

Machine-Actionable Data Management Plans

Feb. 28

Data Management Plans Interest Group #1, 2023

We are an Australasian interest group for people interested in discussing, building and improving Data Management Plans (DMPs).



By Australian Research Data Commons

485 followers

Follow

When and where



Date and time

Tue., 28 February 2023,
2:00 pm – 3:00 pm AEST



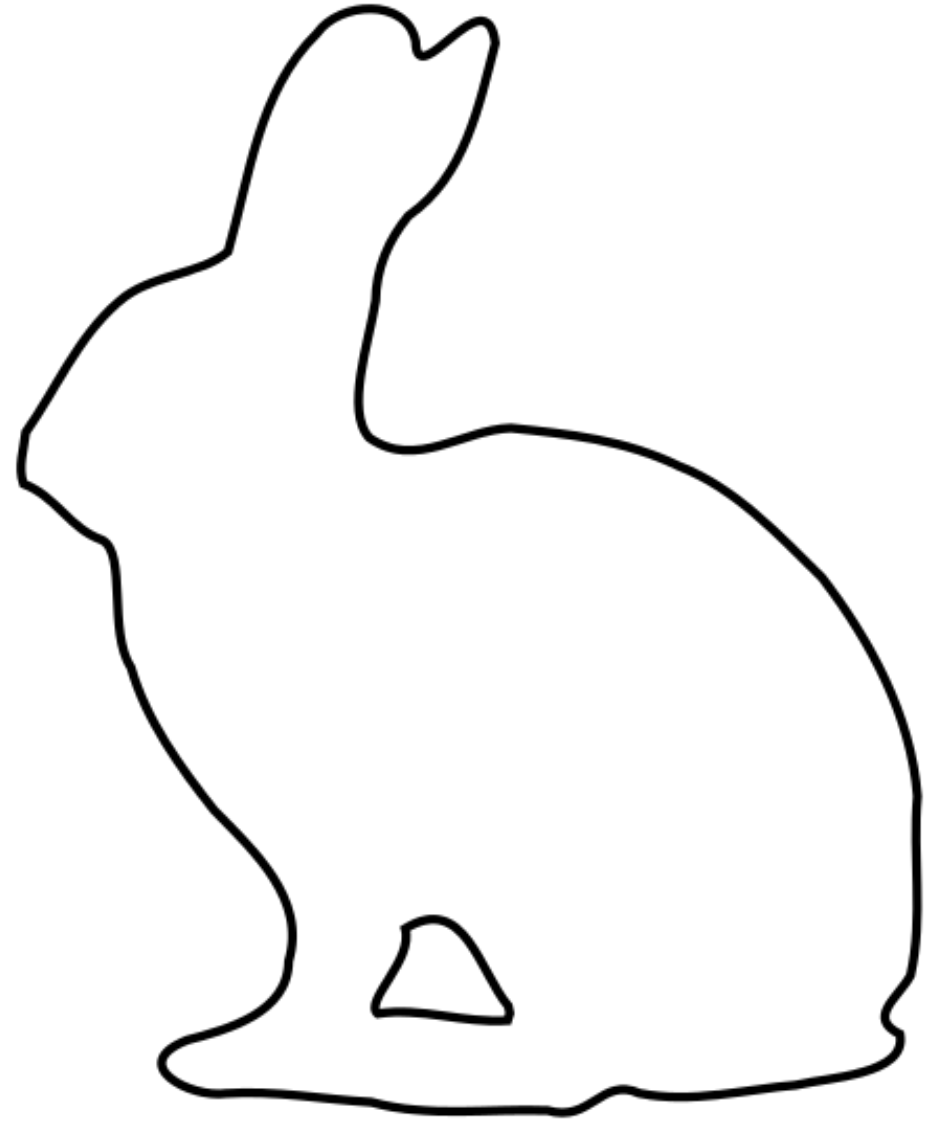
Location

Online

[Access link](#)

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*

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

400+ members!

DMP Common Standards *Working Group*

Maintenance mode

Updates the recommendation when necessary

Can add new serializations

JSON, RDF, ...

Supports adopters of the recommendation

240+ members!

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	
Foreword by Dr Thierry Damerval	2
Introduction	4
GUIDANCE FOR ORGANISATIONS: CORE REQUIREMENTS FOR DATA MANAGEMENT PLANS	7
GUIDANCE FOR ORGANISATIONS: CRITERIA FOR THE SELECTION OF TRUSTWORTHY REPOSITORIES	11
GUIDANCE FOR RESEARCHERS: Translating the Core Requirements into a DMP template Guiding the Selection of Trustworthy Repositories	15
GUIDANCE FOR REVIEWERS: Evaluation Rubric for Data Management Plans	31
Notes and References	51
Annex: Compatibility with the FAIR Data Principles	52

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>



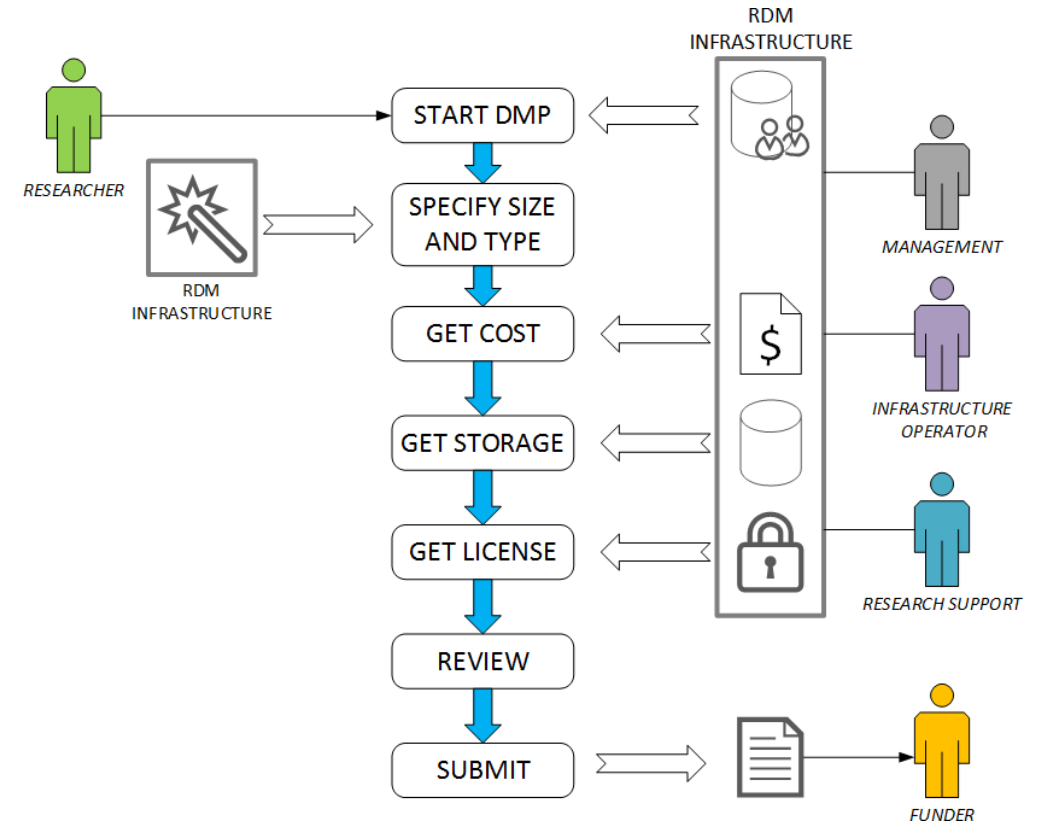
Machine-actionable DMPs (maDMPs)

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

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.

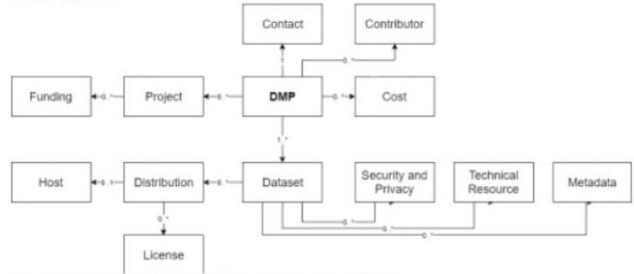
DMP - Provides high level information about the DMP, e.g. its title, modification date, etc. It is the root of this application profile.

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 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



```
graph TD
    Funding --> Project
    Project --> DMP
    DMP --> Contact
    DMP --> Contributor
    DMP --> Cost
    DMP --> Dataset
    Dataset --> Host
    Dataset --> Distribution
    Distribution --> License
    Dataset --> Security[Security and Privacy]
    Dataset --> Technical[Technical Resource]
    Dataset --> Metadata
```

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

1

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>

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"
```

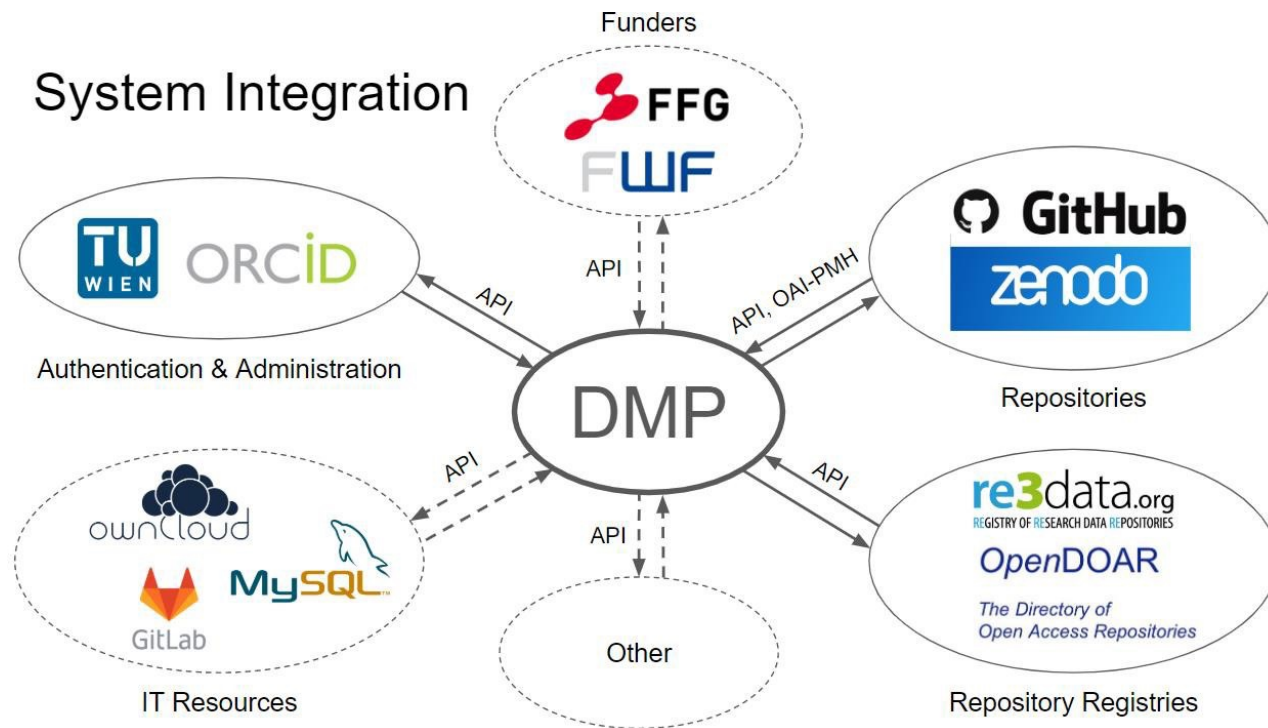
Each **dataset** has a **title** and a human readable **description**.

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

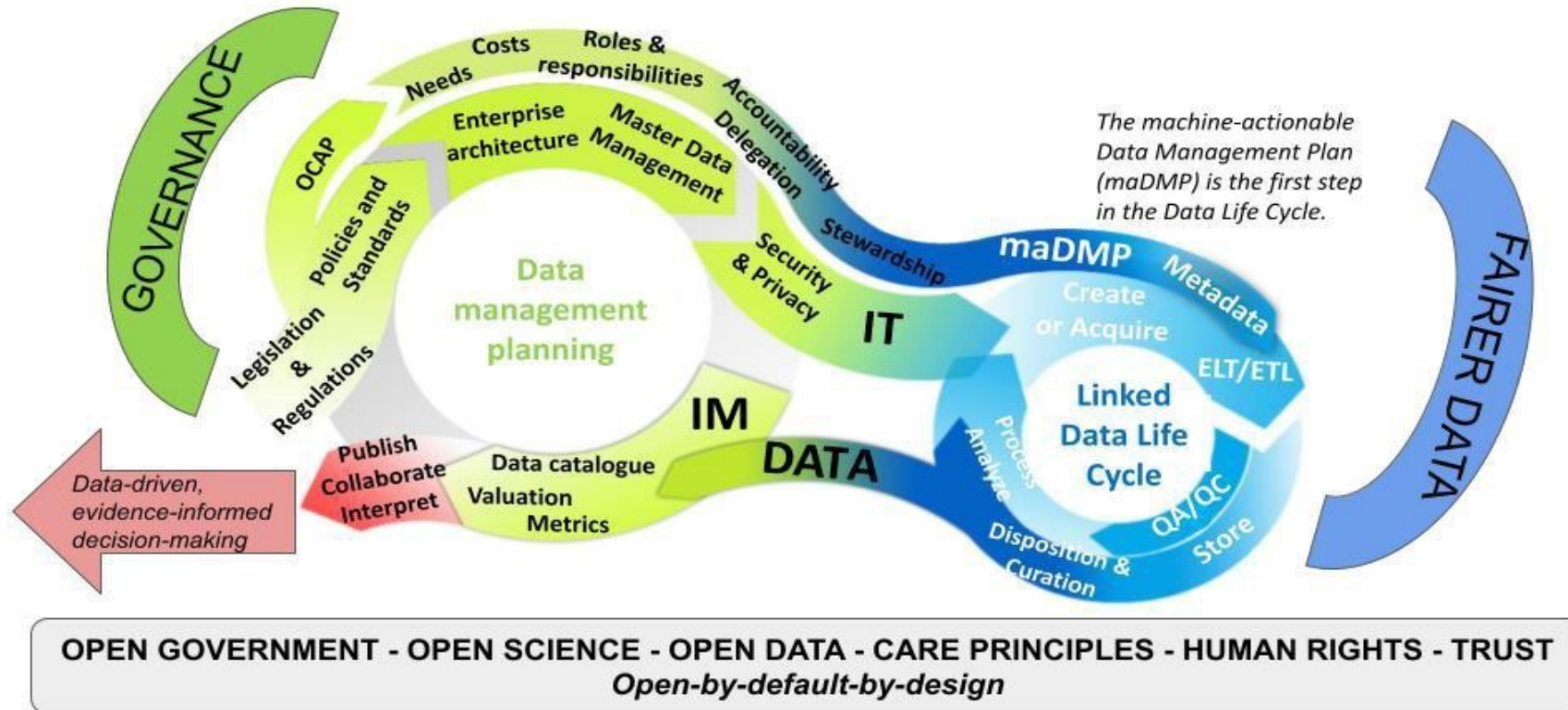
- maDMPs are the 'glue' between different systems
 - Automate getting information **in** and **out**



Adoptions (selected)



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



Findable, Accessible, Interoperable,
Reusable, Ethical, Reproducible

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

GENERAL INFORMATION

Administrative information

- Provide information such as name of applicant, project number, funding programme, version of DMP.

[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 .
}
```

Evaluation - Reasoning

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

The image displays two side-by-side screenshots of a query interface, likely from a software tool used for evaluating Data Management Plans (DMPs) using Description Logic (DL) queries.

Left Interface:

- Query (class expression):** `{open.access} and inverse(hasCost) some {genomept}`
- Buttons:** "Execute" and "Add to ontology"
- Query results:** Instances (1 of 1)
 - open.access** (indicated by a purple diamond icon)

A green callout box points to the "open.access" result, stating: "Cost of Open Access included in DMP".

Right Interface:

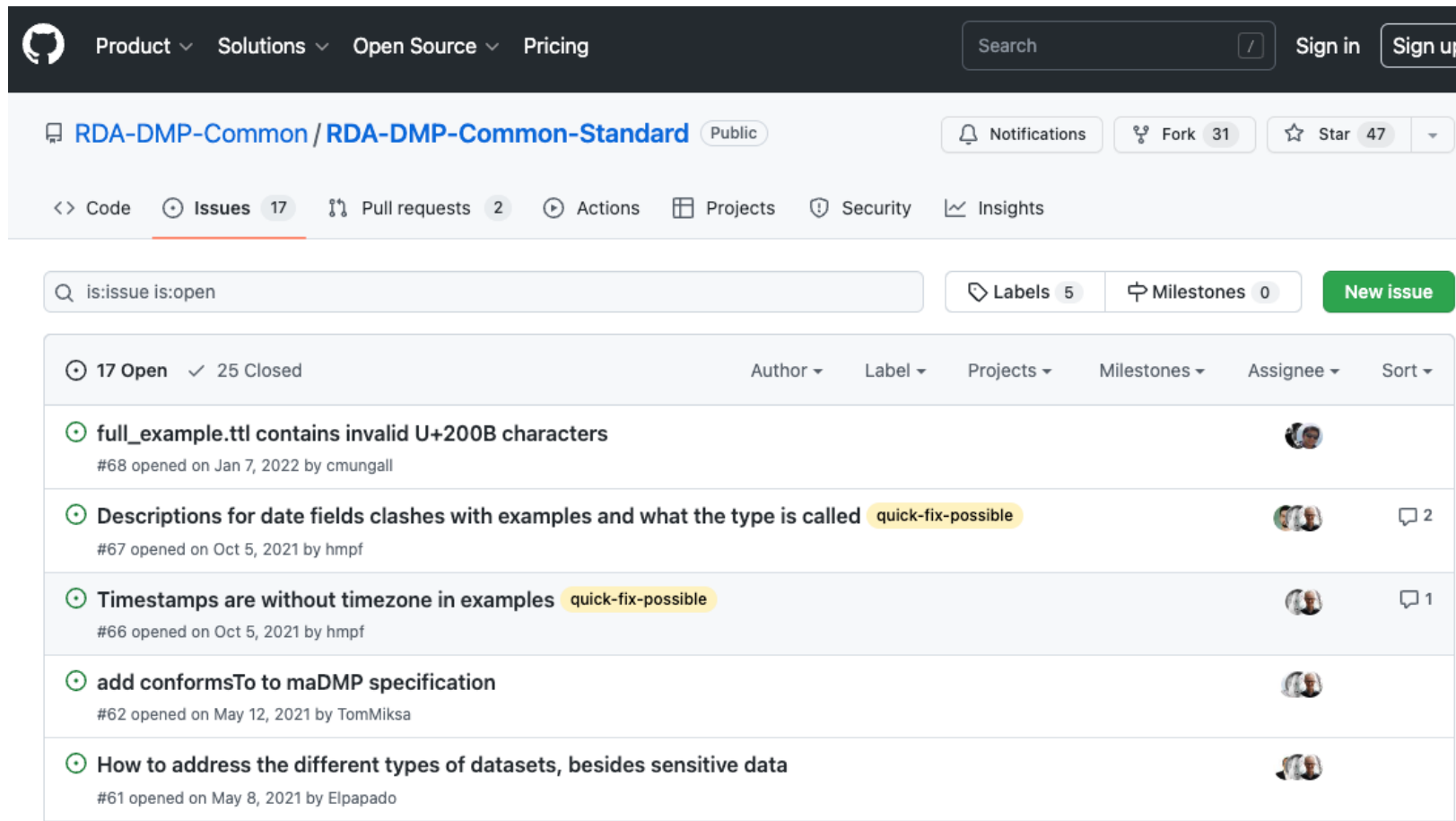
- Query (class expression):** `{open.access} and inverse(hasCost) some {oneida}`
- Buttons:** "Execute" and "Add to ontology"
- Query results:** Instances (0 of 0)

An orange callout box points to the "Instances (0 of 0)" result, stating: "Cost of Open Access missing in DMP".

Maintenance of Standard

Issues tracked through GitHub

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



The screenshot shows the GitHub interface for the repository `RDA-DMP-Common / RDA-DMP-Common-Standard`. The repository is public and has 31 forks and 47 stars. The `Issues` tab is selected, showing 17 open issues. The issues are listed in a table with columns for Author, Label, Projects, Milestones, Assignee, and Sort. The first five issues are visible:

Issue Title	Label	Assignee	Comments
<code>full_example.ttl</code> contains invalid U+200B characters #68 opened on Jan 7, 2022 by cmungall		cmungall	0
Descriptions for date fields clashes with examples and what the type is called #67 opened on Oct 5, 2021 by hmpf	quick-fix-possible	hmpf	2
Timestamps are without timezone in examples #66 opened on Oct 5, 2021 by hmpf	quick-fix-possible	hmpf	1
add conformsTo to maDMP specification #62 opened on May 12, 2021 by TomMiksa		TomMiksa	0
How to address the different types of datasets, besides sensitive data #61 opened on May 8, 2021 by Elpapado		Elpapado	0

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: [f](#) [t](#) [g](#) [in](#)

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 Praetzelis, 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., Praetzelis, 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



O&A Members
71
Active Organisational & Affiliate members

MEMBERSHIP
Members: 13303
Becoming a member of RDA is simple and open to both individuals and organizations
[Register now](#)

RDA Groups
WG & IGs: 101
Discover what RDA Working and Interest Groups and all other Groups are up to and find out how to join them. [Explore Groups](#)

[ABOUT RDA](#) [GET INVOLVED](#) [GROUPS](#) [RECOMMENDATIONS](#) [RDA FOR DISCIPLINES](#) [PLENARIES & EVENTS](#) [NEWS & MEDIA](#) [SEARCH](#)

Interest Groups

[Home](#) » [Groups](#) » [Interest Groups](#)

Group Title	Chair(s)
Active Data Management Plans IG	David Giarretta, Kevin Ashley, Tomasz Miksa, Elli Papadopoulou, Maria Praetzelis
Archives and Records Professionals for Research Data IG	Rebecca Grant, Laura Molloy, Sarah Ramdeen, Hea Lim Rhee
Biodiversity Data Integration IG	Wouter Addink, Hamish Holewa, Sridhar Gutam, Libby Ellwood
Chemistry Research Data IG	Leah McEwen, Stuart Chalk, Ian Bruno
CODATA/RDA Research Data Science Schools for Low and Middle Income Countries	Hugh Shanahan, Bianca Peterson, Raphael C��be, Sara El jadid
Data Conservation IG	Denise Hills, Stephen Diggs
Data Discovery Paradigms IG	Mingfang Wu, Fotis Psomopoulos, Kathleen Gregory
Data Economics IG	Yuri Demchenko, Jane Greenberg
Data for Development IG	Ingvill Constanze ��degaard, Norman Mukasa, Robert Sentamu, Mahadia Tunga, Agapiti Manday
Data policy standardisation and implementation IG	Iain Hrynaskiewicz, Natasha Simons, Simon Goudie, Rebecca Grant
Data Versioning IG	Jens Klump, Lesley Wyborn, Mingfang Wu, Kirsten Elger
Digital Practices in History and Ethnography IG	Kim Fortun, Mike Fortun, Lindsay Poirier
Domain Repositories IG	Kerstin Lehnert, Peter Doorn
Early Career and Engagement IG	Devan Ray Donaldson, Fotis Psomopoulos, Elli Papadopoulou, Maria Tsagionoulou

Creating and Managing RDA Groups

Creating or Joining an RDA Working Group

WG Case Statement Process

Creating or Joining a Community of Practice

Creating or Joining an RDA Interest Group

Group Chairs: Roles and Responsibilities

Birds of a Feather

Find a Group

All Groups

Working Groups

Interest Groups

Historical Groups

Coordination Groups

National Groups

Communities of Practice