

DMP Tools Actionable

Gerald Weber, Helmuth Breitenfellner

01.07.2019

Make DMP Tools Actionable

- **194.045 Data Stewardship 2019S**
- Gerald Weber, 0125536
- Helmuth Breitenfellner, 08725866

Link to GitHub Pages

<https://helmuthb.github.io/dmp-tools-actionable/>

Set Up Instance of RDMO

We have been setting up an instance of RDMO on a virtualized environment, using Docker. The installation can be used at rdmo.helmuth.at.

For this part of the exercise we have been using the docker-compose version of RDMO as available on GitHub.

Re-created DMP in RDMO

RDA DMP Export for RDMO

Approach to Connect with RDMO

We were investigating the following possible ways of connecting with an RDMO instance:

- Database connection
- RDMO API
- XML export from RDMO

Using the database access was considered the last resort, as this would be dependent on internal data structures of the tool.

The RDMO API did originally sound very promising. However, we found that existing installations often disabled the API (e.g. rdmo-demo.uibk.ac.at/), or the API is not accessible for regular users.

We settled on the third option, which is using the XML which can be exported from a project in the RDMO user interface. This allows using the export tool for regular users, even if the API has been disabled in the installation.

Using the Tool

The tool is now a simple command line tool. As a precondition, the user has to export the project as XML from RDMO. Then the tool can be started with

```
python3 xml2madmp.py <rdmo-export.xml> <ma-dmp.json>
```

License

The tool is licensed under the MIT license.

Mapping existing DMP templates to RDA DMP Common Standard

For this part of the exercise we have used a two-part mapping.

First, we created a questionnaire based on the Horizon 2020 template and FWF template, which maps the questions to the corresponding fields and attributes of the RDMO model.

Then, we were applying our mapping from RDMO attributes to the RDA DMP Common Standard.

Map Horizon 2020 template to RDA DMP Common Standard

Some of the questions from Horizon 2020 do not clearly map to attributes in the RDMO data model. We were using guidelines from the existing Horizon 2020 view which performs the mapping into the other direction.

Fermont Bridge DMP

- Horizon 2020 PDF version
- RDA DMP Common Standard JSON version

Salzburg - US Wheat DMP

- Horizon 2020 PDF version
- RDA DMP Common Standard JSON version

Map FWF template to RDA DMP Common Standard

Similar issues as in the case of Horizon 2020 were encountered. In addition, we did not cater for the last case when no data is processed, as a DMP is then not needed at all.

Fermont Bridge DMP

- FWF PDF version
- RDA DMP Common Standard JSON version

Salzburg - US Wheat DMP

- FWF PDF version
- RDA DMP Common Standard JSON version

Field Mapping from RDMO to RDA DMP Common Standard

RDA DMP Field	RDMO source
title	title
description	description
language	* <i>always set to en</i>
created	created
modified	* <i>latest modification date of any value</i>
ethical_issues_exist	* <i>loop through datasets</i>
contact	coordination/name split into name and email
project.start	schedule/project_start
project.end	schedule/project_end
cost	* <i>loop through cost-entries</i>
dataset.title	dataset/id

RDA DMP Field	RDMO source
<code>dataset.type</code>	<code>dataset/format</code>
<code>dataset.description</code>	<i>* combined from various fields</i>
<code>dataset.data_quality_assurance</code>	<code>dataset/quality_assurance</code>
<code>dataset.personal_data</code>	<code>dataset/sensitive_data/personal_data_yesno/yesno</code>
<code>dataset.sensitive_data</code>	<code>dataset/sensitive_data/other/yesno</code>
<code>dataset.distribution.description</code>	<i>* combined from various fields</i>
<code>dataset.distribution.license.license_ref</code>	<code>dataset/sharing/sharing_license</code>
<code>dataset.distribution.license.start_date</code>	<code>dataset/data_publication_date</code>

Special Cases

Language Field `language`

As the RDMO data model does not have a field for the language of a project this field is set to English for all DMPs.

Field `modified`

In RDMO, the field `updated` more or less corresponds to the `modified` field in the RDA DMP Common Standard. However, since RDMO uses a relational database model, the `updated` field of a project is only changed if the fields in the corresponding table are updated.

To get a more natural interpretation of the last modification date the tool is looking for the latest modification of any field or value in the RDMO data and uses this value as the `modified` field in the RDA DMP Common Standard.

Field `ethical_issues_exist`

The tool uses the following logic:

- If for *all* datasets the questions for both personal data issues and sensitive data issues are answered with *no* then this field is set to **no**.
- If for *at least one* dataset one of the questions is answered with *yes* then the field is set to **yes**.
- Otherwise, if the question is never answered with *yes* but is not answered for at least one dataset, the field is set to **unknown**.

Substructure `cost`

All cost-related entries in RDMO are parsed. The entries are added as sub-entries to the `cost` element, using the following translation logic:

- `title`: This is mapped to the key in RDMO, e.g. `ipr/non_personnel`.
- `cost_value`: This is mapped to the value in RDMO.

Field `dataset.description`

Since there are fields in RDMO with no direct mapping to RDA DMP Common Standard, the field `dataset.description` is filled with data from other RDMO fields:

- `dataset/description`
- `dataset/interoperability`
- `dataset/creation_methods`
- `dataset/metadata`

Field `dataset.distribution.description`

Since there are fields in RDMO with no direct mapping to RDA DMP Common Standard, the field `dataset.distribution.description` is filled with data from other RDMO fields:

- `dataset/size/number_files`
- `dataset/versioning_strategy`
- `dataset/structure`
- `dataset/reuse_scenario`
- `dataset/sharing/conditions`