Machine-actionable Data Management Plans

Tomasz Miksa, RDA Austria
tmiksa@sba-research.org

DMPs currently

- 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**	
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?*

https://www.fwf.ac.at/fileadmin/files/Dokumente/Open_Access/FWF_DMPTemplate_e.pdf

Data Management Plans

How to discover these tools?

Which one do I need to use?

Why do I have to provide the same information again?

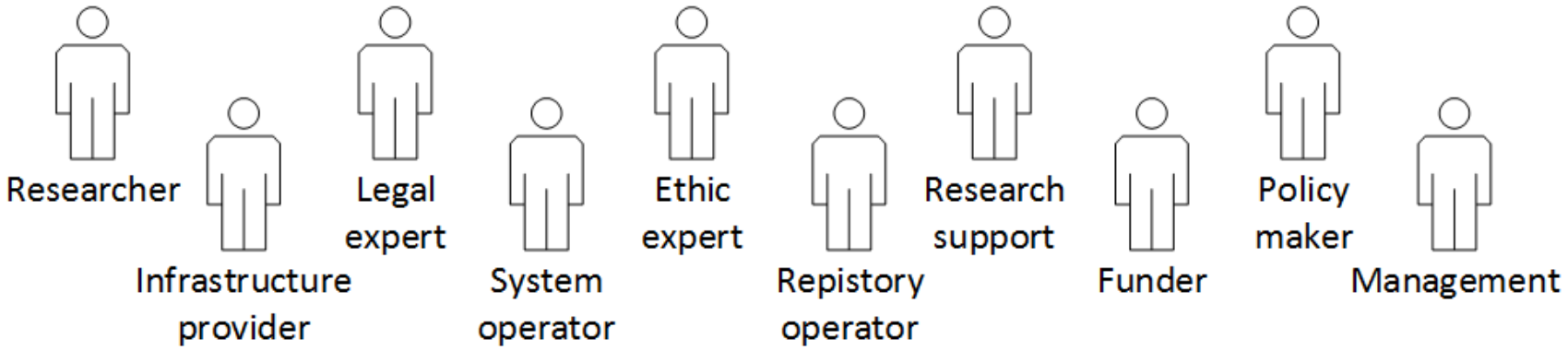
Why haven't they consulted us before?

Who is going to pay for this?

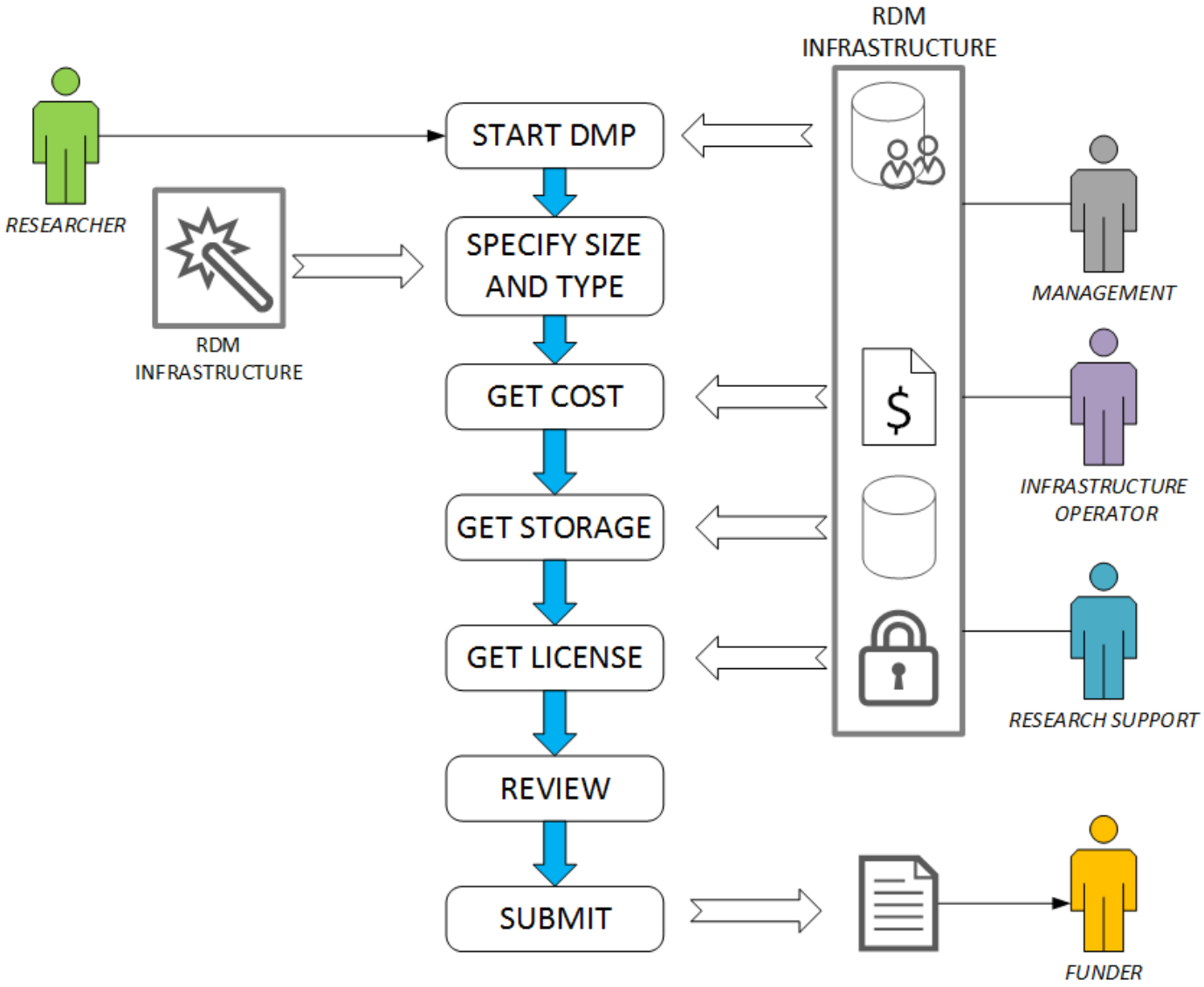
We don't have enough people for that!

Research data lifecycle

- Stakeholders involved in research data management
  - require information at certain stages
  - can provide information if requested at a proper stage
- Many problems can be avoided when
  - timing is right
  - information flow is ensured



Automated Data Management Workflow



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 data model

DMP Common Standards - Outputs

- **Common standard for machine-actionable DMPs**
  - to model information from standard DMPs
  - NOT a template
  - NOT a questionnaire
  - modular design
    - core set of elements
    - domain specific extensions
- **Reference implementations**
  - ready to use models
    - JSON, RDF, etc.

Example

- Current DMPs – model questionnaires

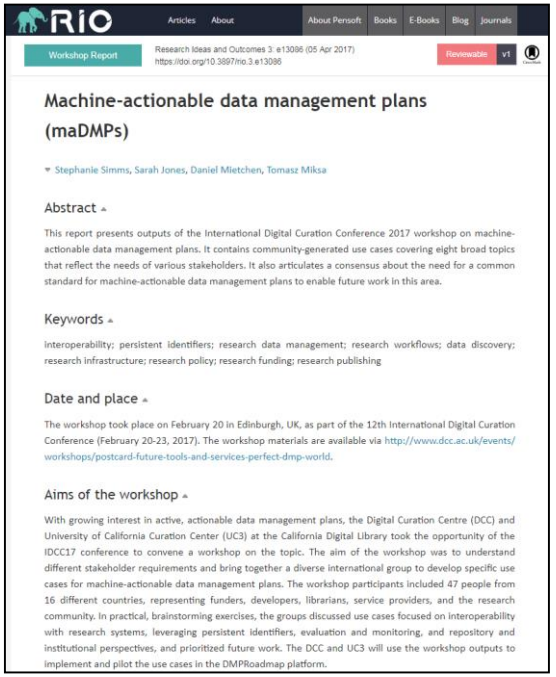
```
<administrative_data>
    <question>Who will be the Principle Investigator?</question>
    <answer>The PI will be John Smith from our university.</answer>
</administrative_data>
```

- Machine-actionable DMPs – model information

```
"dc:creator":[ {
            "foaf:name":"John Smith",
            "@id":"orcid.org/0000-1111-2222-3333",
            "foaf:mbox":"mailto:jsmith@tuwien.ac.at",
            "madmp:institution":" AT-Vienna-University-of-Technology"
 } ],
```

DMP Common Standards WG

- Launched in October 2017
- Result of a consultation made by Active DMPs IG
- Focus on machine-actionable DMPs
- 200 members from all continents
- DMP tool owners are part of it


RiO

Workshop Report

Research Ideas and Outcomes 3: e13086 (05 Apr 2017)
https://doi.org/10.3897/rio.3.e13086

Machine-actionable data management plans (maDMPs)

Stephanie Simms, Sarah Jones, Daniel Mietchen, Tomasz Miksa

Abstract

This report presents outputs of the International Digital Curation Conference 2017 workshop on machine-actionable data management plans. It contains community-generated use cases covering eight broad topics that reflect the needs of various stakeholders. It also articulates a consensus about the need for a common standard for machine-actionable data management plans to enable future work in this area.

Keywords

interoperability; persistent identifiers; research data management; research workflows; data discovery; research infrastructure; research policy; research funding; research publishing

Date and place

The workshop took place on February 20 in Edinburgh, UK, as part of the 12th International Digital Curation Conference (February 20-23, 2017). The workshop materials are available via http://www.dcc.ac.uk/events/workshops/postcard-future-tools-and-services-perfect-dmp-world.

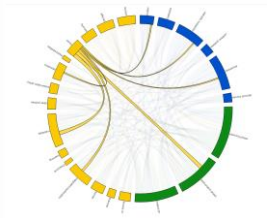
Aims of the workshop

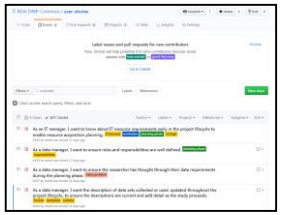
With growing interest in active, actionable data management plans, the Digital Curation Centre (DCC) and University of California Curation Center (UC3) at the California Digital Library took the opportunity of the IDCC17 conference to convene a workshop on the topic. The aim of the workshop was to understand different stakeholder requirements and bring together a diverse international group to develop specific use cases for machine-actionable data management plans. The workshop participants included 47 people from 16 different countries, representing funders, developers, librarians, service providers, and the research community. In practical, brainstorming exercises, the groups discussed use cases focused on interoperability with research systems, leveraging persistent identifiers, evaluation and monitoring, and repository and institutional perspectives, and prioritized future work. The DCC and UC3 will use the workshop outputs to implement and pilot the use cases in the DMPRoadmap platform.

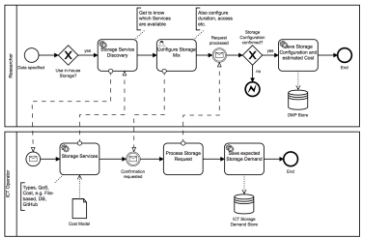
https://doi.org/10.3897/rio.3.e13086

Summary of actions till now

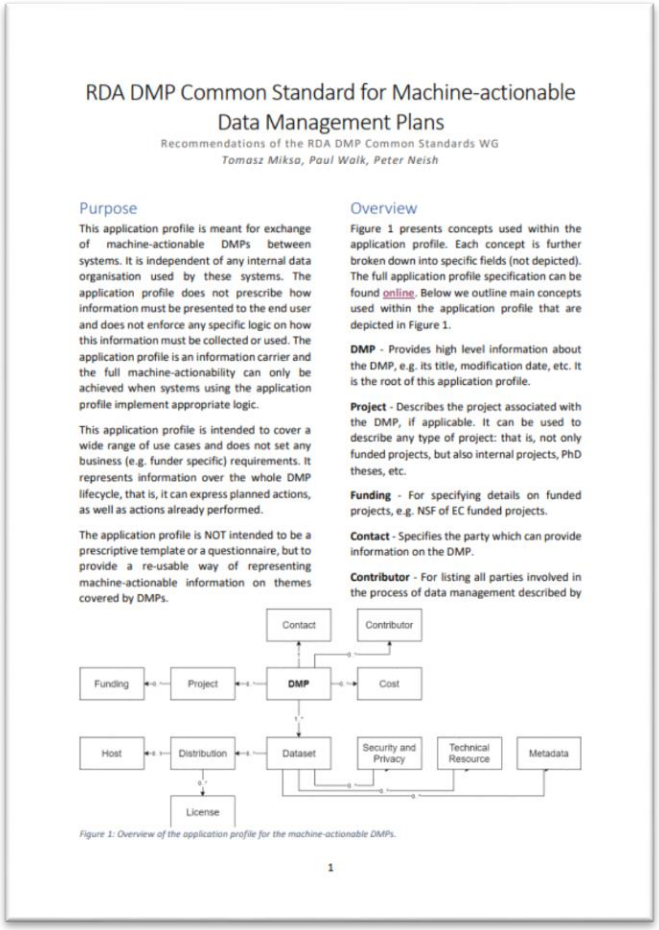
- 1st consultation (user stories) went broad
  - to define scope of maDMPs
- 2nd consultation went deep
  - to identify models for specific requirements
- Proof of concept tools
  - to demonstrate how model can be used to automate tasks
- BPMN processes
  - to identify systems and stakeholders involved
- Model development







Official RDA output



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

Model - documentation

Properties in 'contact'

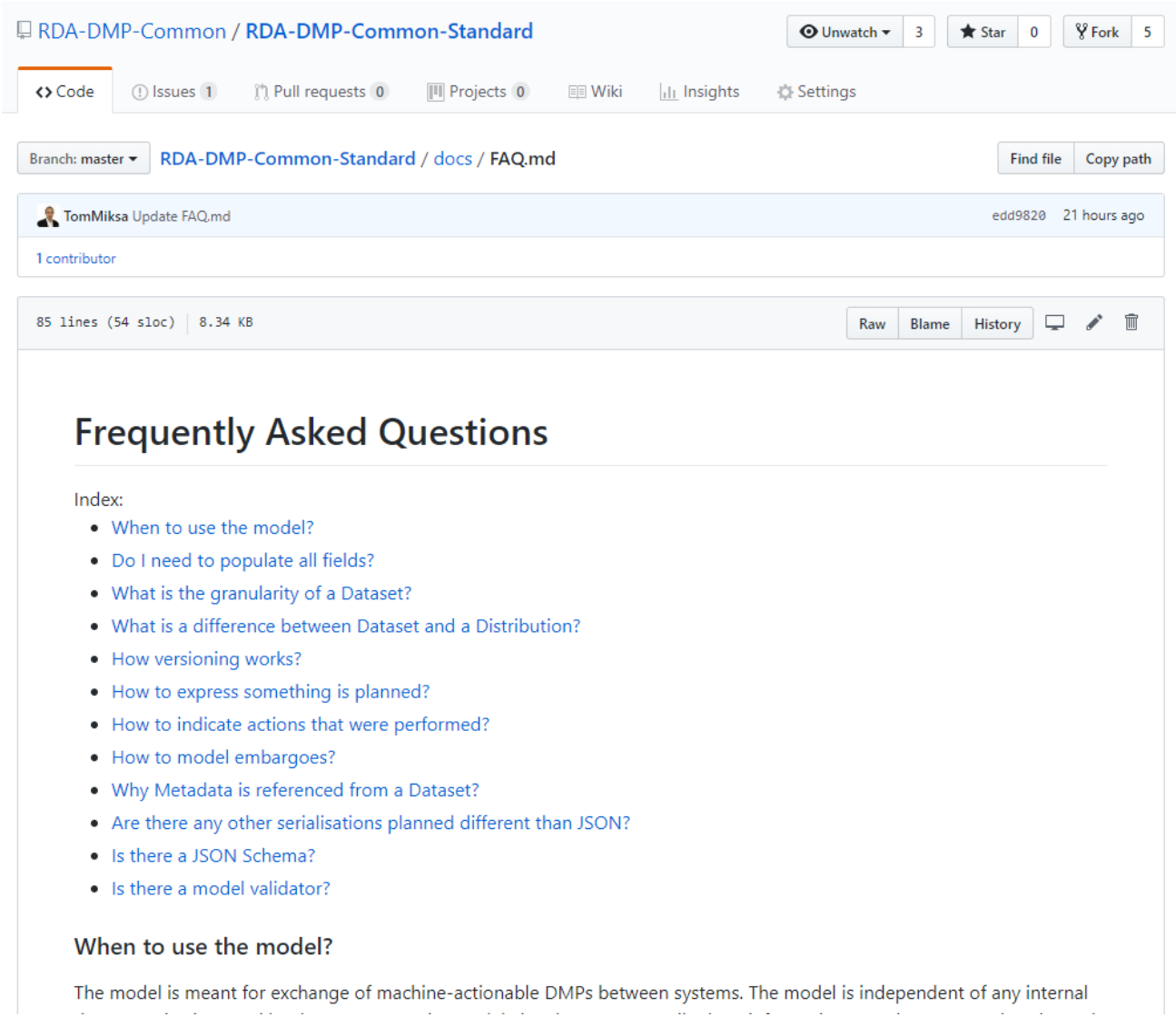
Name	Description	Data Type	Cardinality	Example Value
contact_id	Identifier for a contact person	String	Exactly One	http://orcid.org/0000-0000-0000-0000
mail	E-mail address	String	Exactly One	cc@example.com
name	Name of the contact person	String	Exactly One	Charlie Chaplin

Properties in 'cost'

Name	Description	Data Type	Cardinality	Example Value
currency_code	Allowed values defined by ISO 4217.	Term from Controlled Vocabulary	Zero or One	EUR
description	Description	String	Zero or One	Costs for maintaining....
title	Title	String	Exactly One	Storage and backup
type	Type	Term from Controlled Vocabulary	Zero or One	
value	Value	Number	Zero or One	1000

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

Model – FAQ



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

Unwatch 3 | Star 0 | Fork 5

Code | Issues 1 | Pull requests 0 | Projects 0 | Wiki | Insights | Settings

Branch: master ▾ RDA-DMP-Common-Standard / docs / FAQ.md — Find file | Copy path

TomMiksa Update FAQ.md — edd9820 21 hours ago

1 contributor

85 lines (54 sloc) | 8.34 KB — Raw | Blame | History

Frequently Asked Questions

Index:

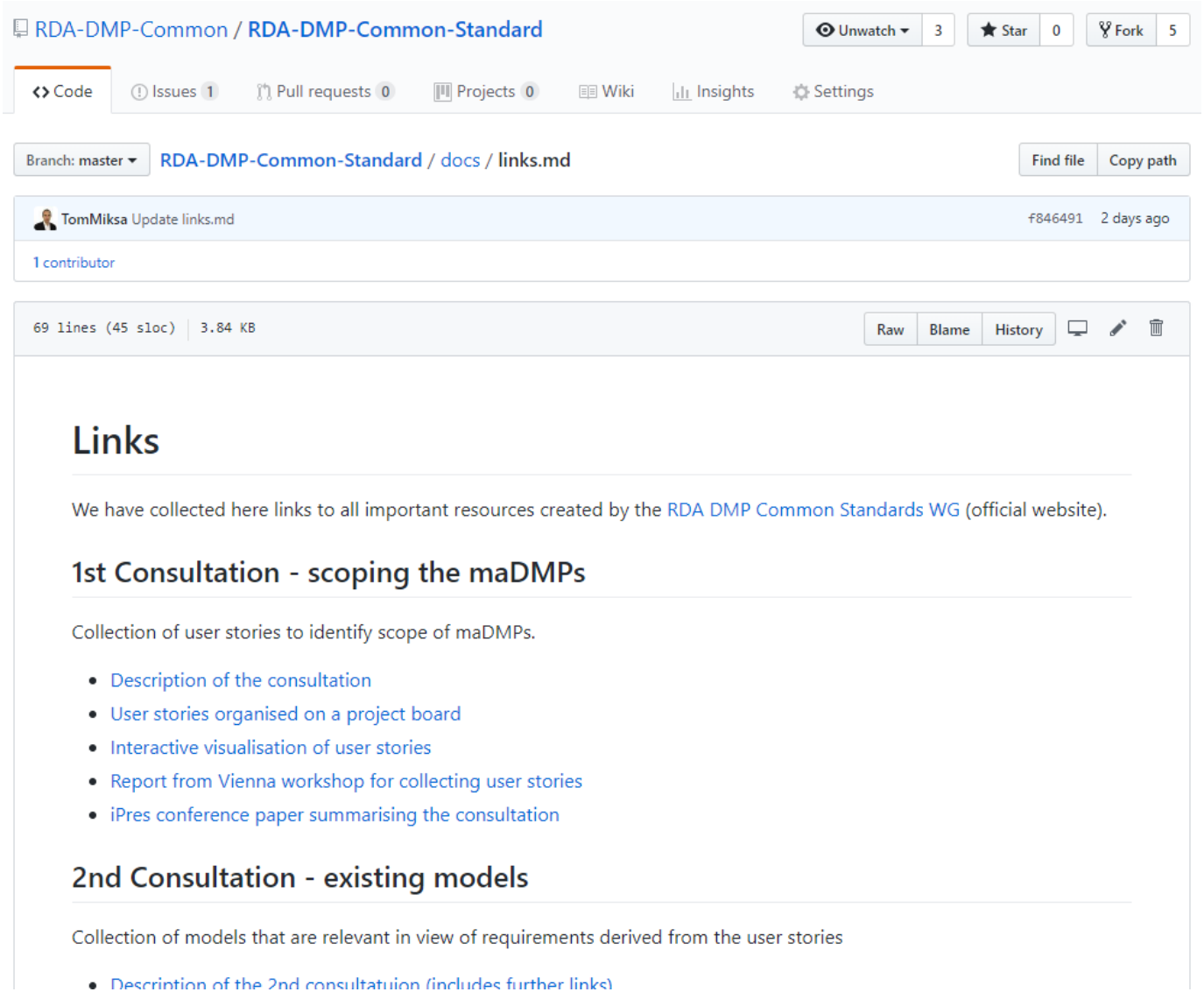
- When to use the model?
- Do I need to populate all fields?
- What is the granularity of a Dataset?
- What is a difference between Dataset and a Distribution?
- How versioning works?
- How to express something is planned?
- How to indicate actions that were performed?
- How to model embargoes?
- Why Metadata is referenced from a Dataset?
- Are there any other serialisations planned different than JSON?
- Is there a JSON Schema?
- Is there a model validator?

When to use the model?

The model is meant for exchange of machine-actionable DMPs between systems. The model is independent of any internal

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

Model – useful links



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

Unwatch 3 | Star 0 | Fork 5

Code | Issues 1 | Pull requests 0 | Projects 0 | Wiki | Insights | Settings

Branch: master ▾ | RDA-DMP-Common-Standard / docs / links.md | Find file | Copy path

TomMiksa Update links.md — f846491 — 2 days ago

1 contributor

69 lines (45 sloc) | 3.84 KB | Raw | Blame | History

Links

We have collected here links to all important resources created by the RDA DMP Common Standards WG (official website).

1st Consultation - scoping the maDMPs

Collection of user stories to identify scope of maDMPs.

- Description of the consultation
- User stories organised on a project board
- Interactive visualisation of user stories
- Report from Vienna workshop for collecting user stories
- iPres conference paper summarising the consultation

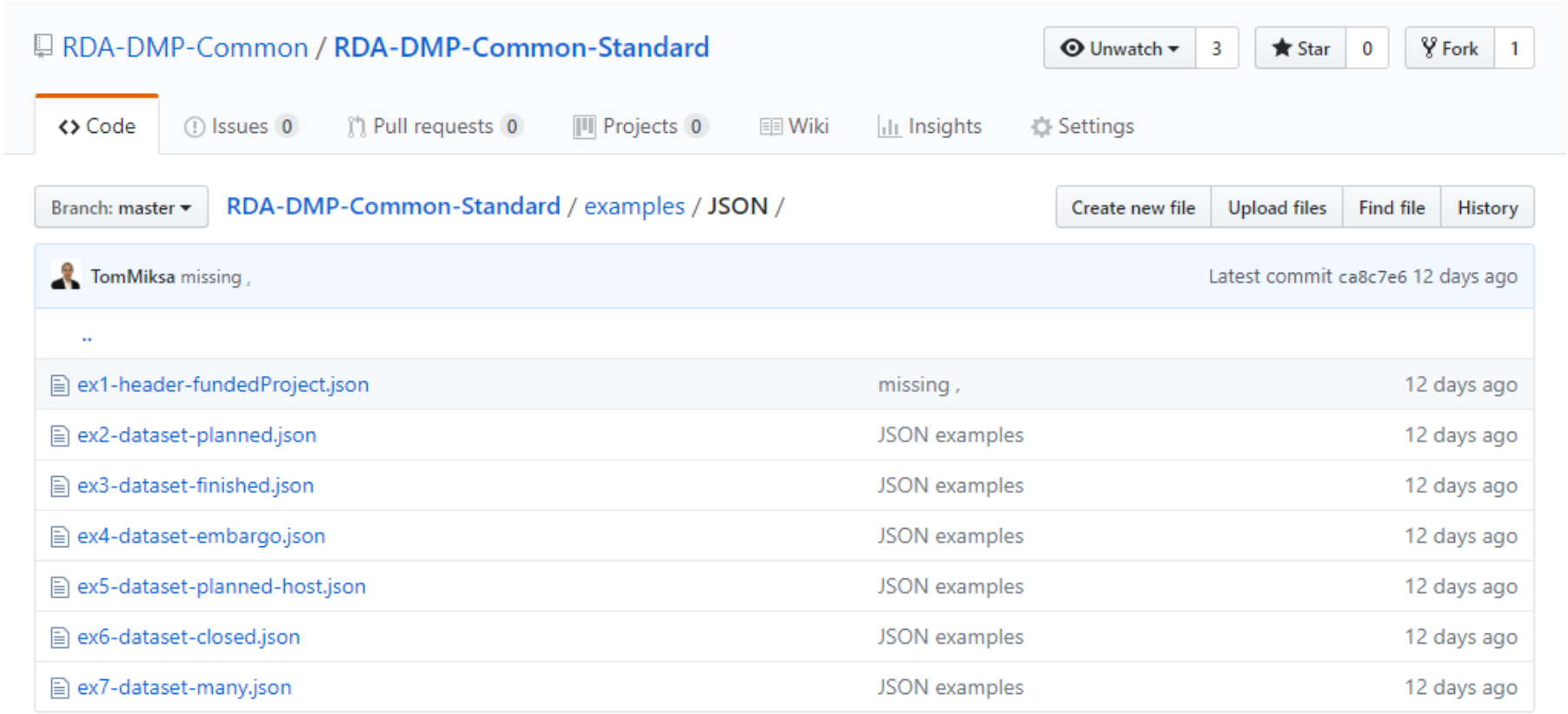
2nd Consultation - existing models

Collection of models that are relevant in view of requirements derived from the user stories

- Description of the 2nd consultatuion (includes further links)

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

Model – JSON examples



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

Unwatch 3 | Star 0 | Fork 1

Code | Issues 0 | Pull requests 0 | Projects 0 | Wiki | Insights | Settings

Branch: master ▾ | RDA-DMP-Common-Standard / examples / JSON / | Create new file | Upload files | Find file | History

TomMiksa missing , — Latest commit ca8c7e6 12 days ago

..		
ex1-header-fundedProject.json	missing ,	12 days ago
ex2-dataset-planned.json	JSON examples	12 days ago
ex3-dataset-finished.json	JSON examples	12 days ago
ex4-dataset-embargo.json	JSON examples	12 days ago
ex5-dataset-planned-host.json	JSON examples	12 days ago
ex6-dataset-closed.json	JSON examples	12 days ago
ex7-dataset-many.json	JSON examples	12 days ago

https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard/tree/master/examples/JSON

Minimal maDMP

32 lines (32 sloc) | 679 Bytes

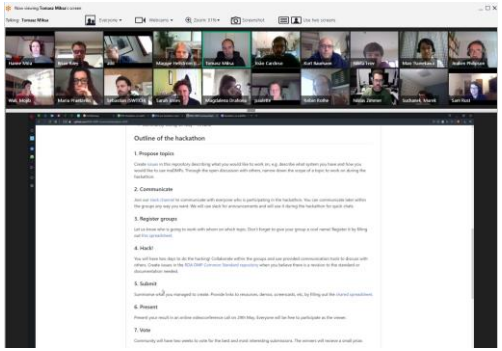
```
{
        "dmp": {
                "title": "Minimal DMP",
                "contact": {
                        "contact_id": {
                                "identifier": "http://orcid.org/0000-0000-0000-0000",
                                "type": "orcid"
                        },
                        "mbox": "cc@example.com",
                        "name": "Charlie Chaplin"
                },
                "created": "2018-07-23T10:10:23.6",
                "dmp_id": {
                        "identifier": "https://doi.org/10.0000/00.0.1234",
                        "type": "doi"
                },
                "dataset": [
                        {
                                "dataset_id": {
                                        "identifier": "https://doi.org/10.0000/00.0.5678",
                                        "type": "doi"
                                },
                                "title": "Placeholder dataset",
                                "personal_data": "unknown",
                                "sensitive_data": "unknown"
                        }
                ],
                "ethical_issues_exist": "unknown",
                "language": "eng",
                "modified": "2019-02-06T15:30:42.1"
        }
}
```

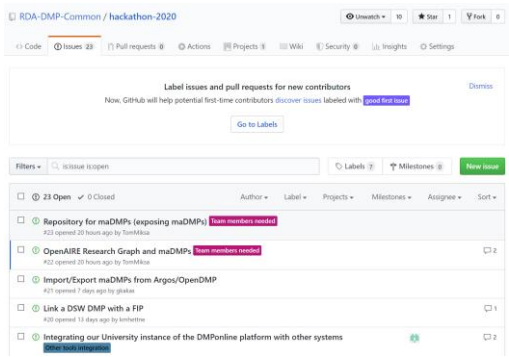
(Pending) adoptions

- **DMP Online** by Digital Curation Centre (DCC) in the UK
- **DMP Tool** by California Digital Library (CDL) in the US
- **DMP OPIDoR** by Centre national de la recherche scientifique (CNRS) in France
- **RDMO** by Leibniz-Institut für Astrophysik Potsdam in Germany
- **Data Stewardship Wizzard** by Elixir research infrastructure in the EU
- **Argos** - OpenDMP by OpenAIRE and EUDAT research infrastructures in the EU
- **F1000Research** open research publisher in the UK
- **NSD DMP** in Norway
- **Haplo** repository in the UK
- TU Wien, TU Graz, Uni Wien via **FAIR Data Austria** project
- ….

RDA Hackathon on maDMPs

- 80 participants, 12 teams
- Starts tomorrow, sign up still possible!
- Participation
  - Join teams!
  - Join grand finale on zoom!
  - Vote!



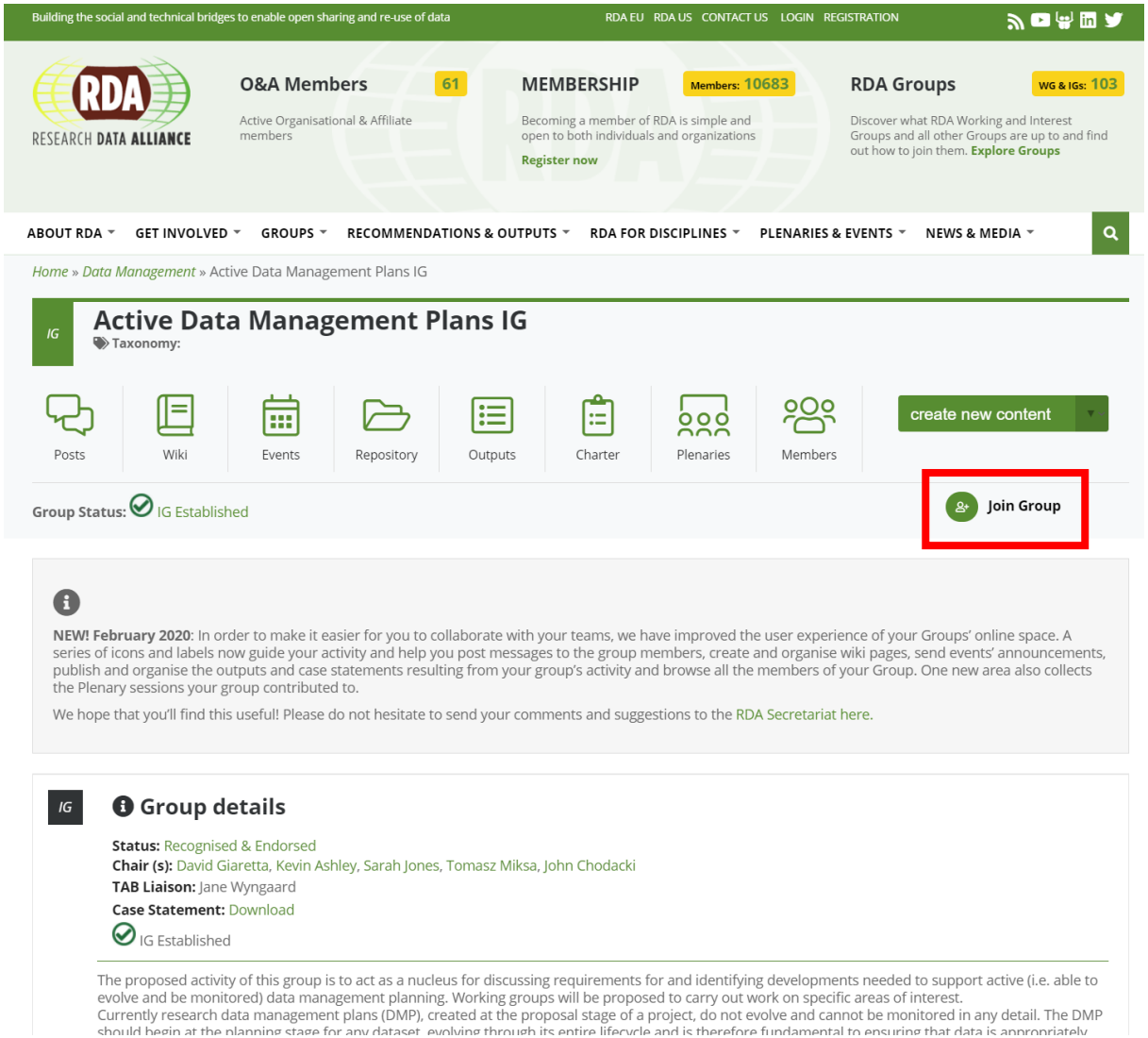


https://github.com/RDA-DMP-Common/hackathon-2020

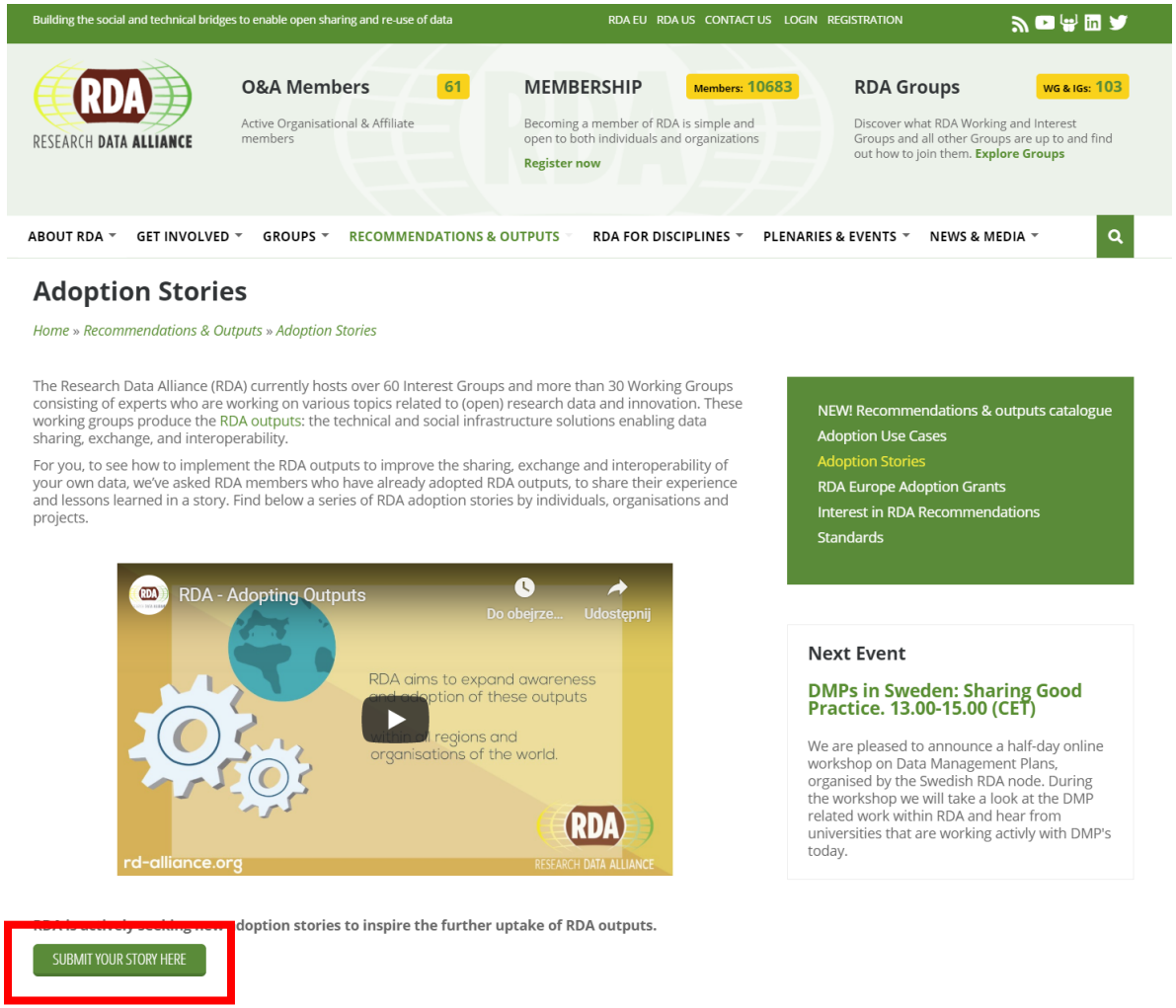
Staying in touch

- Active Data Management Plans IG
  - Umbrella group
  - Place to bring your new ideas!
- Exposing Data Management Plans WG
  - Opening DMPs for public
  - Use cases for sharing DMPs
- DMP Common Standards WG
  - Standard specific discussion
  - Adoptions of the standard

Join groups now!



Adopt RDA recommendations!



https://www.rd-alliance.org/recommendations-outputs/adoption-stories

Publications

- Tomasz Miksa, Peter Neish, Paul Walk, Andreas Rauber: **Defining requirements for machine-actionable Data Management Plans**. iPres 2018

- 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, 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