DMPonline user group

https://dmponline.dcc.ac.uk

Sarah Jones, Sam Rust & Magdalena Drafiova

13 March 2019, University of Amsterdam
Preferred topics

• Machine-actionable DMPs
  ➢ Priority system connections e.g. ethics review, RIM, repository...

• Plan lifecycle
  ➢ Versioning of DMPs
  ➢ Notification of changes
  ➢ Archiving and publishing plans

• Assessing the FAIRness of data plans

• Usage statistics

• Other?
Machine-actionable DMPs
- Not this!
Why make DMPs machine-actionable?

Promote information flow across systems to support:

- Data discovery
- Capacity planning
- Aggregation/integration
- Policy compliance
- ....

From Flickr by highwaysengland, CC BY 2.0
What should connect? Priorities...

Planning & administration
Create, analyse, manage data
Publishing & reuse

Identifiers:
Research Data Alliance (RDA) is a global organisation supporting collaboration to promote data management and open sharing.

Active DMPs Interest Group

DMP Common Standards Working Group
- develop a common data model with a core set of elements
- provide reference implementations of the data model
https://www.rd-alliance.org/groups/dmp-common-standards-wg

Exposing DMPs Working Group
- develop use cases of what can be shared, when and with whom
- provide a reference model and alternative strategies for exposing DMPs
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
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"
} ],

Principles

• Reuse existing standards, e.g. Dublin Core, PREMIS
• Use PIDs whenever possible, e.g. ORCID
• Use controlled vocabularies
• Develop own concepts and vocabularies only when needed
Model in consultation: DCC plans to test

Interchange Format for maDMPS

This is a minimum set of universal terms that we agree on which ensure basic interoperability of systems using maDMPS. Further fields can be added in specific deployments, but they do not guarantee interoperability. DMP tools can use any other fields in their internal data models.

Cardinalities: each field in this model can be expected to appear as part of a message exchanged between systems. However, not all of fields or classes are required [0..1] or [0..*].

https://www.lucidchart.com/invitations/accept/ee26bc71-01a6-442a-b946-5b9c910fb926
Machine-actionable DMP discussion

• What are the priority systems to connect with?
  ➢ Research information systems? Which?
  ➢ Repositories? Which?
  ➢ Ethics processes?
  ➢ Other?

• What information do you want to extract from DMPs?

• What information can feed into DMPs and when?

• Other questions / comments?
Potential lifecycle

Phases based on NERC lifecycle

Phases based on proposed DCC default lifecycle

Icon identifies a plan creation phase that will generate a DMP following the Plan Creation Lifecycle.

ISSUE:
Who/what should trigger this type of transition between phases? The DMP forms only part of a wider project grant application? Should the DMP being complete be sufficient to move on to next phase, or should we wait until a user explicitly signals that the application has been submitted?
Questions...

What stages / statuses need to be defined?
➢ Draft
➢ Submitted (to funder? for internal review/approval?)
➢ ‘Active’ (an *accepted* DMP in progress / post-award)
➢ Final DMP / research project finished
➢ Others?

Version control
➢ Should versions save at given status points?
➢ Should users be able to save a snapshot at any time?
➢ Should users / administrators archive a copy of DMPs?

Other requirements
➢ Should funders / admins / others be notified of changes?
Publishing plans

- Publish via DMPonline
  - Does this need to be more formal with DOIs?
- Export to repositories e.g. Zenodo
- Submit to journals e.g. RIOjournal
- Others?
Assessing the FAIRness of data plans
What makes data FAIR?

FOSTER advice to EC project officers on the six main questions in H2020 DMP to emphasise in review:

- 1.c description of data types and formats
- 2.1.a unique and persistent identifiers (PIDs)
- 2.1.d rich metadata
- 2.2.c repository deposit
- 2.3.a data and metadata standards for interoperability
- 2.4.a data licensing

ZonMw key items:

- Persistent identifier
- Digital repository, preferably certified
- Online catalogue where data is registered (may be repository)
- Terms and conditions of use
- Data format (preferably machine-readable)
- The terminology for the data (code, classification, ontology)
- Metadata schema
Lessons: some aspects are more important

- Rich documentation
- Standardised metadata, use of controlled vocabularies
- Accepted (open, well-known) data formats
- Deposit in a trusted repository
- Clear licensing, ideally open / unrestrictive
- Assigning PIDs

Can we collect this data in a structured way in DMPs so potential FAIRness can be automatically assessed?
MSD added as an answer type

Tool pulls in information from the API to let users browse and select metadata standards

What are the metadata standards you will use?

Your Selected Standards:

Please select a subject

Multidisciplinary

Browse Standards

**CERIF (Common European Research Information Format)**

The Common European Research Information Format is the standard that the EU recommends to its member states for recording information about research activity. Since version 1.6 it has included specific support for recording metadata for datasets.

**Data Package**

The Data Package specification is a generic wrapper format for exchanging data. Although it supports arbitrary metadata, the format defines required, recommended, and optional fields for both the package as a whole and the resources contained within it.

A separate but linked specification provides a way to describe the columns of a data table; descriptions of this form can be included directly in the Data Package metadata.
Select standards and add comments

**What are the metadata standards you will use?**

Your Selected Standards:
- EML (Ecological Metadata Language)

**Please select a subject**
- Science

**Please select a sub-subject**
- Environmental sciences and e

**Browse Standards**

**DIF (Directory Interchange Format)**

An early metadata initiative from the Earth sciences community, intended for the description of scientific data sets. It includes elements focusing on instruments that capture data, temporal and spatial characteristics of the data, and projects with which the dataset is associated. It is defined as a W3C XML Schema.

Sponsored by the Global Change Master Directory, the DIF Writer's Guide Version 6 is from November 2010.

**EML (Ecological Metadata Language)**

Ecological Metadata Language (EML) is a metadata specification particularly developed for the ecology discipline. It is based on prior work done by the Ecological Society of America and associated efforts (Michener et al., 1997, Ecological Applications).

Sponsored by ecoinformatics.org, EML Version 2.1.1 was released in 2011.

**Comment**

EML is one of the accepted formats used in ecology, and works well for the types of data we will be producing. We will create these metadata using Morpho software, available through KNEI.
Potential to integrate others

- Suggest standards
- Recommend repositories
- Pulling in grant numbers
- ....
Utilising EC grant IDs in plans

- Harvest grant IDs from OpenAIRE API
- Provide look up when entering project details
- Enables join up of DMP with other outputs
FAIR and DMPs?

• Can we get insights into the prospective FAIRness of the data via DMPs?
• What are the key aspects to check?
• Where should we focus efforts?
• What are your needs?
Usage statistics

• DMPonline currently offers:
  ➢ Total users (by month and adjustable date range)
  ➢ Total plans (by month and adjustable date range)
  ➢ Global usage statistics for comparison
  ➢ Access to plan content for review
  ➢ Download of basic plan metadata

• What else do you need to know?

• How else do you want to use the API?
Other things we need to discuss?
# Connect & stay in touch

<table>
<thead>
<tr>
<th></th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Helpdesk</td>
<td><a href="mailto:DMPonline@dcc.ac.uk">DMPonline@dcc.ac.uk</a></td>
</tr>
<tr>
<td>Twitter</td>
<td>@DMPonline and #ActiveDMPs</td>
</tr>
<tr>
<td>Blog</td>
<td><a href="http://www.dcc.ac.uk/news/DMPonline">http://www.dcc.ac.uk/news/DMPonline</a></td>
</tr>
<tr>
<td>User group</td>
<td><a href="http://tiny.cc/DMPonline-user-group">http://tiny.cc/DMPonline-user-group</a></td>
</tr>
<tr>
<td>Github</td>
<td><a href="https://github.com/DMPRoadmap">https://github.com/DMPRoadmap</a></td>
</tr>
<tr>
<td>Slack</td>
<td><a href="http://tiny.cc/DMPRoadmap-slack">http://tiny.cc/DMPRoadmap-slack</a></td>
</tr>
</tbody>
</table>