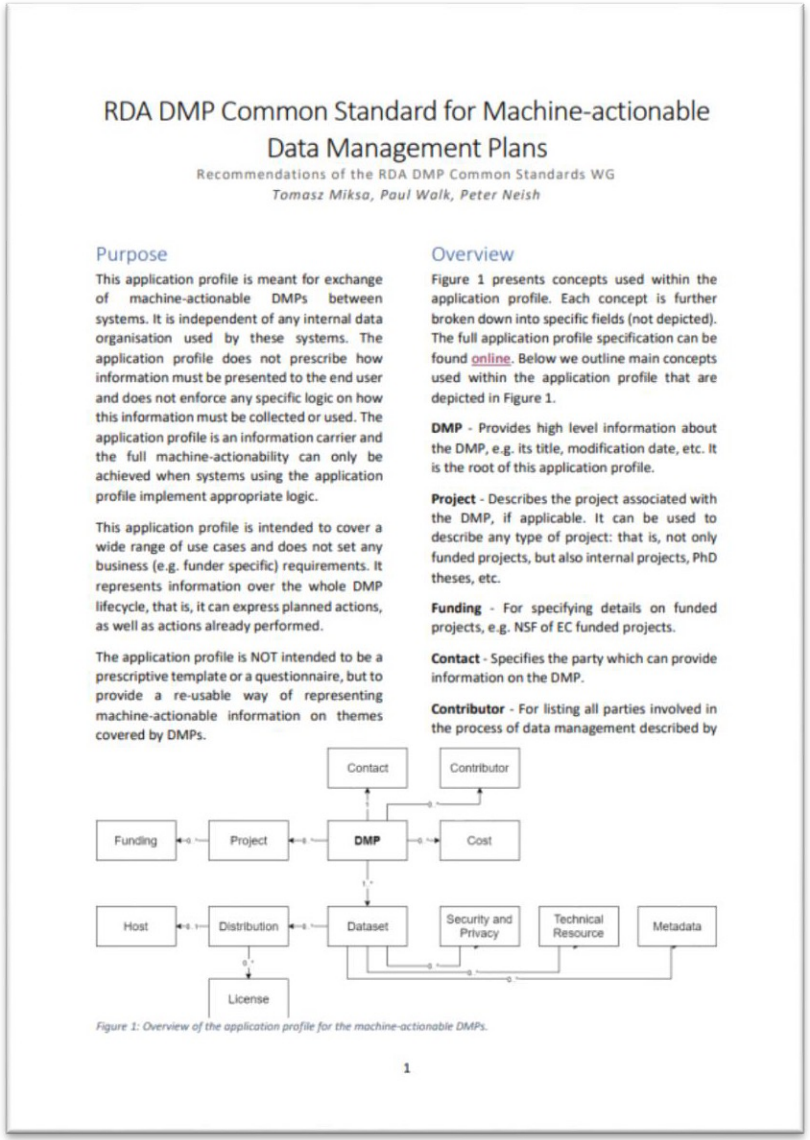
This requires

- well-defined RDM workflows data management infrastructure
- <u>common standard to represent information</u>



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

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.

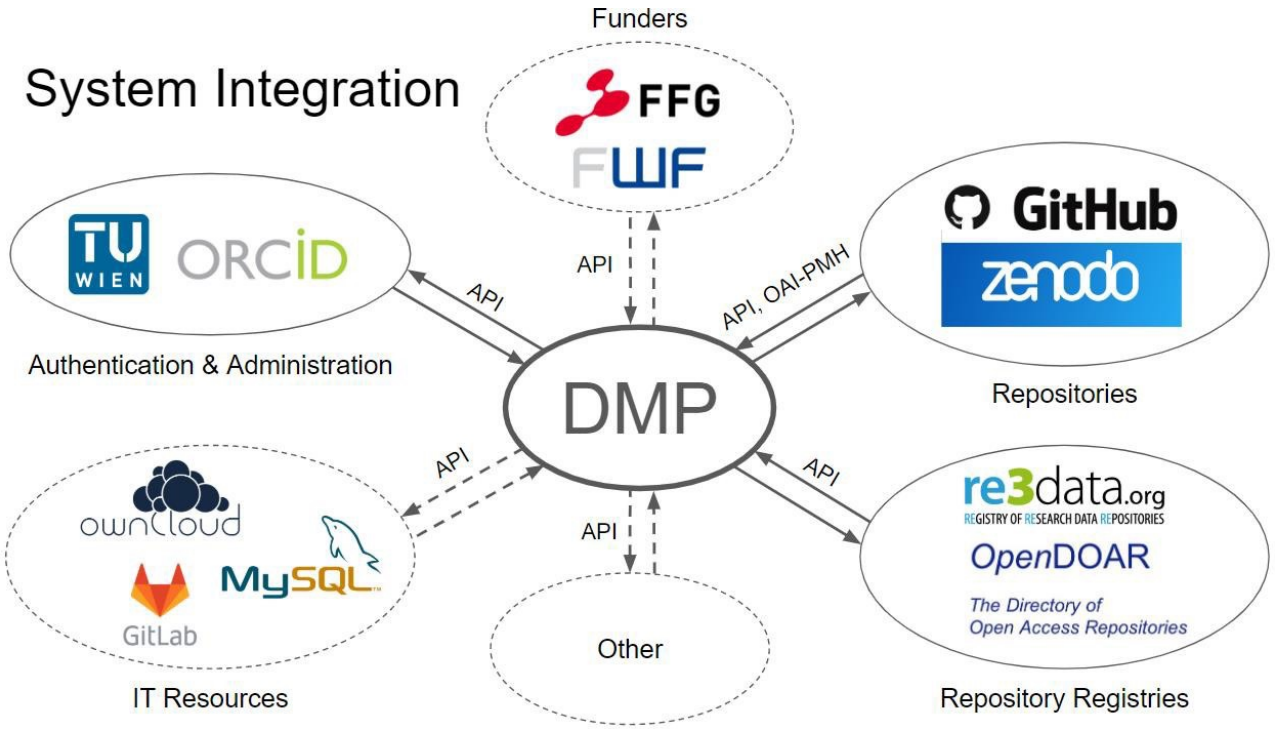
Machine-actionable DMP


```
"dataset" : [ {
  "description" : "For each dataset (fish and employee) the original dataset will be split into two subsets, one for training and one for testing the
  performance.",
  "distribution" : [ {
    "access_url" : "https://zenodo.org/record/6467615",
    "byte_size" : 2999302,
    "data_access" : "open",
    "description" : "For each dataset (fish and employee) the original dataset will be split into two subsets, one for training and one for testing the
    performance.",
    "format" : [ "STRUCTURED_TEXT" ],
    "host" : {
      "description" : "ZENODO builds and operates a simple and innovative service that enables researchers, scientists, EU projects and institutions to share
      and showcase multidisciplinary research results (data and publications) that are not part of the existing institutional or subject-based repositories of
      the research communities.\nZENODO enables researchers, scientists, EU projects and institutions to:\neasily share the long tail of small research results
      in a wide variety of formats including text, spreadsheets, audio, video, and images across all fields of science.\ndisplay their research results and get
      credited by making the research results citable and integrate them into existing reporting lines to funding agencies like the European
      Commission.\neasily access and reuse shared research results.",
      "pid_system" : [ "doi" ],
      "storage_type" : "other",
      "support_versioning" : "unknown",
      "title" : "Zenodo",
      "url" : "https://zenodo.org/"
    },
    "license" : [ {
      "license_ref" : "https://creativecommons.org/licenses/by/4.0/",
      "start_date" : "2022-05-01 22:00:00.0"
    } ],
    "title" : "Training and Test Subsets for Performance Comparison of kNN and GD"
```


Each **dataset** has a **title** and a human readable **description**.
It is also clear what the **format**, **size** and the **location** of the dataset are.
License and mode of **access**, including any exact **embargo** periods, are specified as well.

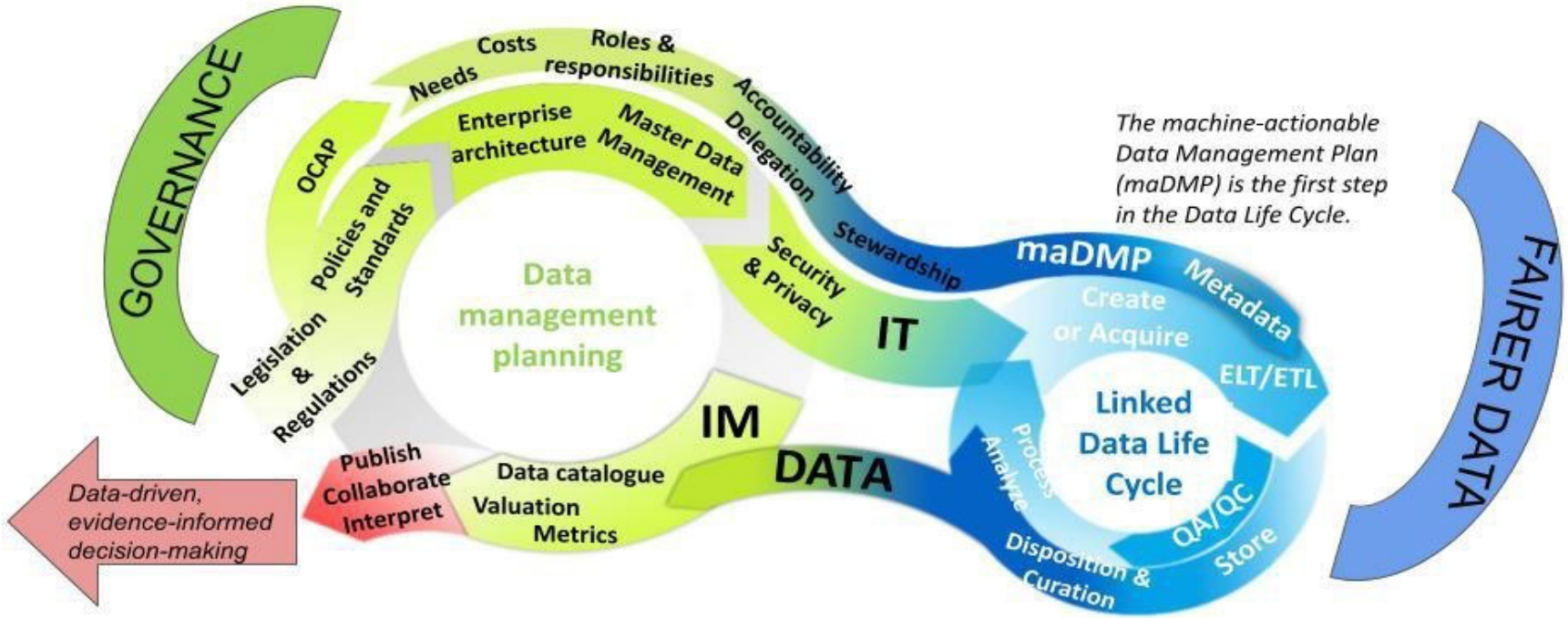
RDM Infrastructure

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



Adoptions (selected)

Governance and Data Management Planning vs. Data Management Plan (DMP) and Data Life Cycle



OPEN GOVERNMENT - OPEN SCIENCE - OPEN DATA - CARE PRINCIPLES - HUMAN RIGHTS - TRUST
Open-by-default-by-design

Findable, Accessible, Interoperable, Reusable, Ethical, Reproducible

https://www.rd-alliance.org/system/files/documents/2%20-%20Austin%20CC%20-%20Update%20on%20maDMPs%20in%20a%20government%20context.pptx.pdf

DMP Common Standard Ontology (DCSO)

DMP Common Standard Ontology (DCSO)

This ontology aims to represent the DMP Common Standard model, through the usage of semantic web technology. It represents the DMP Common Standard model using the Web Ontology Language (OWL).

This is still a work in progress. The idea behind having the DMP Common Standard model represented in OWL, is to explore the potential behing a machine-readable version of the model.

Directory Organization

```
├── core - Core elements of the DCSO
├── diagrams - DCSO diagrams
├── examples - Examples of DMPs represented using the DCSO
├── extensions - Context specific extensions of the DCSO
└── validation - Shex representation of the DCSO constraints
```

Authors

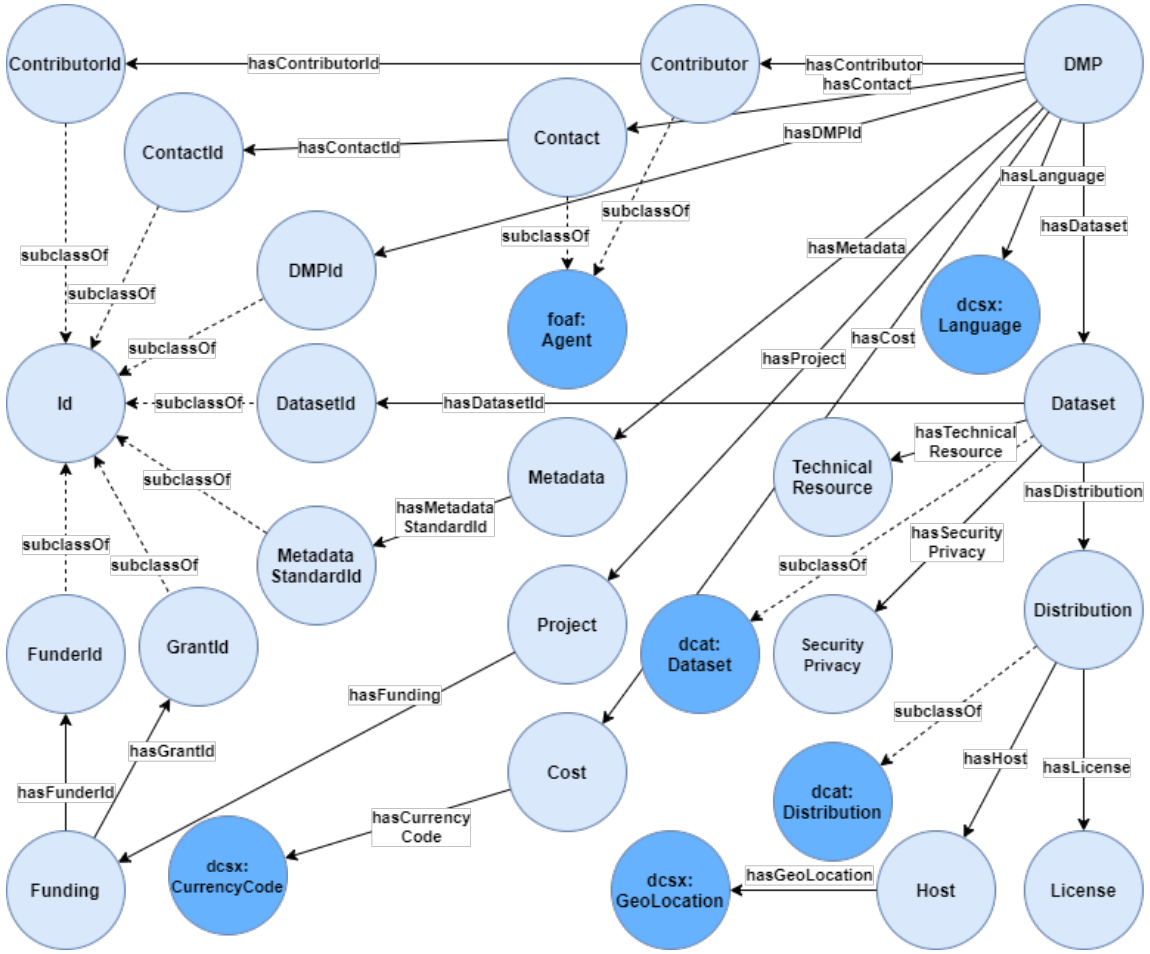
- João Cardoso
- Fajar J. Ekaputra
- Marie Christine Jacquemot
- Leyla Garcia

Contributors

- Daniel Faria

Acknowledgments

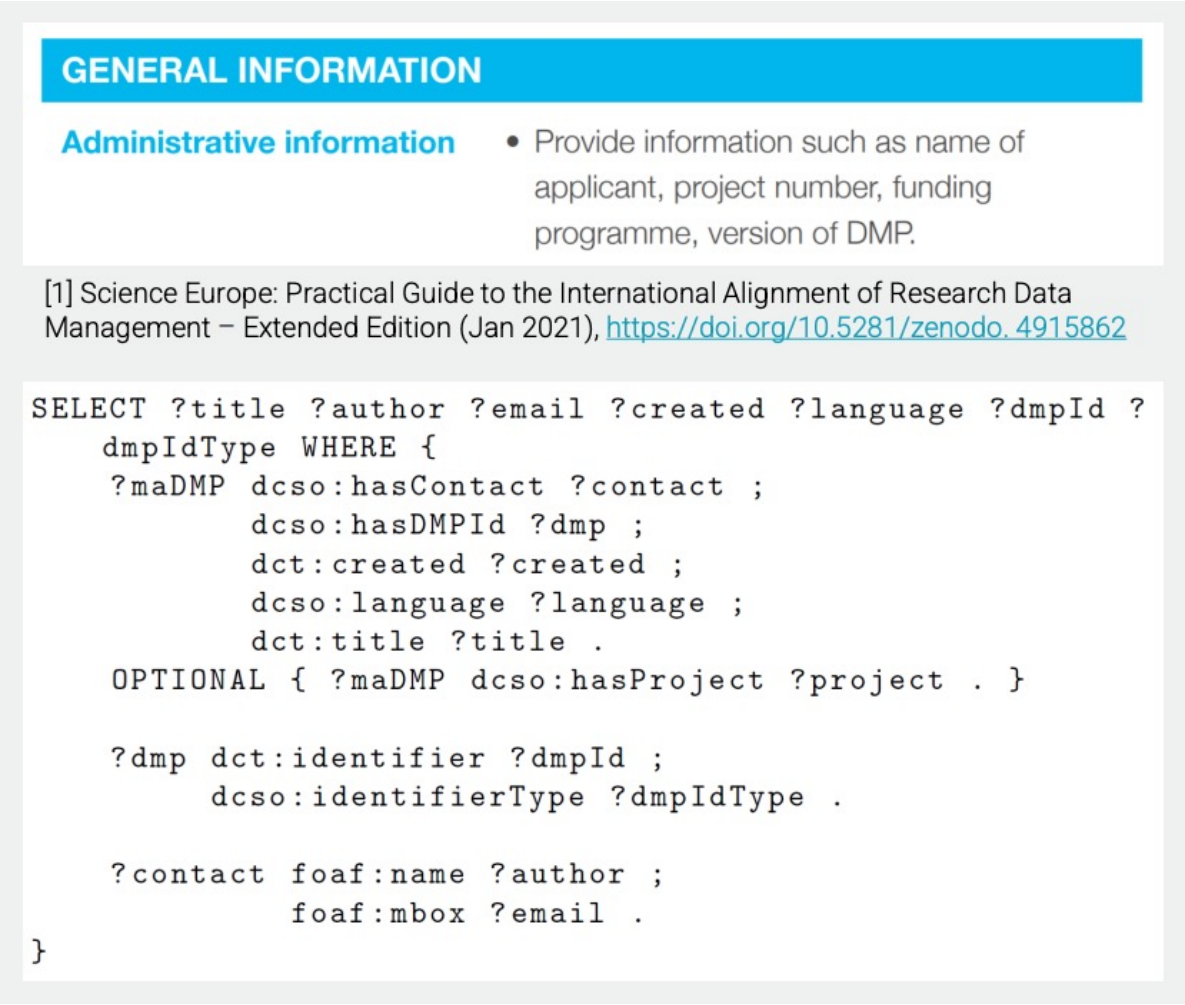
- Tomasz Miksa
- Marek Suchánek



https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard

Evaluation - SPARQL

- Researchers from TU Wien used DCSO representation to evaluate against funder DMP rubric
- 28/48 subitems largely covered (58%)
- Measured completeness and ‘Satisfaction Value’ for rubric elements
- Semi-automated process helps with evaluation, but could not replace manual evaluation



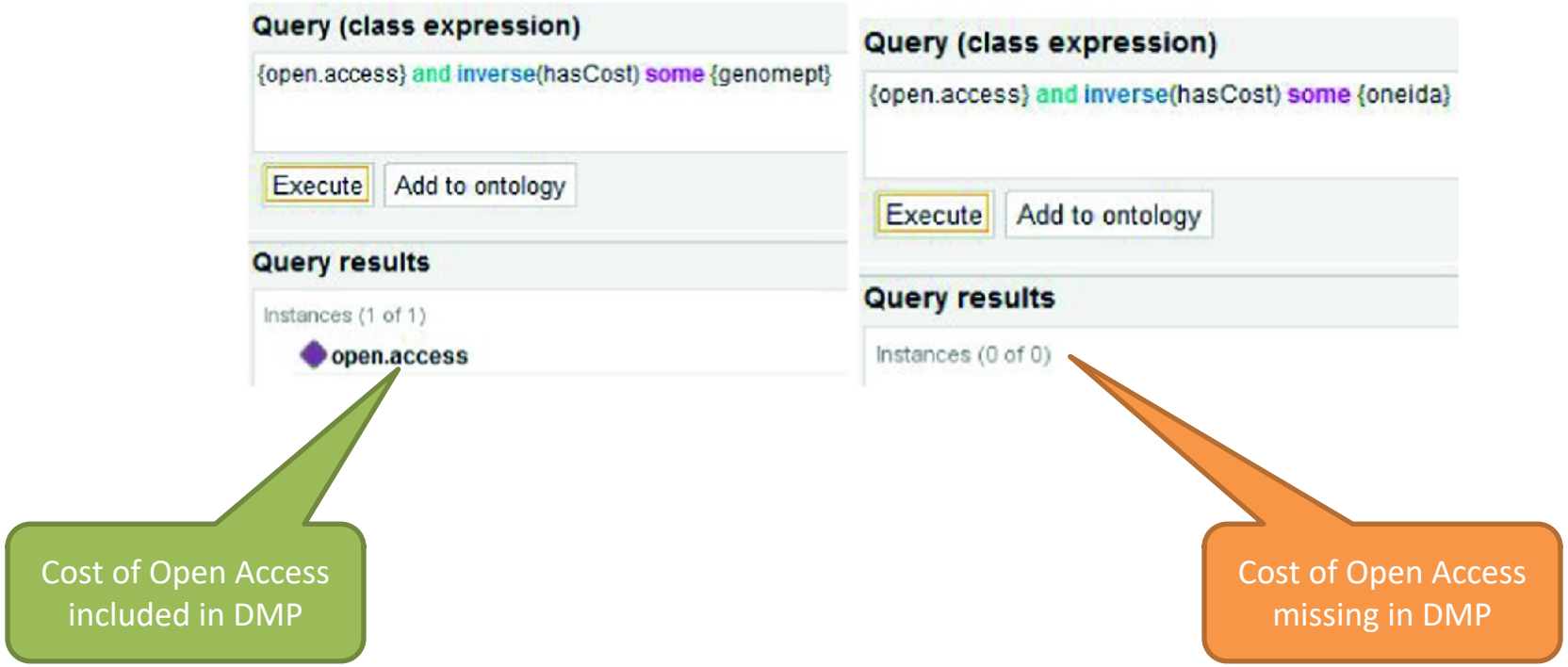
[1] Science Europe: Practical Guide to the International Alignment of Research Data Management – Extended Edition (Jan 2021), https://doi.org/10.5281/zenodo. 4915862

```
SELECT ?title ?author ?email ?created ?language ?dmpId ?
    dmpIdType WHERE {
    ?maDMP dcso:hasContact ?contact ;
           dcso:hasDMPId ?dmp ;
           dct:created ?created ;
           dcso:language ?language ;
           dct:title ?title .
    OPTIONAL { ?maDMP dcso:hasProject ?project . }

    ?dmp dct:identifier ?dmpId ;
         dcso:identifierType ?dmpIdType .

    ?contact foaf:name ?author ;
             foaf:mbox ?email .
}
```

Foidl, R., Brugger, L., Miksa, T.: Automating Evaluation of Machine-Actionable Data Management Plans with Semantic Web Technologies. In DaMaLOS – Second Workshop on Data and Research Objects Management for Linked Open Science: Co-located at the International Semantic Web Conference ISWC 2021. PUBLISSO (Nov 2021), https://doi.org/10.4126/FRL01-006429413

Evaluation - Reasoning

Cardoso et al. used Description Logic (DL) queries to check validity of DMP

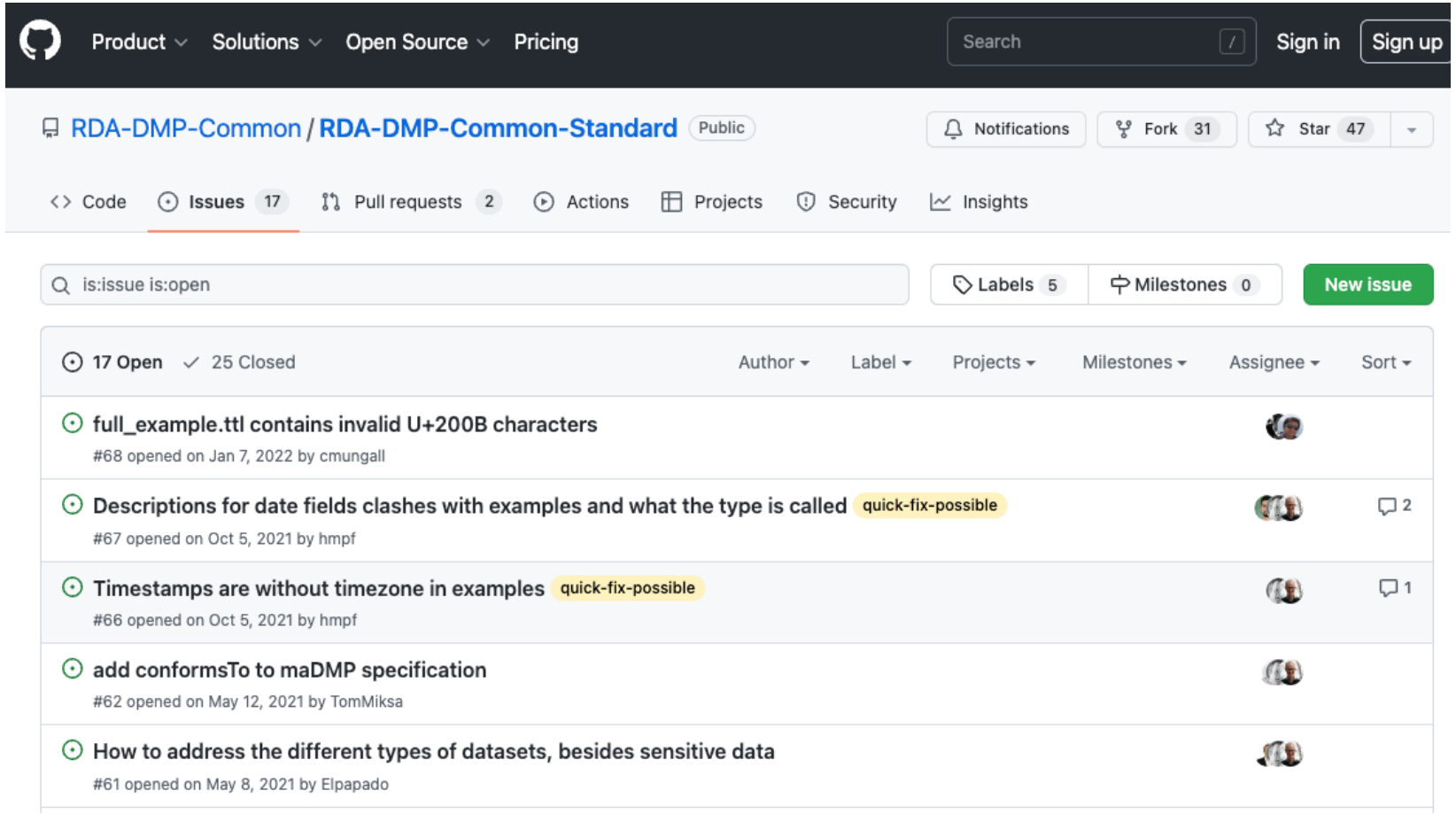


Cardoso, J., Proença, D., Borbinha, J. (2020). Machine-Actionable Data Management Plans: A Knowledge Retrieval Approach to Automate the Assessment of Funders' Requirements. In: , *et al.* Advances in Information Retrieval. ECIR 2020. Lecture Notes in Computer Science(), vol 12036. Springer, Cham. https://doi.org/10.1007/978-3-030-45442-5_15

Maintenance of Standard

Issues tracked through GitHub

https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard/issues/



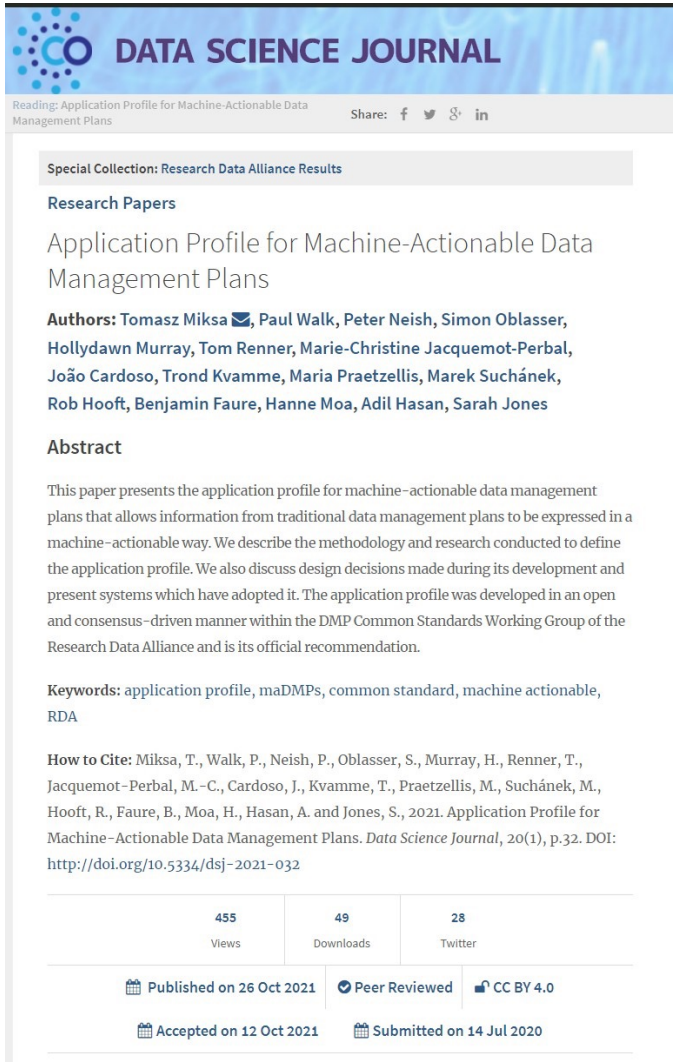
Read more in…

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



DATA SCIENCE JOURNAL

Reading: Application Profile for Machine-Actionable Data Management Plans

Share:

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

455 Views | 49 Downloads | 28 Twitter

Published on 26 Oct 2021 | Peer Reviewed | CC BY 4.0

Accepted on 12 Oct 2021 | Submitted on 14 Jul 2020

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

Find out more

DMP Common Standards Working Group
https://www.rd-alliance.org/groups/dmp-common-standards-wg

Active DMP Interest Group
https://www.rd-alliance.org/groups/active-data-management-plans.html

All RDA Groups (101 current groups)
https://www.rd-alliance.org/groups

Next Plenary 20-24th March 2023
https://www.rd-alliance.org/rdas-20th-plenary-programme

DMP Workshop 20th 7:30pm – 11pm AEDT
https://www.dcc.ac.uk/events/RDAcolocated_machine_actionable_DMPs