Making Data Management Plans Machine Actionable: Templates and Tools

JOAKIM PHILIPSON [ID]

ADIL HASAN [ID]

HANNE MOA [ID]

*Author affiliations can be found in the back matter of this article

]u[ubiquity press

ABSTRACT

Since September of 2019, a task group within the European Open Science Cloud - EOSC Nordic Project, work-package 5 (T5.3.2), has focused its attention on *machine-actionable Data Management Plans* (maDMPs). A delivery working-paper from the group *(Hasan et al.* 2021) concluded in summary that extracting useful information from traditional free-text based DMPs is problematic. While maDMPs are generally more FAIR compliant, and as such *accessible* to both humans and machines, more *interoperable* with other systems, and serving different stakeholders for processing, sharing, evaluation and *reuse*. Different DMP tools and templates have developed independently, to a varying degree, allowing for the creation of genuinely machine actionable DMPs. Here we will describe the first three tools or projects for creating maDMPs that were central parts of the original task group mission. We will get into a more detailed account of one of these, specifically the Stockholm University – EOSC Nordic maDMP project using the DMP Online tool, as described by *Philipson* (2021). We will also briefly touch upon some other current tools and projects for creating maDMPs that are compliant with the RDA DMP Common Standard (RDCS), aiming for integration with other research information systems or research data repositories. A possible conclusion from this overview is that the development of tools for maDMPs is progressing fast and seems to converge towards a common standard. Nonetheless, there remains an immense amount of work to get there.

INTRODUCTION

Data management plans (DMPs) have been in use since the late 1960s as research and development project management tools in disciplines with complex data management needs. At this early stage, the development of DMPs was driven mainly by researchers with specific requirements on data management (*Hasan et al.* 2021). Later, from beginning of this century, with the advent of the Open Science Movement, other stakeholders became involved with DMPs. Such as, funding organisations issuing their own DMP templates expressing their requirements on data management of funded projects to be transparent for results to be verifiable. Other stakeholders that emerged were academic institutions (universities) involved in research administration, and society at large represented by governmental bodies, demanding publicly funded research to be as openly available as possible. Part of this movement was also the issue of the FAIR principles (*Wilkinson et al.* 2016). This in turn called for DMPs to become machine-actionable (maDMPs), as a requirement for complying notably with the FAIR principles of *accessibility* and *interoperability*, thereby facilitating integration with the whole research data management infrastructure, as suggested by *Miksa et al.* (2019). Responding to the call for maDMPs, the Research Data Alliance (**RDA**) developed from 2018 the *RDA DMP Common Standard for machine-actionable Data Management Plans* (**RDCS**), the latest release of which (at the time of writing) is Version 1.1 from November 11, 2020. Compliance with the RDCS has been one of the important guiding principles in the development of online tools for the creation of maDMPs.

METHODS

In the EOSC Nordic T5.3.2 task group, we explored three different available DMP tools individually and the possibility of creating maDMPs that are, at least to some extent, compliant with the RDCS. In each case, the method employed had to be adapted to the structure and metadata of that particular framework. The fulfilment of this task also required, at least in two separate cases (DMP Online and easyDMP), development of software scripts and templates to make the output of DMPs within the frameworks of these services' machine-actionable and RDCS compliant, at least to the extent of validating against the RDCS madmp-schema-1.1.json.

RDCS AND THE MADMP-SCHEMA-1.1. JSON

The RDA DMP Common Standard (**RDCS**) is embodied in a data model (https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard) and a validation schema, the current maDMP-schema-1.1.json (2020). Both of these define the data elements and their properties in this model, including which elements and properties are mandatory (required, that is having a cardinality of 1 or more). The required top elements, or properties, of a DMP are only 8 in RDCS: "contact", "created", "dataset", "dmp_id", "ethical_issues_exist", "language", "modified" and "title". This means that such presumably important information elements as contributors, or even the research project with its funding (often constituting the most important reason why a DMP is first created, as a direct response to a requirement from the funder organizations) could be left out of a DMP that would still validate against the maDMP-schema.

DMP TOOLS AND TEMPLATES

DMP ONLINE

DMP Online (2022) is a tool supporting DMP creation and management offered by the Digital Curation Centre (DCC), based in the UK, but now serving individual users and institutions all over Europe. The DCC and the California Digital Library (CDL) are collaborating on a development plan, DMP Roadmap, that is being successively implemented in their tools (DMP Online and DMPTool) respectively at a different pace. DMP Online offers access to over twenty public funder templates that the individual users can choose between for the creation of their DMP. These different templates tend to vary considerably in what information they ask for. What they have in common is that they are almost exclusively free, and text based (requiring shorter or longer essay answers to questions). To make these machine-actionable, if possible, at all, they would at least involve recourse to natural language processing (NLP) techniques. In addition to the public funder templates, DMP Online offers the possibility of creating your own local institutional templates which are only fully accessible to the other local users, others affiliated at your institutions. There

are only some further functional limitations, compared to the funder templates. Both partners of DMP Roadmap also offer APIs for extracting filled out DMP outputs in JSON format.

SU-EOSC NORDIC 5.3.2 MADMP PROJECT

In this project, Stockholm University (SU) developed a custom template in DMP Online that was named the *SU-VR template* as it is based on the Science Europe and Swedish Research Council (VR) sections and questions, but the output of which is machine-actionable to the extent that it validates against the RDCS maDMP-schema. To achieve this proved to require some further processing of the output received via the DMP Roadmap API (v0). However, the purpose of the project has become threefold, where the first of these was the most important to us:

1. To ease the administrative burden on our SU researchers to fill in a DMP, by means of multiple-choice answer options, drop-down menus, radio buttons etc., including extensive local guidance.

2. Facilitate semi-automated review and evaluation of potential FAIR-ness while retrieving information required for RDM administration elsewhere in the local infrastructure.

3. Conform to the RDCS by validation against maDMP-schema-1.1.json.

There were mainly practical reasons for performing this pilot work in DMP Online. SU had already earlier signed up as an institution for the DMP Online service at DCC, and even before the beginning of the EOSC Nordic project, SU was offering it as a service to our researchers. Thus, there was a perceived local need to develop this service as RDM administrators also realised they would never be able to keep up with the growing demand for reviewing DMPs, if these were largely in the form of full text essays and there were no common evaluation criteria for these. Another reason for this choice was later the free provision of DMP Online as a test service to all Swedish universities and institutes of higher education for a limited test period by the Swedish Research Council. There was a perception that other institutions, in Sweden or elsewhere, could also benefit from the work done at SU on maDMPs in DMP Online (*Hasan et al.* 2021).
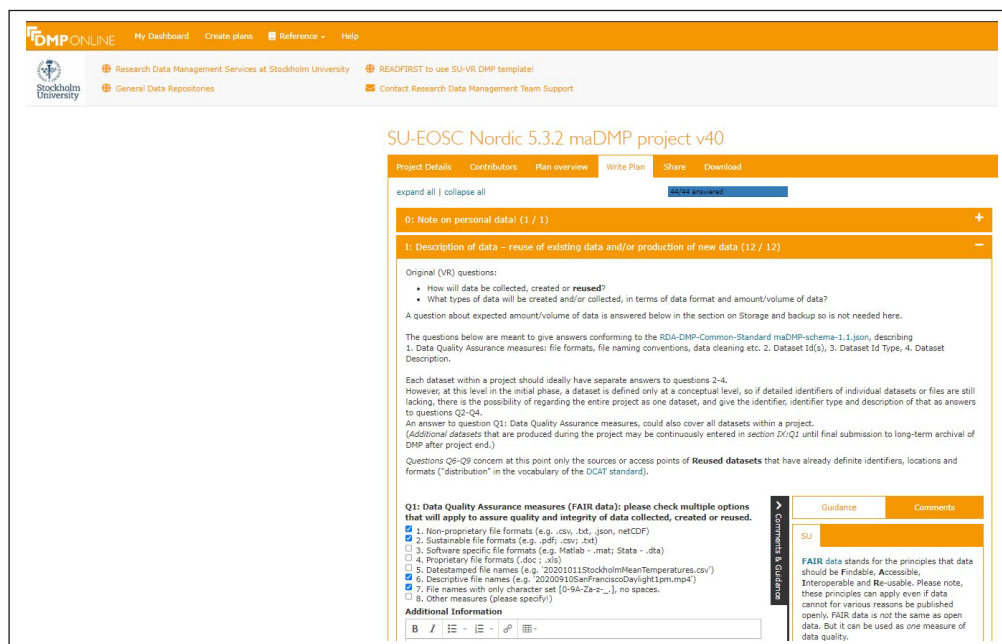

Figure 1 Showing Section I, Question 1 with multiple-choice answering options in the SU-VR template.

The SU-VR template in DMP Online is based on the Swedish Research Council (*Vetenskapsrådet*, VR) and Science Europe model (sections I-VI and original questions), but with more specified questions and answer options by means of multiple choice checkboxes, dropdown menus and radio buttons to make the output more machine-actionable. The template has answers formatted with respect to the SU Open Science Policy, local research data management rules, and the RDCS. Replacing an earlier two-phase model led to introducing the recent versions of a final section of the SU-VR template, *IX: Full DMP – additional Datasets and identifiers, Reference list and Project end,* to mark the completion of the research project described in the DMP. Among other features of later versions is a mapping of funder names to fundRef ID allowing enhanced compliance with the RDCS madmp-schema. Also included is a Schematron schema (working on the transformed xml output files) for assessment of prospective FAIR-ness

of the RDM measures described by the DMP. The Dataverse draft record for the project includes: DMP Online instances of DMPs that used the SU-VR maDMP template, with raw API (v0) JSON output, converted to XML, then transformed using an XSLT-file and converted back to JSON for validation against the RDCS madmp-schema-1.1.json. The transformation file *SUDMP2maDMP1-1.xsl* further uses: a direct download of the DMP export.json, converted to xml, as a parameter document for the provision of some information elements lacking in the API v0 output, notably start and end-date of project. Recent versions of the SU-VR DMP template, the transformation file SUDMP2maDMP1-1.xsl, and the Schematron schema SUDMP-FAIReval.sch have been updated and are all available in the Dataverse record. This also includes a refined validation of consistency between dataset identifier and identifier type.

Since its inception, and first version in November 2020, the SU-VR template has become the DMP template most frequently used by Stockholm University researchers among the different templates provided. The template is more used than the Swedish Research Council general template which comes in second place, even when excepting test plans.

CHALLENGES ENCOUNTERED

The framework for constructing a maDMP template within DMP Online has its limitations. Among the challenges encountered were the dependence on a well-functioning API (v0) from DMP Roadmap, which at times was not fully operable. The API v0 gives the full content, questions and answers, except for the cover sheet with Project Details such as Funder and Grant Number (in DMP Online as a separate tab of the DMP). This is a serious limitation because this information might be required by the funders demanding a DMP from their grantees. Moreover, DMPs created from one of the public funder templates get the Funder field on the cover sheet automatically pre-populated. Whereas for DMPs created from local templates, it has to be filled in manually and only from the selected list of previously acknowledged funder organisations. In response to this, DMP Online has mentioned the possibility of eventually "decoupling" templates and funders in the system to solve these problems by showing greater flexibility. API v1, which is still under development and not fully implemented in DMP Online, on the other hand delivers an output of information content exclusively meant to satisfy and comply with the RDCS. This means, in practice, that parts of the original content may be left out while other parts are satisfied by producing e.g. a generic or "fake" dataset entry, since this is a mandatory element in the RDCS that is still lacking in perhaps most funder templates.

To make output (in JSON) compliant with RDCS, it first has to be converted to XML, transformed (by means of XSLT), and then reconverted to JSON in order to validate against the RDCS madmp-schema-1.1.json. However, the conversion from JSON to XML and back to JSON cannot handle or distinguish pure numbers (integers) from strings, which are required for dataset id-s and this made it necessary to prefix the output from these input fields with an 'ID-' to ascertain getting a string datatype for validation against the RDCS maDMP-schema-1.1.json.

Further, the DMP Online framework does not allow for pre-ingest validation of data types or pattern matching of input strings in the template web form. This means there is little possibility of controlling or influencing the type of input/output received, other than through very clear guidance that may not always get the attention necessary to avoid less meaningful answers.

RDCS is based on the DCAT 2.0 model for datasets, allowing repeated entries for datasets and their distributions. This is not easily accommodated within the DMP Online framework for templates where there is no accommodation of repeated entry fields.

Copying and sharing the template web-form, other than as a PDF, in a more editable format proved difficult. Already from start there was an ambition to share with others, as much as possible, of the work in this project, which has been presented in other RDM and research environments (Swedish, Nordic and European) on several occasions. Within DMP Online, you can share individual DMP instances with other users by invitation, and individual DMPs could also be made publicly available, but then again only as PDFs (as was done with an earlier version of the SU-maDMP project). However, the maDMP template itself, the basis of our work, can only be shared as a PDF which is not ideal for allowing editing and adaptation to local needs at other universities.

Though, recently a full dataset,including: several versions of the SU-VR template, example DMPs created with this template, and processing scripts (xslt and a Schematron schema) were finally published in Dataverse (*Philipson* 2022).

Other challenges were not directly dependent on the DMP Online framework, but resulted from the RDCS and different FAIR metrics tools. This means that assessment of (potential) FAIR-ness of RDM measures in the research project, as described by the DMP, is not part of the RDCS proper and constitutes an extension.

The FAIR metrics schemes developed so far, e.g. the FAIRsFAIR v0.4 (doi: 10.5281/zenodo.4081213), used by the F-UJI tool, are more adapted to evaluation of the machine-actionability structure of general data repositories rather than individual datasets or RDM measures in research projects. For this reason, we had to develop and adapt, largely, our own metrics scheme for the assessment of potential future FAIRness should the DMP be implemented fully as planned.

COMPARISON WITH OTHER APPROACHES

There are some advantages to an approach based on a local template, instead of trying to find a smallest common denominator in general funder templates in order to achieve compliance with the RDCS, as e.g. DMP Roadmap/DMPTool and tools like Argos seem to aim for.

An obvious advantage is the possibility to tailor output of DMPs to local RDM administration needs and policies, thereby allowing for at least semi-automated FAIR-assessment and potential self-evaluation. In this way, responding in part to the well-founded criticism by *Smale et al.* (2020) of DMPs hitherto not achieving their stated mission of serving the needs of different stakeholders.

There is also less need to make up dummy entries of input fields (e.g. datasets and identifiers) that are required output in the RDCS, but are not commonly part of general funder templates.

While our local SU-VR model, using DMP Roadmap/DMP Online API v0, conversion to XML, transformation with XSLT and conversion back to JSON actually produces RDCS compliant output that validates fully against the maDMP-schema-1.1, the current export to JSON within DMP Online, using instead API v1 does not yet fully validate against the maDMP-schema-1.1. Required keys, such as *contact_id* and *dataset_id* are missing and the data formats used, for example, for start and end date of project are incorrect, apparently irrespective of template used.
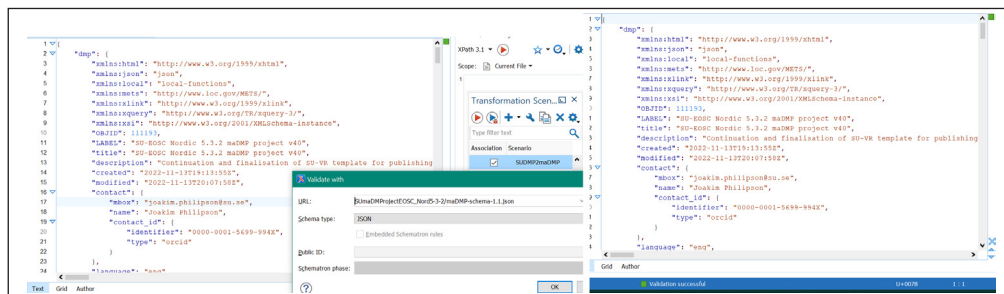

Figure 2 Successful validation against madmp-schema-1.1.json of DMP output created with SU-VR template v40.
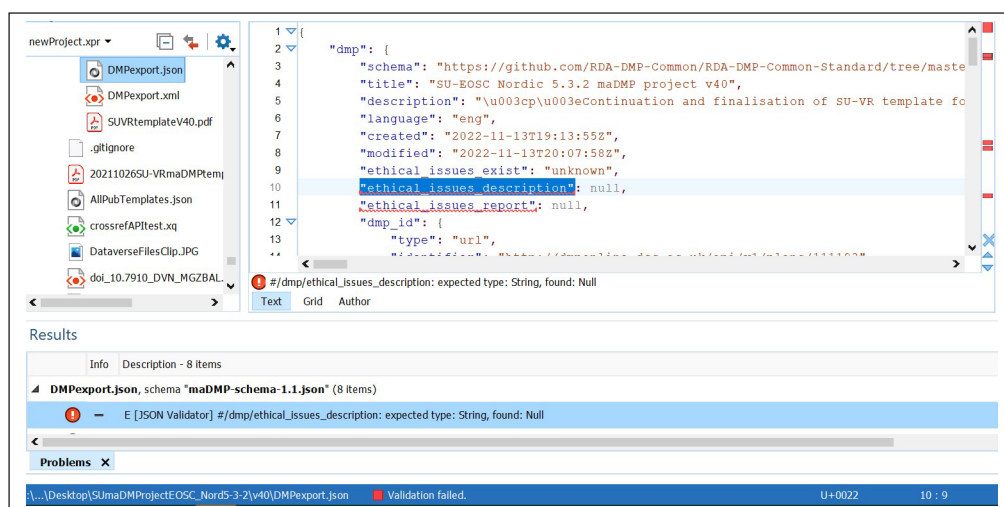

Figure 3 Failed validation against madmp-schema-1.1.json of direct DMPexport.json from DMP Online.

However, a negative aspect of the SU-EOSC Nordic maDMP model, compared to e.g. DSW and Argos OpenAIRE (below), is that the outputs and processed results for evaluation is only accessible to SU administrators of the DMP Online service, but not to the actual end users – but the SU researchers themselves, other than on demand.

All this, including the remarks about the current limited functioning of the API v1 (above), is said in the awareness that DMP Online is also developing a new tool for creating DMPs by means of a "wizard" (perhaps more in the fashion of DSW, below). This is to be presented together with a Research Outputs feature very soon at a user group meeting (https://www.dcc.ac.uk/events/DMPOnline-user-group-new-plan-creation-wizard). Perhaps these new functions in DMP Online might render at least parts of the SU-EOSC Nordic maDMP project and the SU-VR template features and processing steps obsolete. This remains to be seen.

DSW – DATA STEWARDSHIP WIZARD

The Data Stewardship Wizard (**DSW**) was initially developed for the European Infrastructure for Life Sciences (Elixir) project, and is heavily used by the life sciences community. It continues strongly to support DMPs for the life sciences, but also from a wider range of disciplines. The DSW calls DMP templates 'knowledge models' and DMP template designers (such as RDM administrators) create those 'knowledge models' for their domain that satisfies the relevant data management guidelines supplied by the institution and/or funding agency. The DSW allows for a hierarchy of knowledge models by inheriting from other models. 'Knowledge models' that meet the Horizon 2020 Science Europe and Elixir data management guidelines exist as well as many others. (Hasan et al. 2021)

As was observed earlier, the 'DSW includes a first implementation of a JSON export that is compliant with the RDCS schema' (ibid.).

The DSW has since then developed considerably. It has recently introduced pre-ingest validation of 'types of value questions like date, datetime, time, URL, email, or color. That leads to improved input fields in questionnaires, such as a date picker or warnings if an invalid value for URL or email is provided.' (*DSW Newsletter 2022–10–27*). Furthermore, DSW v3.15 was introduced as a project importer allowing users to prefill questionnaires with imported answers, which was a first for machine-actionable DMPs according to the RDCS in JSON. All answers in a DMP instance produced with DSW may also be exported from a project to a JSON file (via the Questionnaire Report template), and then imported into a different project, even in a different DSW instance.

A negative aspect of DSW to an end user might be that the inherent RDCS maDMP knowledge model is very comprehensive, with well over 200 questions to answer (compared to maximum 47 with the SU-VR template in DMP Online), where some of the questions appear to ask for duplicated content. It is also not obvious that all questions are relevant; some questions might have been removed based on previous answer options (as in DMP Online).

Although maDMP output as native JSON in DSW is also available to end users, the relationship between the extensive questionnaire and the rather meagre (86 lines in json) RDCS maDMP output is not obvious. For example, in our test use case it proved very difficult to, *post factum,* locate the missing answer in the *questionnaire* that makes the output fail validation against the RDCS madmp-schema-1.1.json in the key value pair *"ethical_issues_exist": ""*, given the fact that 217 of 217 questions are indicated as answered in the "current phase" and 266 /268 in total. The question is then: where are those two missing answers to be found?
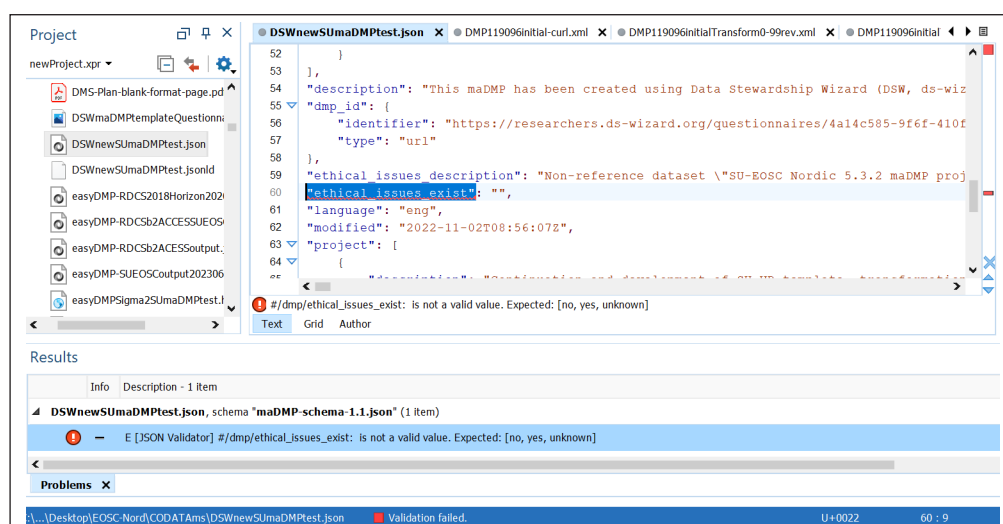

Figure 4 Failed validation of DSW maDMP against madmp-schema-1.1 due to missing answer.

EASY DMP

EasyDMP was developed by Uninett Sigma2 in Norway. At the *Sigma2* (2022) web site easyDMP is described as

> 'a service offering researchers with minimal experience in data management a simple way of creating a Data Management Plan (DMP). We do this by transforming any funding agency or institution's data management guidelines and policies into a series of easy-to-answer questions, many containing a simple list of canned answers you can easily choose from when creating your plan.'

The resulting DMP produced with easyDMP claims to be "machine-readable" and even "machine actionable" (*ibid*.)

EasyDMP is accessible free of charge to all researchers on the Sigma2 infrastructure (https://easydmp.sigma2.no/). European users outside Norway can log in and authenticate via the B2ACCESS portal. "OpenIdP login, such as Facebook, Google or Linkedin" is also supported.

The following is an extract of the description of EasyDMP in *Hasan et al.* (2021):

> One of the original goals of EasyDMP was to provide a simpler interface for creating DMPs with questions containing canned responses, or with controlled vocabularies. Another goal was to integrate with the services provided by the Norwegian Infrastructure for Research Data that is managed by Sigma2, such that a researcher would fill in the DMP, answering the appropriate questions for services. Once the plan is approved this will trigger the allocation of those services. In common with the other tools, a RDM administrator can create a DMP template for a community. New templates can be derived from existing templates through cloning. RDM administrators have the flexibility to create highly structured DMPs that are easier to transform into machine-actionable plans, or narrative plans that would be easier for a human to comprehend. EasyDMP currently provides templates for the Science Europe guidelines, Horizon 2020 and the Sigma2 guidelines for new projects.

EasyDMP started as a reaction to existing systems that needed domain specific knowledge on how to write an acceptable DMP, in an essay format. Instead of an essay, easyDMP uses a wizard format where the researchers choose among a small set of answers. The answers can be converted to an essay by inserting the answers into a frame, that is, a blank space to fill into a natural language full sentence. In easyDMP, a collection of questions, frames, and their allowable answers make up a template. A template is divided into one or more sections, and each section contains questions, their answer-alternatives according to answer type, and their frames. When a researcher has answered all the required questions of a template, the result is a plan, which is available both as an essay and as JSON-structure of question/answer pairs.

The actual answers are stored in a machine-readable format so they can be easily edited. This results in a new essay generated. Since the original use case was mass-generating essays, according to funders' needs, the answers are not machine-actionable because the question/answer pairs do not carry any semantics/metadata.

RELATIONSHIP BETWEEN EASYDMP AND RDCS

Hasan et al. (2021) further described the complex relationship between the design of easyDMP and that of RDCS:

> "While EasyDMP and similar funder-driven systems are organized in a linear, very open structure, completely in control of the template designer, the RDCS is structured around objects containing other objects that are not designed to be converted into free-flowing text. According to the standard, a plan contains at least one contact, one or more datasets, zero or more contributors, zero or more projects, and zero or more cost-objects. Each project-object may contain zero or more founding objects, each dataset may contain zero or more distribution-objects etc.

In order for EasyDMP to be able to interact with plans built according to RDCS, it was necessary to adapt it to support the objects-within-objects structure. This was done by adapting the sections: these were changed to be able to be optional, to be able to contain other sections, and to be answered more than once. This stage of the adaptation is complete. The next stage is to design a complete template following RDCS, and build a system to import and export plans using this template.

Once there exists one such complete plan, the framing-system can be adjusted so that an RDCS plan can be turned into an essay format. When this stage is finished we hope new templates will use the RDCS template as a basis so that the standard is followed by default."

The EasyDMP tool when tested recently (via B2ACCESS portal, 2022–11–08), both with the Science Europe (SE v2) and the Sigma2 (Sgma2DMP) templates now provide export of output in addition to PDF, HTML, and as RDCS compliant JSON, thus offering the possibility of validating output against RDCS madmp-schema-1.1.json. However, among the templates in use (Horizon 2020, Science Europe, Sigma 2 RFK and institutional templates) none (of those accessible to all users) appears to have questionnaires formatted to give answers to questions about e.g. dataset identifiers that are required (mandatory) in RDCS. For this reason, the RDCS compliant json export from EasyDMP contains dummy answers to some of these questions. (The same "trick" using "dummy" values for compliance with RDCS is not unique to EasyDMP. It is also used in other DMP tools and templates, e.g. Argos and DMP Online API v1). The user documentation of EasyDMP currently appears to lack of information about an API to get the DMP output in machine-actionable form. (*EasyDMP User Documentation* 2022)

A new version of EasyDMP was released at the end of November 2022, which allows users to *import* plans, in both RDCS format and EasyDMP's own format.

The short-term goal for EasyDMP is to finalize the import and export of plans, conforming to the RDCS schema, which also requires a RDCS conformant DMP template in EasyDMP. Part of this goal is to allow plans conforming to the RDCS schema to be exchanged between the ELIXIR DSW and EasyDMP. The mid-term goal is to carry out an EasyDMP v2 project to provide a DMP tool that aims to "...better support FAIR data and meet the requirements of the research and education sector" (*Sigma2* 2022).

SND Checklist

The SND checklist is an extensive document, which can be downloaded in both Swedish and English (*SND Checklist for Data Management Plan* 2021). Formats are pdf and for an editable version as docx or rtf. The checklist is updated regularly and complies with, among others, the recommendations from Horizon 2020 and Science Europe. The Association of Swedish Higher Education Institutions (SUHF) and the Swedish Research Council also follow this.

The latest version of the SND checklist was released in 2021. The resulting documents from filling out the checklist are still not machine-actionable. The intention has been to map the latest version of the checklist to the EasyDMP and DSW tools during the EOSC Nordic project, which would then allow export of RDCS compliant plans. As of end November 2022, this remains to be implemented.

OTHER TOOLS AND PROJECTS

There are a number of other tools and projects globally professing to create maDMPs, some of which are listed at the activeDMPs website (https://activedmps.org/). We have not been in a position to evaluate and try them all out. Below follows just a short description of a small selection of these tools and projects.

ARGOS OPENAIRE [https://argos.openaire.eu]

This DMP tool has come quite far in provisioning for integration with datasets (e.g. through import of metadata from Zenodo.org, and soon from other data repositories as well, as we learned during a community call session 2022–10–26). Although there is clearly an ambition to produce machine-actionable output that complies with the RDCS through a direct export

format option "RDA JSON", presently the resulting output file sometimes does not appear to validate completely against the RDCS madmp-schema-1.1.json. For example, a proper formatting of *roles* seems to be lacking, and sometimes the same appears to be the case for the start and end dates of projects. Like some other DMP tools, Argos also appears to use for the occasion "made-up" identifiers for datasets and other items that are required in the RDCS, but are lacking corresponding entries in the DMP funder template questionnaires.

But Argos continues its development, announcing that the next *Argos Community Call* (2022–12–14) will focus "on new features that are to be released in ARGOS: a. Machine actionable table input in DMP Templates, and b. Deposit in many repositories for OpenDMP instances."

This turned out to be a very promising plan with the potential to set a new standard for DMP tools to produce truly machine-actionable and RDCS compliant DMPs, while still being very flexible and adaptable to local needs, also allowing for prefilling of metadata by means of APIs *(Argos Community Calls,* 2022). In particular, the possibility for admins of the tool to add special dataset table templates with repeated entries and mapping directly to RDCS compliant json output, as an integral part of general DMP funder template, will reduce or make obsolete the need for using "dummy" values to comply with RDCS required entries for e.g. datasets and identifiers.

DAMAP (https://damap.org)

DAMAP describes itself in a README file as a tool that is currently being developed by TU Wien and TU Graz as part of the FAIR Data Austria project. It is based on the idea of machine actionable data management plans (maDMPs) and aims to facilitate the creation of data management plans (DMPs) for researchers. The tool aims to be closely integrated into the institutional environment, collecting information from various established systems, to perceive project information, research data and personnel data from existing systems. This saves DMP authors from having to enter the same data several times. Finally, DAMAP delivers both a DMP that can be read and edited as a Word document, and an maDMP whose information can be used at machine level. The current content of DAMAP is based on Science Europe's Practical Guide to the International Alignment of Research Data Management and is compatible with the RDA recommendation on machine actionable DMPs. (*DAMAP Readme* 2022)

Thus, in common with DSW and DMPTool (below), DAMAP appears to invest heavily into integration with other research information systems to enable reuse of data from other sources. This, however, makes for a rather complex system where it is difficult for a lay outsider, who is not a systems developer, to view individual parts such as templates or questionnaires without first installing the whole system. It also seems that to the end user of DAMAP, output of DMPs is only delivered as Word documents, while the maDMP output (presumably in json) is only for machine consumption.

DMPTOOL (CDL) (https://dmptool.org/)

This is a federated service hosted by California Digital Library, leading partner and developer of the DMP Roadmap (shared with DMP Online and others), with a pronounced ambition to integrate with other research information systems as electronic lab notebooks, like RSpace and data repositories. "This new integration enables tri-directional data flows between RSpace, DMPTool, and data repositories, facilitating higher quality and more comprehensive research data capture and tracking." (DMPTool Blog 2022–01–14) This integration work represents a conscious effort to promote the publication of **FAIR** research data, with particular focus on **I**nteroperability and **R**e-usability.

The look and feel of DMPTool, judged from outside without being able to sign in, is otherwise similar to that of DMP Online, not surprisingly so since they share a common DMP Roadmap, where DMPTool seems to have advanced further than its partner.

GO FAIR AND LEIDEN UNIVERSITY: FIP2DMP

Representatives of the GO FAIR foundation and Leiden University are exploring the possibility of integration of a DMP with a so-called FAIR Implementation Profile (**FIP**), that is "a list of technology choices, so-called FAIR Enabling Resources, declared by a community to implement each of the FAIR principles" (Hettne et al. 2022). On a more specific level a FAIR Enabling Resource (**FER**), is a "machine-readable output of internal agreements represented

as a nanopublication containing an assertion linked to provenance and publication information for each declaration made by members of this community." (ibid.) **FER**s are thus machine-readable nanopublications accessible through purl-links that are used in **FIP**s to describe e.g. choices of Persistent Identifiers (PIDs) (like DOI [link], ORCID [link], ePIC [link] or Handle [link]). Mapping the questions and answer options of a DMP to those of selected FIPs would then allow the DMP form to be at least partly pre-filled, and in this way, more machine-actionable. *Hettne et al.* (2022) describes how this can be implemented, at least in part for the Leiden University DMP template, by using a specific knowledge model in the DSW tool.

When tested at this early stage of development (2022–10–31), the JSON output of the DMP created with this knowledge model (*KM for Leiden FIP2DMP, 0.0.5*), exported (via the *Questionnaire Report template*) was still far from validating against the RDCS madmp-schema-1.1.json, with at least nineteen error items. This might be due to the fact that the Leiden knowledge model, at this point, does not seem to provide access to the *maDMP (RDA Common Standard) 1.13.1* document template for export as an alternative to the *Questionnaire Report template*. The possible results of the validation against the RDCS madmp-schema-1.1.json will be more accurate with the use of the maDMP document template for export of output.

However, as indicated, this is still work in progress and thus not yet ready to be evaluated properly. The results are heavily dependent on the machine actionability of the DMP template used. The FIP2DMP team is still working on the enhancement of the Leiden University DMP template to enable a smooth mapping to the FIPs.

SND PROJECT: CHALMERS UNIVERSITY AND KTH

Within the framework of the *Swedish National Data Service* (**SND**), *Chalmers University of Technology* and *KTH Royal Institute of Technology* have a joint "flagship project" exploring possibilities with maDMPs in DSW and DMP Online (SND 2022). A main objective of this project, as we understood it, (*Personal communication* 2022–09–15) is integration with other information systems (CRIS, data repositories, local data management systems etc.) by means of APIs developed specifically for these purposes. This remains yet to develop fully, but when we tested the staging instance for compliance with the RDCS, the results were promising, with only two validation errors against the madmp-schema-1.1.json. One of these errors, the missing required value for key "ethical issues exist" (yes, no, unknown) might be a bug inherited from the *maDMP (RDA Common Standard) 1.13.1* document template for export, which returns the same error in the generic DSW knowledge model (see above). The other error detected, the missing keys for "contact" should not be too difficult to fix "internally", by adjusting the mapping from contributors and role list accordingly, allowing also multiple choice of roles for the same contributor.

The SND Chalmers and KTH project envisages integration and automation to the fullest extent possible of data flows from e.g. DMP to repositories and metadata catalogues (here the SND DORIS system), enabled by an API that produces RDCS madmp-schema-1.1.json compliant output, which can then be searched on ORCiD ID, ROR Id, e-mail etc. on the receiving end (e.g. SND DORIS). This kind of "pull-integration" is made possible on the assumption that the creators of DMPs, the researchers themselves, initially give their consent to the reuse of information (metadata) in their DMPs by other information systems. Their benefit from such a consent is naturally to avoid having to fill in the same information/data, repeatedly, in several different information system questionnaires.

DISCUSSION

We have seen that different DMP tools, to a varying degree, seem to focus on:

1. End-user (researcher) friendliness (e.g. by ease of use, number and type of questions and answers in questionnaires, guidance, prefilling/automation options, integration with other research information documents).

2. Adaptation to funder and / or institutional administrative interests (e.g. completeness and level of detail of DMP content, facilitating review and evaluation of DMPs by semi-automatic means).

3. Machine-actionability and compliance with the RDCS.

It may be difficult to strike the right balance between these three stakeholder interests, although to some extent all stakeholders might benefit from a higher level of automation and integration with other research information systems (e.g. CRIS systems, ethical vetting etc.). From our brief overview it seems Damap, DSW, and the SND Chalmers-KTH project have the most far-reaching integration and automation ambitions. However, at least with some knowledge models in DSW, this seems to come at the expense of ease of use and too many questions to answer in the DMP questionnaire, even if a good number of these might be prefilled or manually imported from other information sources. It also appears that the possibility of inferring answers to some questions, from other questions and answers is missing in some of these extensive DSW knowledge models, where sometimes the same answers have to be repeated several times. An example of such a possible *internal integration* could be, to give the end user the opportunity to answer a question about metadata standards simply by inference from the choice of a known repository, where metadata is best created and through which data is made accessible. From our experience, researchers are often not familiar with different common metadata standards. That is why we try to make it easier for them to answer such a question simply by ticking a box in the SU-VR template. We believe this is for mutual benefit because in this way we, as RDM administrators or reviewers, probably get more precise information about the metadata standards that are available for a given dataset than what might be expected otherwise. Possibly, it also contributes to an early assessment of the potential level of FAIR-ness resulting from the data management measures described in the DMP.

In the given example, a question about metadata standards could also be answered by means of *external integration* of DMP questionnaires e.g. with FIPs or directly with data repositories, but that would probably still demand more from the end user than the simple ticking of a box.

Another burning issue regarding DMPs, discussed recently at a *Community cross-fertilisation workshop: RDA for Data Management Planning* (2022–11–10), is how to make evaluation and review of DMPs scalable with a potentially large increase of DMPs to be scrutinised. This also involves the question of later follow up on the fulfilment of the professed data management measures stated in a DMP. One of the objectives of maDMPs could be to help in this process by enabling to the extent possible (at least semi-automation) of evaluation, review, and follow-up. This has already, from the start, been part of the SU-EOSC Nordic 5.3.2 maDMP project by developing a validation schema for evaluation of potential FAIR-ness met by the full implementation of the data management measures described in a DMP created with the SU-VR template. Subsequent archiving at two different points in the DMP lifecycle, initially at time of approval and finally at project end, will allow for assessment of possible improvement and increase in FAIR score during the elapsed period between these two events. The next step then should be to check-up on the fulfilment of promises made in the DMP by accessing the repositories and other information systems referred to in the DMP and see that the datasets and items described there can also be found, where they are supposed to be.

CONCLUSION

Development of maDMP tools is generally progressing through increased integration opportunities with other research information systems, and convergence to a common standard that might prove to be a revised and enhanced version of the RDCS. It remains yet to be seen if integration with FIPs will also play a proper part in that convergence process.

For the SU-EOSC Nordic 5.3.2 maDMP project, with its SU-VR template limited by the framework of DMP Online, questions and answer options are on a more specific level. These are often tailor-made to validate against the RDCS maDMP-schema, while FIPs tend to be more general in their perspective, with questions and answers listing resources used by a certain research community. Therefore, we foresee that a mapping directly to FERs rather than to FIPs will be more feasible here. The FERs would then only be made accessible "under the hood" as part of the processing to make the transformed output of the DMP itself more machine-actionable and FAIR, as an extension of the present RDCS. As a first step in this direction, in the latest version of the transformation file *SUDMP2maDMP1-1.xsl* (v0.98), published as part of the dataset in *Philipson* (2022), we have included the links to three FER nanopublications that describe DOI, Handle, and ORCID identifiers. These links will then appear as part of the transformed output files (xml and json) of any DMP that can be processed with our project tools, thereby at least potentially serving to enhance machine consumption of these output files.

One lesson learned from several of the DMP tools described above seems to be the essential role of the DMP *templates* used for achieving truly machine-actionable DMP output. The responsibility for these templates largely still rests with the *funders* of research. In any event, the further progress in the development of better, smarter, integrated maDMP tools, converging to a common standard should serve the multiple purpose of benefitting most stakeholders – researchers, funders, reviewers, and institutional administrators alike.

DATA ACCESSIBILITY STATEMENT

The RDCS madmp-schema-1.1.json that was used for validation of maDMPs is found through a direct reference below, and is also available together with examples of processing and processed files in another reference:

Philipson J (2022) SU-EOSC Nordic 5.3.2 maDMP project. Harvard Dataverse/Stockholm University Library Dataverse. https://doi.org/10.7910/DVN/MGZBAL.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

Joakim Philipson orcid.org/0000-0001-5699-994X
Stockholm University, Sweden

Adil Hasan orcid.org/0000-0003-2373-7112
Sigma2 AS, Norway

Hanne Moa orcid.org/0000-0002-6635-0033
Norwegian Agency for Shared Services in Education and Research (Sikt), Norway

REFERENCES

Argos Community Calls. 2022.[Available at: https://www.openaire.eu/argos-community-calls].

Community cross-fertilisation workshop: RDA for Data Management Planning. (common notes from RDA DMP WG meeting 2022–11–10). [Available at: https://docs.google.com/document/d/1_hXZpPg5cWfl-msbYKQgjFhcpcwQ9uMlaMg-ic04NMY/edit#].

DAMAP Readme. 2022. [Available at: https://github.com/tuwien-csd/damap-backend]

DMP Online. 2022. [Available at: https://DMPOnline.dcc.ac.uk/].

DMPTool Blog. 2022–01–14. [Available at: https://blog.dmptool.org/2022/01/14/supporting-fair-data-with-integrations].

EasyDMP User Documentation. 2022. [Available at : https://documentation.sigma2.no/services/easydmp-user-documentation.html] [Accessed 2022–10–31].

Hasan, A, Meerman, B, Moa, H, Jakobsson, U and **Philipson, J.** 2021. D5.5 Actionable Modular Data Management for Brokering Distributed Resources. *Zenodo*. DOI: https://doi.org/10.5281/zenodo.5713787

Hettne, K, Magagna, B, Gambardella, A, Schultes, E and **Suchánek, M.** 2022. FIP2DMP: Linking Data Management Plans with FAIR Implementation Profiles. *FAIR Connect*. [Available at: https://docs.google.com/document/d/13XO4cHnEYfyYqDY4sQ07lEkIf5EDPWeMEIW00qtDBMw/edit#]. DOI: https://doi.org/10.3233/FC-221515

maDMP-schema-1.1.json. 2020. [Available at: https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard/blob/master/examples/JSON/JSON-schema/1.1/maDMP-schema-1.1.json].

Miksa, T, Simms, S, Mietchen, D and **Jones, S.** 2019. Ten Príncipes for Machine-Actionable Data Management Plans. *PLoS Comput Biol*, 15(3): e1006750. [Available at: https://doi.org/10.1371/journal.pcbi.1006750].

Philipson, J. 2021. Presentation of Stockholm University – EOSC Nordic maDMP project at DMP Online Dropin session 2021–10–26. [Available at: https://doi.org/10.5281/zenodo.5802060].

Philipson, J. 2022. SU-EOSC Nordic 5.3.2 maDMP project. *Harvard Dataverse/Stockholm University Library Dataverse*. DOI: https://doi.org/10.1371/journal.pcbi.1006750

RDA DMP Common Standard for Machine-Actionable Data Management Plans. [Available at: https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard].

Sigma2. 2022. *Data-planning*. [Available at: https://www.sigma2.no/data-planning] [Accessed 2022–10–31].

Smale, N, et al. 2020. A Review of the History, Advocacy and Efficacy of Data Management Plans. *International Journal of Digital Curation*, 15(1). DOI: https://doi.org/10.2218/ijdc.v15i1.525

SND Checklist for Data Management Plan. 2021. [Available at: https://snd.gu.se/en/manage-data/guides/dmp-checklist].

SND Launches Flagships …. 2022. Published: 2022–03–17. [Available at: https://snd.gu.se/en/news/snd-launches-flagships-and-targeted-initiatives-two-pilot-projects-year].

Wilkinson, M, Dumontier, M, Aalbersberg, J, Appleton, G, Axton, M, Baak, A, Blomberg, N, et al. 2016. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3(March 2016): 160018. DOI: https://doi.org/10.1038/sdata.2016.18