Import and Export of maDMPs from Argos/OpenDMP

Diamadis Tziotzios (OpenDMP / Argos), Georgios Kakaletris (OpenDMP / Argos), Georgios Kalabokis (OpenDMP / Argos), Elli Papadopoulou (Athena Research Center/ OpenAIRE)

The DMP Ninjas worked on the [OpenDMP software](#), an open source software for Data Management Planning developed in collaboration between OpenAIRE and EUDAT CDI with installations already provided by both infrastructures: [Argos](#) in OpenAIRE and [easyDMP](#) in EUDAT CDI.

During the hackathon, we focused on the following issues:

1. [Import/ export function](#) to test how maDMPs react during an exchange with other machine-actionable tools.

Revision of Argos H2020 template specification and refinement of the import/export mechanism to meet the needs of the updated maDMPs standard were performed. The focus was the latest version of the maDMP specification and the following are the major areas of work:

- Extension of Argos template semantics to match specific elements of maDMP which although optional for Argos/OpenDMP may be essential for other tools, in order to exchange the DMPs, e.g., licensing.

- Creation of default entities for items that do not directly match the OpenDMP data model, e.g., dataset distributions, hosts etc.

- Introduction of placeholder for auxiliary DMP data, which may be used to store all data not currently in the model, yet managed by OpenDMP, so as to be able to maintain all DMP information across imports/exports.

Furthermore, successful tests were conducted with DMPs imported from or exported to other maDMP tools, e.g., the DSW, as seen in the demo of the grand finale video [here](#).

2. [Information exchange between OpenAIRE Research Graph and maDMPs](#).

Preparatory work to map the two models of [RDA DMP Common Standard](#) and [OpenAIRE](#) started during the course of the hackathon and was presented at the grand finale session. It should be noted that this activity took into consideration the latest version of the DCS at the time of the hackathon while a new release was simultaneously refined by another hacking team.

The mapping concentrated in identifying common entities and properties between the two models while paying attention to major deviations in their cardinality. Relationships available in the Research Graph were included too. This exercise showed that:

- Most of the maDMP entities can be mapped directly to the Research Graph, such as contact, contributor, dataset, and project.

- There are few entities, including their properties, that cannot be directly fit into the Research Graph model, cost being a typical example for being totally absent. In OpenAire, this information might be found at more abstract/general fields, covered by Argos DMP outputs either as they come or tweaked to accommodate the needs of maDMPs documentation, or, rarely, omitted.

- There is a need for new entities, properties and relationships to be included in the Research Graph facilitating information exchange with maDMPs. Metadata, for instance, was identified to be missing from this environment. DMPs themselves are in the process of being included as independent entities in the Research Graph, hence enhancing it with additional information coming from DMPs.

More details about the initial mapping can be found here. During the mapping exercise, a couple of issues came to the surface which could not be dealt with at the time, but have been noted and were shared with other colleagues-participants of the hackathon to further engage in/ trigger discussions.

Thoughts and lessons learned concern:

a. Identifying and claiming DMP outputs

Although global vocabularies, such as COAR, have introduced a new **resource_type** exclusively for DMPs, there seems to be a lack of use of the field. Hence, DMPs are still claimed as resource_type publications or as other outputs, like reports, and validated by searching for keywords in the title and in the description.

OpenAIRE and Zenodo are already working to update their metadata schema based on COAR and other vocabularies that treat DMPs as a separate resource_type. As soon as this enhancement is ready, DMP outputs will be available in OpenAIRE Explore.

b. Pre-filling information in DMPs

There is some complexity and sensitivity regarding whether **re-used datasets** should be also handled in a DMP. In addition, differences are identified according to (i) the purpose and context of the DMP creation and (ii) the time of the DMP lifecycle when the pre-filling takes place.

Based on the aforementioned, a quick distinction can be made between DMPs created for a project or for an institution. For example, institutions can set their own services on a DMP template by pre-filling it with information as appears in the institution's RDM policy. A very common example is the institutional repository where the datasets will reside. However, even in that case, there are dependencies concerning, among other things, the model and the type of data documented. Some examples are:

- Information about the author of DMPs and affiliation, which might not be automatically filled when entering the DMP or creating new templates via the given configured Authentication and Authorisation mechanisms.

- Information about the projects, particularly when it is the first DMP for a given project and no prior data exist for it.

- Information about datasets, highlighting the differences between new versus re-used data. Existing datasets can be easily claimed by a repository or registry and return back metadata about their title, authors, formats, licences, etc. assigned to them. Those information can be pre-filled in the DMP tool and respective institutional DMP templates. It is apparent that this is not the case for new datasets which have not been described or deposited following Open and FAIR practices, yet.

Among further questions arising from that scenario is how flexibility of researchers is ensured when choosing other services than those offered by the institution. For example, how can researchers involved in a research community specific project select a thematic repository outside the institutional spectrum? Argos can facilitate that through its APIs.

Moreover, to pre-fill information in DMPs there are some dependencies that lie with the given status of the DMP lifecycle. Some information can be pre-filled immediately when creating a DMP, such as the organisation or storage for DMPs deployed by institutions, and others might be the derivative of a first selection which has the ability to unlock dependent fields as they appear in a consequent way, such as the datasets that are included in a given repository.

Below there is a description of a use-case scenario for projects where DMP Common Standard is used to pre-fill information from the view of the Argos model.

Basic DMP Information
Contact (lies with the DMPtool) -> basic information from one initiating the DMP writing process can be pre-filled from login/profile information

Contributors -> can't be identified yet

Project and Funding (from the Research Graph)

DMP and Cost -> information about DMPs will come from Argos to enhance the Research

Graph once the DMP is finalised; some might be already there for DMPs under Resource_Type Publications or Other Outputs

Dataset (check below) -> this is possible only for datasets that are being re-used

- *Re-used data only*
  - Basic info:

DMP, Contact, Contributors and Cost (provided that there is a DMP in place and that the DMPtool assigns roles)

Project and Funding (from the Research Graph)

  - More specific to datasets:

Dataset, License, Host, Distribution, Technical Resource

No **Security** and **privacy** and **metadata** at the moment in the Research Graph (ongoing)

In parallel to the hackathon, coordination between OpenAIRE and Zenodo was happening to accommodate the new resource_type in OpenAIRE Explore and to introduce new relationships in the Research Graph.

Next steps involve updating the mapping with the newest version of both the DCS and the Research Graph so as to test it to random datasets included in OpenAIRE Explore. Then, respective issues on the RDA Github page will be updated, worked on and closed. Communication with all participants of the hackathon is hoped to be continued and maintained.