DMP Common Standards WG at the 13th Plenary in Philadelphia

Tomasz Miksa, Paul Walk, Peter Neish
Agenda

Part 1 – Introduction for newcomers
  › goals of this WG
  › summary of what we did so far

Part 2 – Common model for machine-actionable DMPs
  › Presentation of the model
  › Discussion

Part 3 – Ongoing and future pilot implementations
  › Evaluations and prototypes developed to date
  › New ideas for model adoption
  › Discussion

Part 4 – Wrap up and Future of the WG
  › Timeline
  › Discussion
Collaborative notes

https://www.rd-alliance.org/wg-dmp-common-standards-rda-13th-plenary-meeting
Introduction for newcomers

Part 1
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

START DMP

SPECIFY SIZE AND TYPE

GET COST

GET STORAGE

GET LICENSE

REVIEW

SUBMIT
Why do we need this WG?

- Shortcomings of existing DMPs
  - manually completed, vague, not updated, ...

- 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 data model 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, XML, RDF, etc.

‣ **Guidelines for adoption of the common data model**
  ‣ requirements for supporting systems
  ‣ pilot studies

17/04/2019
www.rd-alliance.org - @resdatall
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"
 }
```
Example

• Currently available – not very useful

<administrative_data>

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

Reuse existing standards, e.g. Dublin Core, PREMIS, etc.

"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"
  }
]
Example

• Currently available – not very useful

<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 DMP

"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"
}

Use PIDs whenever possible, e.g. ORCID
Example

• Currently available – not very useful

<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 DMP

"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"
}
]
Example

• Currently available – not very useful

<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 DMP

"dc:creator":[
  {
    "foaf:name":"John Smith",
    "@id":"orcid.org/0000-0001-111-2222-3333",
    "foaf:mbox":"mailto:jsmith@tuwien.ac.at",
    "madmp:institution":"AT-Vienna-University-of-Technology"
  }
],

Develop own concepts and vocabularies only when needed
Launched in October 2017
Result of a consultation made by Active DMPs IG
Focus on machine-actionable DMPs
100+ members from all continents
DMP tool owners are part of it

https://doi.org/10.3897/rio.3.e13086
Summary of actions till now

› 1\textsuperscript{st} consultation

› 2\textsuperscript{nd} consultation

› Proof of concept tools

› BPMN processes

› Model development
  ‣ Part 3
1st consultation – user stories

Goals

- identify stakeholders at each lifecycle stage
  - define which information they provide
  - define which information they expect

As a <stakeholder>, I want <goal> so that <reason>.

As a researcher, I want to inform repository operator on the amount of data in the planning phase, so that they provide information on costs.

2\textsuperscript{nd} consultation – existing models

\textbf{2\textsuperscript{nd} consultation goes deep}

how do we model specific requirements?

which specific fields are needed?

which models exist?

\begin{itemize}
  \item Format
  \begin{itemize}
    \item Text (r), 20, 30, 40, 50, 60, 70, 80
  \end{itemize}
  \item Volume
    \begin{itemize}
      \item Data use estimate [1, 77, 80, 100]
      \item For specific type of data [102]
      \item Data size real [94]
  \end{itemize}
  \item Provenance [54]
  \item Metadata
    \begin{itemize}
      \item Taxonomy/Classification [14, 111]
      \item Link to metadata of the real data [94, 96]
      \item Link publication to data [95]
      \item Authorship [90]
      \item Multilingual metadata [60]
    \end{itemize}
  \item Rights
    \begin{itemize}
      \item Link to metadata location [19, 20, 30, 35, 60]
      \item Rights [40]
      \item Persistent identifier for data [70]
      \item Link publisher to data [35, 60]
    \end{itemize}

Please help us:
\begin{itemize}
  \item Break down existing requirements into more specific requirements.
  \item Add missing requirements.
  \item Provide examples of existing models, vocabularies, etc. that can be used to model these.
\end{itemize}

Please provide your suggestions below.

- **Models**
  - **Format**: rtf, text, latex
  - **Volume**: data size real [94]
  - **Provenance**: information about the source of the data
  - **Metadata**: includes metadata of the real data [94, 96]
  - **Rights**: persistent identifier for data [70]

- **Requirements**
  - **Quality**: description of quality the dataset and how to assess it
  - **Data Dimensions and units of measurement**: data dimension, data measure

- **Other comments**
  - [Link to comments]

---

\begin{itemize}
  \item [Link to previous consultation]
  \item [Link to user stories]
\end{itemize}
Proof of concept tools

Requirements

- Provide minimum input
- Import as much as possible from existing systems to help in creating maDMPs

Tools available as Docker containers on GitHub

- [https://github.com/TomMiksa/DMPGenerator](https://github.com/TomMiksa/DMPGenerator)
- [https://github.com/TomMiksa/digital_preservation_ex_1_2](https://github.com/TomMiksa/digital_preservation_ex_1_2)
- [https://github.com/TomMiksa/tu-dpue-lab2-ss18](https://github.com/TomMiksa/tu-dpue-lab2-ss18)
- [https://github.com/TomMiksa/DigitalPreservation_2](https://github.com/TomMiksa/DigitalPreservation_2)
- [https://github.com/TomMiksa/digitalpreservation-dmp-generator](https://github.com/TomMiksa/digitalpreservation-dmp-generator)
- [https://github.com/TomMiksa/DMPlanner](https://github.com/TomMiksa/DMPlanner)

Example of a landing page for maDMPs

- [https://oblassers.github.io/fair-data-science/](https://oblassers.github.io/fair-data-science/)
- [https://github.com/oblassers/fair-data-science](https://github.com/oblassers/fair-data-science)
Planning phase

- Goal: get **estimations** and **recommendations** (which are feasible to implement later)
Project and Post-project phases

- Goal: **update** DMP with **real** information by **re-using** (linking) information provided elsewhere

Diagram:

- **BASIC INFO**
  - John Smith

- **ADMINISTRATIVE DATA**

- **GET METADATA**

- **PRESERVATION**
  - 10 YEARS

- **GENERATE DMP**

- **GitHub**

- **OAI-PMH**
Name
Please provide your full name.

- **full name**: Tomasz Miksa
- **orcid**: 0000-0002-4929-7875
- **current employment name**: SBA Research

Resources
Add as many Github repositories or OAI-PMH compliant DOIs as you like.*

- **Zenodo**: Ten Simple Rules For Machine-Actionable Data Management Plans (Preprint)
  - documentation
- **Github**: TomMiksa/DMPlanner
  - software

Preservation Time
Choose how many years the data for each group should be kept.

- **Software**: 10 years
- **Documentation**: 20 years
TUW DMP

A Data Management Plan created using DMPlanner.

Creator
Name: Tomasz Miksa
ORCID: 0000-0002-4929-7875
Current Work: SBA Research

How will you manage copyright and Intellectual Property Rights (IPR) issues?
The software which was created in the course of the project has the license restrictions “MIT License”.

Which data are of long-term value and should be retained, shared, and/or preserved?
In this project especially the documentation, as well as the software has a long-term value and should at least be as long preserved as the targeted preservation time specifies. The targeted preservation time for the documentation is 20 years. The targeted preservation time for the software is 10 years.

What is the long-term preservation plan for the dataset?
One of the main strategies of the long-term preservation plan is the use of public accessible repositories to save the components of the project. The documentation resource “Ten Simple Rules For Machine-Actionable Data Management Plans (Preprint)” is hosted on Zenodo. The software resource “DMPlanner” is hosted on Github.

How will you share the data?
The data will be primarily shared through the public repositories listed above. This way the data is openly accessible and findable, as well as searchable. The data is available at the repositories as of this moment.

Are any restrictions on data sharing required?
The restrictions on data sharing are composed of the used licenses together with the long-term preservation plan. With this in mind the following restrictions for the resources of the project apply. The documentation resource “Ten Simple Rules For Machine-Actionable Data Management Plans (Preprint)” will be hosted on Zenodo for at least 20 years. The software resource “DMPlanner” will be hosted on Github for at least 10 years.

Who will be responsible for data management?
The creator of this data management plan is Tomasz Miksa. Therefore Tomasz Miksa is also the reference person for possible reviews and revisions regarding this data management plan in the future. Unless amended Tomasz Miksa is additionally responsible for the adherence to the plan.
Processes

- Processes help identify
  - **tasks** performed by stakeholders
    - e.g. ICT operator provide costs of storage
  - **systems** needed to be put in place
    - e.g. maDMP repository or costing service
  - **concepts** to be developed or agreed
    - e.g. cost model for storage
Processes

▶ Useful in deploying maDMPs
▶ Allow us to narrow down focus of this WG
  ▶ common model does not contain business logic
    ▶ e.g. cost estimation is done by a service that provides a value
  ▶ common model is an information carrier
  ▶ tools, services, processes make maDMPs *machine-actionable*

http://doi.org/10.5281/zenodo.2607556
Summary of actions till now

› 1\textsuperscript{st} consultation (user stories) went broad
     › to define scope of maDMPs

› 2\textsuperscript{nd} 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
Common model for maDMPs

Part 2
Common model for maDMPs
Common model for maDMPs

Common model for maDMPs

Common model for maDMPs
## Properties in 'contact'

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Data Type</th>
<th>Cardinality</th>
<th>Example Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>contact_id</td>
<td>Identifier for a contact person</td>
<td>String</td>
<td>Exactly One</td>
<td><a href="http://orcid.org/0000-0000-0000-0000">http://orcid.org/0000-0000-0000-0000</a></td>
</tr>
<tr>
<td>mail</td>
<td>E-mail address</td>
<td>String</td>
<td>Exactly One</td>
<td><a href="mailto:cc@example.com">cc@example.com</a></td>
</tr>
<tr>
<td>name</td>
<td>Name of the contact person</td>
<td>String</td>
<td>Exactly One</td>
<td>Charlie Chaplin</td>
</tr>
</tbody>
</table>

## Properties in 'cost'

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

[Source](https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard/blob/master/docs/index.md)
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
Model – useful links

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 consultation (includes further links)

**Model – JSON examples**

<table>
<thead>
<tr>
<th>File Name</th>
<th>Description</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>ex1-header-fundedProject.json</td>
<td>missing</td>
<td>12 days ago</td>
</tr>
<tr>
<td>ex2-dataset-planned.json</td>
<td>JSON examples</td>
<td>12 days ago</td>
</tr>
<tr>
<td>ex3-dataset-finished.json</td>
<td>JSON examples</td>
<td>12 days ago</td>
</tr>
<tr>
<td>ex4-dataset-embargo.json</td>
<td>JSON examples</td>
<td>12 days ago</td>
</tr>
<tr>
<td>ex5-dataset-planned-host.json</td>
<td>JSON examples</td>
<td>12 days ago</td>
</tr>
<tr>
<td>ex6-dataset-closed.json</td>
<td>JSON examples</td>
<td>12 days ago</td>
</tr>
<tr>
<td>ex7-dataset-many.json</td>
<td>JSON examples</td>
<td>12 days ago</td>
</tr>
</tbody>
</table>

DMP and Project – JSON example

```
{
  "DMP": {
    "title": "Funded DMP",
    "description": "Example of a DMP header for a funded project.",
    "created": "2019-02-22T13:28:15.5",
    "modified": "2019-02-22T13:10:56.9",
    "contact": {
      "name": "First Last",
      "email": "test@test",
      "contactID": {
        "identifier": "https://orcid.org/0000-0002-4929-7875",
        "identifierType": "HTTP-ORCID"
      }
    },
    "ethicalIssuesExist": "false",
    "project": {
      "title": "Making maDMPs awesome",
      "projectStart": "2017-01-01",
      "projectEnd": "2020-12-31",
      "funding": {
        "funderID": {
          "identifier": "501100002428",
          "identifierType": "FUNDREF"
        },
        "grantID": {
          "identifier": "1234567-AT",
          "identifierType": "custom"
        },
        "fundingStatus": "granted"
      }
    },
    "dataset": {}
  }
}
```

DMP and Project – JSON example

```json
{
    "DMP": {
        "title": "Funded DMP",
        "description": "Example of a DMP header for a funded project.",
        "created": "2019-02-22T13:20:15.5",
        "modified": "2019-02-22T15:10:56.9",
        "contact": {
            "name": "First Last",
            "mbox": "test@test",
            "contactID": {
                "identifier": "https://orcid.org/0000-0002-4929-7875",
                "identifierType": "HTTP-ORCID"
            }
        }
    },
    "ethicalIssuesExist": "false"
}
```

DMP and Project – JSON example

```json
"project": {
    "title": "Making maDMPs awesome",
    "projectStart": "2017-01-01",
    "projectEnd": "2020-12-31",

    "funding": {
        "funderID": {
            "identifier": "501100002428",
            "identifierType": "FUNDREF"
        },
        "grantID": {
            "identifier": "1234567-AT",
            "identifierType": "custom"
        },
        "fundingStatus": "granted"
    }
},

"dataset" : {}  
}
```

### Controlled vocabularies

#### Broader problem
- identifier-type vocabulary

#### Can RDA help in solving this?

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Data Type</th>
<th>Cardinality</th>
<th>Example Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>dataset_id</td>
<td>Dataset ID</td>
<td>String</td>
<td>Exactly One</td>
<td><a href="http://doi.org/10.5281/zenodo.1172673">http://doi.org/10.5281/zenodo.1172673</a></td>
</tr>
<tr>
<td>dataset_id_type</td>
<td>Identifier type</td>
<td>Term from Controlled Vocabulary</td>
<td>Exactly One</td>
<td>HTTP-DOI</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Data Type</th>
<th>Cardinality</th>
<th>Example Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>funder_id</td>
<td>Funder ID, recommended to use CrossRef Funder Registry. See: <a href="https://www.crossref.org/services/funder-registry/">https://www.crossref.org/services/funder-registry/</a></td>
<td>String</td>
<td>Exactly One</td>
<td>501100002428</td>
</tr>
<tr>
<td>funder_id_type</td>
<td>Identifier type</td>
<td>Term from Controlled Vocabulary</td>
<td>Exactly One</td>
<td>FUNDREF</td>
</tr>
</tbody>
</table>
## Model – reused standards

<table>
<thead>
<tr>
<th>id</th>
<th>label</th>
<th>uri</th>
</tr>
</thead>
<tbody>
<tr>
<td>ali</td>
<td>Access License and Indicators</td>
<td><a href="http://www.niso.org/schemas/ali/1.0/">http://www.niso.org/schemas/ali/1.0/</a></td>
</tr>
<tr>
<td>dces</td>
<td>Dublin Core Element Set</td>
<td><a href="http://purl.org/dc/elements/1.1/">http://purl.org/dc/elements/1.1/</a></td>
</tr>
<tr>
<td>dct</td>
<td>DCMI Metadata Terms</td>
<td><a href="http://purl.org/dc/terms/">http://purl.org/dc/terms/</a></td>
</tr>
<tr>
<td>foaf</td>
<td>Friend of a Friend (FOAF)</td>
<td><a href="http://xmlns.com/foaf/0.1/">http://xmlns.com/foaf/0.1/</a></td>
</tr>
<tr>
<td>dcat</td>
<td>DCAT</td>
<td><a href="https://www.w3.org/TR/vocab-dcat/">https://www.w3.org/TR/vocab-dcat/</a></td>
</tr>
<tr>
<td>datacite</td>
<td>Data Cite</td>
<td><a href="https://schema.datacite.org">https://schema.datacite.org</a></td>
</tr>
<tr>
<td>cerif</td>
<td>Cerif</td>
<td><a href="https://www.eurocris.org/ontologies/cerif/1.3/index.html#currencyCode">https://www.eurocris.org/ontologies/cerif/1.3/index.html#currencyCode</a></td>
</tr>
<tr>
<td>coar</td>
<td>COAR</td>
<td><a href="http://vocabularies.coar-repositories.org/pubby/resource_type.html">http://vocabularies.coar-repositories.org/pubby/resource_type.html</a></td>
</tr>
<tr>
<td>iso6391</td>
<td>ISO 6391-1</td>
<td>Two letter country code</td>
</tr>
<tr>
<td>iso4217</td>
<td>ISO 4217</td>
<td>Currency code</td>
</tr>
</tbody>
</table>
Model assumptions – relaxed constraints

- Model must be applicable in different settings
  - relaxed constraints within the model
    - e.g. DMP can relate to a project [0..*)
  - constraints introduced at the ‘business level’
    - tool implementing the model
    - e.g. DMP must relate to a project [1..*)
- DMP instances are still compatible
Model assumptions - interoperability

- Model will be pre-dominantly used to exchange information between systems

- Internal representation of information in a DMP tool may differ (physical model)
  - e.g. database may have a different schema

- No ‘meta-fields’ about DMP
  - e.g. no DMP state field ‘final’
Model assumptions - versioning

- DMP versioning done by systems using the model
  - model provides fields allowing to identify DMP version
  - model does not track connections between versions
Model assumptions – evolving information

- Model expresses ‘certainty’ of provided information
  - to support different phases of DMPs
- Example
  - Source code will be issued on 2019-06-30 (planned) in ‘some-repo’. There will be an embargo period till 2019-12-31. Later on the source code will be available on a CC-BY license.

"DMP": {
  "modified": "2019-02-22T13:20:15.5"

  "dataset": {
    "title": "Source Code",
    "issued": "2019-06-30",
    "distribution": {
      "accessURL": "http://some-repo...",
      "license": {
        "license_ref": "https://creativecommons.org/licenses/by/4.0/",
        "startDate": "2019-12-31"
      }
    }
  }
}
Ongoing and future pilot implementations

Part 3
RDM Infrastructure

Pilot project
Mock-up for a tool + prototype

Goal
- generate easily and quickly DMPs
- not a training tool

Mock-ups
- To define requirements of ALL stakeholders

Deployment requires integration with university specific services
- e.g. researchers database, research support ticket system, etc.

There are common services to be co-developed
- e.g. repository recommendation service
System Integration

Funders

API

GitHub

Repositories

API, OAI-PMH

zenodo

Repositories Registries

API

re3data.org

Other

API

MySQL

IT Resources

ownCloud

Authentication & Administration

ORCID

FFG

FWF

DMP

GitLab
Mock-up of a funder view for maDMP

### DMP Funder View

**Home > DMPs > DMP#54365437012341**

**Reuse of pre-existing data**

<table>
<thead>
<tr>
<th>Dataset title</th>
<th>Origin</th>
<th>License</th>
</tr>
</thead>
<tbody>
<tr>
<td>Occurrence records download on 2018-11-05</td>
<td>doi:10.26197/5be00504ec357</td>
<td>CC-BY</td>
</tr>
</tbody>
</table>

**FAIR Data**

**Metadata standards**
- Dublin Core
- DataCite Metadata Schema
- DDI - Data Documentation Initiative
- CIE (Crystallographic Information Framework)
- CSMD (Core Scientific Metadata Model)

**Inferred FAIRness by repository selection**

<table>
<thead>
<tr>
<th>Selected repository</th>
<th>Dataset</th>
<th>Data access</th>
<th>PID system</th>
<th>AID system</th>
<th>Certificate</th>
<th>Quality Mgmt.</th>
<th>Versioning</th>
<th>Location</th>
<th>API</th>
</tr>
</thead>
<tbody>
<tr>
<td>GitHub</td>
<td>Source code for client application</td>
<td>open</td>
<td>none</td>
<td>none</td>
<td>none</td>
<td>no</td>
<td>yes</td>
<td>U.S.</td>
<td>other</td>
</tr>
<tr>
<td>Zenodo</td>
<td>Supplementary material</td>
<td>open</td>
<td>DOI</td>
<td>ORCID</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>EU</td>
<td>REST QAI-PMH</td>
</tr>
<tr>
<td>GESIS Data Archive</td>
<td>Raw data</td>
<td>open</td>
<td>DOI</td>
<td>CoreTrustSeal</td>
<td>no</td>
<td>-</td>
<td>-</td>
<td>Germany</td>
<td>QAI-PMH</td>
</tr>
<tr>
<td></td>
<td>Processed data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Licensing**

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Sharing strategy</th>
<th>Selected license</th>
<th>License planned to be active from</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supplementary material</td>
<td>keep closed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raw data</td>
<td>keep closed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Source code for client application</td>
<td>publish</td>
<td>Apache License 2</td>
<td>2020-01-01</td>
</tr>
<tr>
<td>Processed data</td>
<td>publish</td>
<td>Creative Commons Attribute (CC-BY)</td>
<td>2021-03-01</td>
</tr>
</tbody>
</table>
Mockups

Machine-actionable Data Management Planning Application

Introduction

Currently we are designing a system to make research data management planning machine-actionable. This involves the automation of workflows and exchange among information systems and services. If you are interested in machine-actionable DMPs or are a stakeholder of research data management (e.g. researcher), feel welcome to try out our mockups and give us feedback. Your help is very appreciated!

https://oblassers.github.io/dmap-mockups/
John Chodacki

Idea 1
Heike Görzig
Application of DMP-common-standard model on Horizon 2020 Template (FAIR)

Mapping to InstrumentDB of RDMinfoPool
• Helmholtz Zentrum Berlin
• Photon and Neutron Science, Laboratories
• No personal data
• Data Policy
• Repository – ICAT
Project (description, run time, funding ref., web page) createsDatasets ProjectDataSet (description, keywords, creators, access team, preview creation time, amount) basedOn DataSet (InstrumentDB)
<table>
<thead>
<tr>
<th>What metadata will be created?</th>
<th>RDMinfoPool</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DMP-common-standards</strong></td>
<td><strong>Metadata:</strong></td>
</tr>
<tr>
<td>identifier</td>
<td>description</td>
</tr>
<tr>
<td><strong>Metadata_schema:</strong></td>
<td><strong>Metadata_schema:</strong></td>
</tr>
<tr>
<td><strong>name</strong></td>
<td><strong>description</strong></td>
</tr>
<tr>
<td><strong>namespace</strong></td>
<td><strong>metadata_catalog:</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Will search keywords be provided that optimize possibilities for re-use?</th>
<th>RDMinfoPool</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DMP-common-standards</strong></td>
<td><strong>Dataset:</strong></td>
</tr>
<tr>
<td><strong>Static keywords generated from instrument name, method name, software name,</strong></td>
<td><strong>Dataset:</strong></td>
</tr>
</tbody>
</table>
# FAIR

What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?

<table>
<thead>
<tr>
<th>DMP-common-standards</th>
<th>RDMinfoPool</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metadata: identifier, description, language</td>
<td>file: file_metadata_schema (referencing table Metadata_schema)</td>
</tr>
<tr>
<td>Dataset: description</td>
<td>method:name, description</td>
</tr>
</tbody>
</table>

Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. ...?

<table>
<thead>
<tr>
<th>DMP-common-standards</th>
<th>RDMinfoPool</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distribution: format</td>
<td>file: format</td>
</tr>
<tr>
<td>file_collection: read_software</td>
<td></td>
</tr>
<tr>
<td>For describing provenance: Technical resource: description</td>
<td>For describing provenance: instrument:instrument_metadata software:name, version, configuration</td>
</tr>
</tbody>
</table>
### How long is it intended that the data remains re-usable? Are data quality assurance processes described?

<table>
<thead>
<tr>
<th>DMP-common-standards</th>
<th>RDMinfoPool</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset: dataQualityAssurance</td>
<td>Policy: description where policy type is validation (metadata_schema:metadata_schema, file, format parameter for validations with PolicyDB (which needs to be implemented))</td>
</tr>
</tbody>
</table>
Thank you!
<table>
<thead>
<tr>
<th>Role</th>
<th>Before project / OPA / EPA</th>
<th>Project initiation</th>
<th>Project planning</th>
<th>Project execution</th>
<th>Project finalisation</th>
<th>After project / OPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instrument scientist</td>
<td>Instrument and software description, selection of applicable metadata standards, general description of datasets</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Actualization of instruments and software</td>
</tr>
<tr>
<td>Data manager</td>
<td>Administration of controlled vocabularies and standards, mapping metadata to standards, general data policies, policy execution.</td>
<td></td>
<td></td>
<td></td>
<td>Automatic metadata extraction and validation</td>
<td>Open access of research data, validation of policy execution. Actualization of standards and policies</td>
</tr>
<tr>
<td>User coordination</td>
<td>Initializing proposal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Researchers</td>
<td>Dataset description, metadata schema selection, amount of datasets produced</td>
<td>Dataset description, metadata schema selection, amount of datasets produced</td>
<td>Experiment execution: parameter and configurations</td>
<td>Dataset selection, metadata completion and validation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project administration</td>
<td>Specific policies, DMP</td>
<td>DMP actualisation</td>
<td>DMP actualisation after experiments</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Discussion
Wrap-up and next steps

Part 4
Next steps – long term goals

- Focus on adoption
- Standard is frozen
- Create change requests on GitHub
  - To request new fields or modify existing
- Standard will be reviewed before the next plenary

Model presented

Feedback from adopters

Official RDA recommendation

Further adoptions

Spring 19
- Philadelphia

Fall 19
- Helsinki

Spring 20
- P15

Fall 20
- P16
Next steps – mid-term goals

- JSON Schema development
- Validator for maDMP instances
- Further serialisations
  - XML
  - OWL
  - JSON-LD
- Assistance in complying with the standard
Visit Active DMPs IG session

- More on machine-actionable DMPs

- Active Data Management Plans: Machine-actionable DMPs - Revisit Requirements, Review Outputs, Reflect on Next Steps

- Day 2
- 12:00-13:30
Staying in touch!

› Sign up to the group
  › https://www.rd-alliance.org/groups/dmp-common-standards-wg

› Visit GitHub repository
  › https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard

› Participate in model adoption

› Contact group chairs

Tomasz Miksa

Paul Walk

Peter Neish
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: Ten principles for machine-actionable data management plans. PLOS Computational Biology (in press)

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