

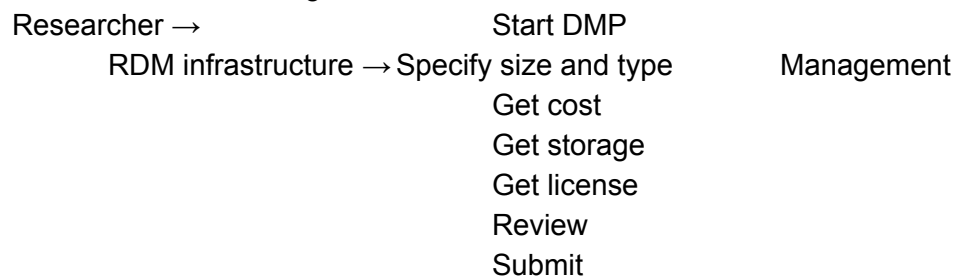
WG DMP Common Standards: DMP Common Model and New Pilots

Intro:

There are things we want to do with DMPs that we currently can't (system integration etc)
Currently it's all free text

Goal: get different stakeholders info at the right time. Better/more efficient information flow.
Make Machine actionable DMPs possible

Automated Data Management Workflow



Multiple actors may have information about different stages of management.

Why do we need this WG?

Shortcomings of existing DMPs

Machine actionable DMPs

Living documents

Automate data management, collect information from systems, trigger actions in system

Common data model allows Machine actionable DMPs to be possible.

Not the same as a template or a questionnaire.

a core set of elements to allow people to add specific extensions within their own domains.

Reference implementations (in JSON, XML etc)

Guidelines for adoption of the common data model

Requirements for supporting systems

Example of current DMPs versus Machine actionable. Incorporates standards already in use.

PIDs, ontologies, controlled vocabularies.

DMP Common Standards WG

Launched in October 2017 as result of consultation Active DMps interest group to focus on Machine-actionable DMPs.

Has 100+ members on all continents.

Paper summarizing activities of the WG within the first 12 months (presented at IEEE CAS workshop): <http://doi.org/10.5281/zenodo.2161855>

Now working on some models.

First consultation user stories

Goals: identify stakeholders at each lifecycle stage; define information they provide

Second consultation: existing models.

How do we model specific requirements? Which specific fields are needed? Which models exist?

Proof of concept tools

See:

github.com/TomMiksa/DMPGenerator

Processes mapping to help identify

- Tasks performed by stakeholders (e.g.

- 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)

Narrow down focus of this WG

Common model does not contain business logic (e.g., cost estimation done by service that provides value)

It is an information carrier

(tools, services, processes make maDMPs machine-actionable)

Common Model for maDMPs (part 2) (Tomasz presenting)

(diagram of the model)

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

<https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard/tree/master/docs/diagrams>

Some parts of model not elaborated because it overlaps with the concerns of other working groups.

Metadata is for what metadata standards are used in the dataset - not a place for extra metadata about the DMP.

Dates - modification dates allow you to distinguish different versions of a DMP.

Identifier types - can use different types. - http doi or is there scope for different types

Interoperability: Model incorporates existing standards and vocabularies.

Model is an interchange format - systems can implement their own internal data structure and functions

Versioning includes an ID but does not reference previous versions of the DMP.

Dates can indicate a future state - eg when source code will be released

Future use of the Model and Pilot project

Use Cases:

Technical University Wien

Mock-up for a tool prototype

System integration to pull in information from funders, repositories, registries, IT resources, authentication and administration.

DMP Funder View

Get view automatically (mock up)

<https://oblassadors.github.io/dmap-mockups/>

Heika Görzig

Helmholtz Zentrum Berlin

Application of DMP-common-standard model on Horizon 2020 Template (FAIR)

Mapping to InstrumentDB of RDMInfoPool

Photon/neutron sources - A good use case for developing a common workflow for DMP integration, as the researchers are using these light sources in similar ways

FAIR

What metadata will be created?

DMP Common standards RDMInfoPool

Will search keywords be provided that optimize possibilities for re-use?

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

Are the data produced in the project interoperable, that is allowing data exchange and reuse between researchers, institutions, organizations, countries, etc??

How long is it intended that the data remains reusable? Are data quality assurance processes described?

Machine Actionable DMPs

Robin Dasler & John Chodacki CDL - Data Cite (CDL and Datacite partnering on this)

Presentation:

https://docs.google.com/presentation/d/1SzEvSEaO5_9z2t8A6TCX9DKuO8ovewtCUTXtCVZ5mbg

If DMPs get DOIs ...

With DataCite DOI, DMPs can get:

Persistent and actionable identifiers

Access to DataCites supporting DOI services

DOIs are static - DMPs are active - how to bridge this gap?

Leverage infrastructure with DOIs

Crossref + DataCite. Capture relationships asserted by DOI.

Event Data <https://datacite.org/eventdata.html>

Captures relationship

PIDGraph. Crosswalk them.

Challenges for maDMPs -- getting stuff in

Event Data stores relations and connects two existing entities identified by PIDs, but does not store the entities themselves. (Both sides need a PID)

Challenge: Not every item in a maDMP's lifecycle has a PID

Event Data gets info from Crossref and Datacite metadata.

Repositories can update their metadata to add relations. Third parties cannot contribute directly.

Challenges:

Who runs the agents to pull in the relations? Could you successfully proxy these instead? (e.g. Crossref grant IDs)

RIOJournal published DMPs. 2 of 8 had relations but types were ambiguous.

Challenges for maDMPs -- getting stuff out

At the end of the day, we just have a pile of relations

How is the maDMP displayed to people? Does the DOI landing page display this? Dynamically?

Privacy concerns. Trust sources, verify or trust assertions?

Next steps

Common standards to represent the core info

How to expose the info in maDMPs

Gump Field Station inputs

WHOI/BCO-DMO as inputs

Confession: wrong slides submitted for this session. See the right slides in the upcoming "Exposing DMPs" session. Here are the correct slides:

https://docs.google.com/presentation/d/1SzEvSEaO5_9z2t8A6TCX9DKuO8ovewtCUTXtCVZ5mbg

Next steps and longer term goals for this working group

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.

JSON schema development

Validator for maDMP instances

Further serialisations

XML

OWL

JSON-LD

Assistance in complying with the standard

QA

<https://journals.plos.org/ploscompbiol/article?id=10.1371%2Fjournal.pcbi.1006750>