

Final Report

Integrating the Metadata Standards Directory into DMPonline

December 2017

Sarah Jones, sarah.jones@glasgow.ac.uk
Jimmy Angelakos, Jimmy.Angelakos@ed.ac.uk

Executive Summary

This project intended to develop a transparent integration between one leading data management planning tool (the Digital Curation Centre's DMPonline) and the RDA Metadata Standards Directory. We have used the API provided by the RDA Metadata Standards Working Group to offer users the ability to browse, search and select metadata standards to use. Organisational administrators within DMPonline can configure this option to include it as a way to answer metadata questions in their own DMP templates, and we have deployed this option in the generic DMP template used within DMPonline to profile how others can implement it.

Feedback from testing has been positive, with users welcoming the addition and finding it intuitive to use. Some bugs were identified and resolved, and suggested improvements have largely been acted on or are planned as part of future work. We are very grateful to RDA Europe for this award to develop an initial proof-of-concept to aid in the transition to machine-readable DMPs. While the project has been ongoing, we have liaised with other database and service providers such as FAIRsharing and re3data to plan similar integrations in future.

Objectives

Data management planning of some form has been accepted good practice for many years as a means to ensure that data outputs are more likely to be discoverable, reusable and preserved. A common question posed in DMPs relates to the data and metadata standards to be used. This is often one of the more challenging questions for researchers to answer as they are not always aware of available standards.

The Metadata Standards Directory (MSD) is an RDA output to help researchers find appropriate standards to use. By integrating the catalogue into DMPonline we aim to raise awareness of existing standards and improve researchers' responses by offering a standardised range of options that could be selected when answering such questions. Using a pre-defined list has the added advantage of generating machine-actionable data that can be more easily analysed and exploited than the typical freetext responses.

The project had a number of specific objectives, all of which have been realized:

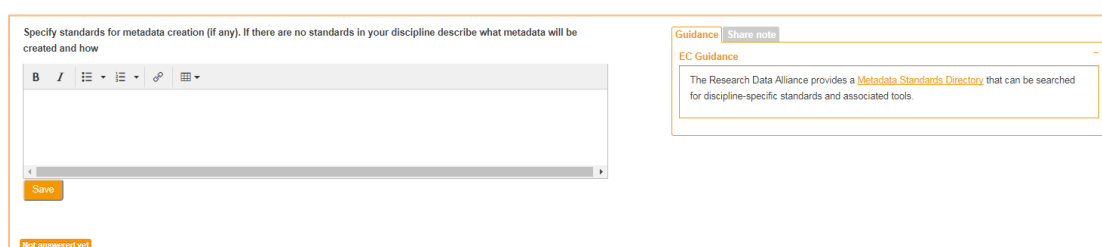
- demonstrate the feasibility of integrating machine-readable knowledge from the MSD into other systems
- produce DMPs more suitable for machine analysis and action, a goal of the RDA Active DMP Interest Group

RDA Europe 3 - Collaboration Project Final Report

- demonstrate the utility of integration between DMP tools in general with other similar catalogues and systems
- researchers will be made aware of relevant metadata standards
- funders, institutions and other administrators will be able to perform compliance checks and analyse responses to DMPs more easily
- information on any other standards being used by researchers that aren't currently listed will be shared with the RDA Working Group to further enhance the MSD

Initial State

Before the project, DMP questions relating to metadata were typically broad, freetext responses. An example from the European Commission's Horizon 2020 DMP template is provided below.



The screenshot shows a web form with a text area for specifying metadata standards. The text area contains a rich text editor toolbar with options for bold, italic, list, link, and table. Below the text area is a 'Save' button. To the right of the text area is a 'Guidance' box with a 'Share note' button. The 'Guidance' box contains the text: 'EC Guidance' and 'The Research Data Alliance provides a [Metadata Standards Directory](#) that can be searched for discipline-specific standards and associated tools.' At the bottom left of the form, there is a status indicator that says 'Not answered yet'.

As noted under the Objectives section, this is not optimum for researchers as they are often unaware of which standards can or should be used. Integrating the catalogue so users can browse this directly within the interface aids more useful responses.

Project Outcomes

The project had a few key activities. Initially we gathered requirements by consulting with user communities and the RDA working Group. We then undertook research to develop an integration plan and used the API to expose MSD content within DMPonline. This work has been delivered in a way that enables other external directories to be integrated in a similar fashion, and part of this project was investigating the possibilities for connecting up with FAIRsharing. The final stage was testing this with the user community and making adjustments in response.

Module 1: requirements gathering

The main consultation was done at the International Digital Curation Conference (IDCC) in Barcelona. We presented a poster in which we outlined the problem and how integrating the MSD would solve this. Options were given for different ways to display and use the content. User feedback indicated that it was important to be able to search the standards, as well as browse via subject listings. It was agreed that the MSD should be provided as a way in which to answer the question, and the option was also felt to be best configured by template owners.

Module 2: integrating the MSD

The first step in integrating the MSD was to review how the data was structured and what was provided via the API. The main developer assigned to this, Sam Rust, started by investigating the calls and data available to the MSD API, and those of re3data and FAIRsharing to ensure the implemented solution would work down the line for those applications. A new type of question was added for remote which would generically hold the address of the API and its title. To store the data, we are using JSON so we can flexibly adapt to the different conventions of various APIs. Finally, a library of JavaScript was created to run multiple queries on API to generate a single data object and display the collected data to any applicable questions on the page.

RDA Europe 3 - Collaboration Project Final Report

Changes were required on the main user interface to allow DMP authors to browse the directory and select standards, as well as on the administrative side of the tool to enable organisational administrators to configure this option if they wish to apply it to their templates. We worked with our usability team to define how best to present the content. Users are given options to browse or search standards. The browse function can be refined by selecting disciplinary areas from the dropdown selectors, and the open search bar is a type-ahead so potential results display after one or more letters is keyed. Users are also able to record an 'Other' value if their standard is not listed. This data will be fed back to the RDA Working Group to enrich their catalogue. A freetext box is provided to allow the user to provide additional descriptive information, and one or more named standards can be selected.

The screenshot shows a web interface titled "Documentation and Metadata (1 question, 0 answered)". The main heading is "What documentation and metadata will accompany the data?". Below this, there is a section for "Your Selected Standards:" with a dropdown menu set to "Multidisciplinary" and an "OR Search:" input field. A "Browse Standards" section is visible, listing three standards: "CERIF (Common European Research Information Format)", "Data Package", and "DataCite Metadata Schema". Each standard entry includes a brief description and an "Add Standard" button. A "Standard not listed? Add your own" link is also present. At the bottom of the main form is a "Comment" section with a rich text editor and a "Save" button. On the right side, there is a "Guidance" panel titled "DCC Guidance" which contains "Questions to consider" and "Guidance" text.

Image 1: How the MSD content can be searched and browsed on the DMPonline user interface

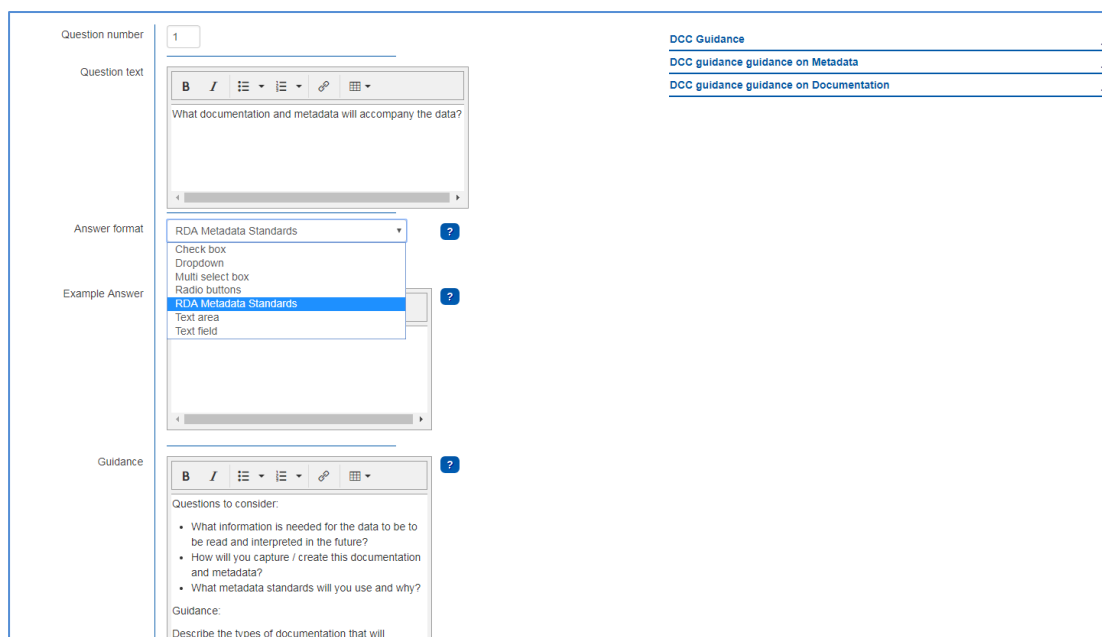


Image 2: On the admin interface, the MSD option is provided as a question format which can be selected when creating (or editing existing) templates.

As part of this work we also engaged with Susanna Sansone and Pete McQuilton of the FAIRsharing initiative.¹ We had a few calls and discussions around the RDA Plenaries to scope out what information could be drawn from FAIRsharing and how to expose it in DMPonline. At present the FAIRsharing API only presents the database directory so this integration would be most useful for questions relating to data sharing, repositories and preservation plans. We're awaiting access to the API so the integration couldn't be delivered within the project timeframes. The model we have developed is intended to suit multiple API sources though, so this work can be undertaken as part of our core DMPRoadmap workplan.

Module 3: testing

We used the existing DMPonline user group to test the integration. This is a very diverse, international group and feedback was received from institutional support staff at UK universities, as well as from France, Germany, Denmark and the USA. Overall the responses were very positive. Users found the functionality intuitive and liked the ability to browse the standards:

"Thank you for this helpful integration. I find it very useful, intuitive and the standards are easy to add. My researchers will be really happy about this!"

"The functionality is intuitive for users and administrators.... We appreciate the 'autosaving' functionality."

"I think the option to choose 'metadata standards directory' should be in all templates by default together with the specific one. I think this could enrich more the DMP, since some of the funders specifications are not very precise. And in any case, many times researchers have no idea where to search this kind of info."

There were naturally some issues to troubleshoot too. Users identified a bug with Firefox which affected the loading of standards due to a failed security certificate. There were also some criticisms of the way in which standards were classified.

¹ <https://fairsharing.org>

RDA Europe 3 - Collaboration Project Final Report

“Generally I found this fairly intuitively and useful, although I was a little surprised to find engineering classified under politics, law and economics.”

“While comparing with the RDA standard metadata directory, we noticed that the subject classification differs.”

The RDA working group has changed its classification approach, moving away from Joint Academic Subject Headings (JACS) which offered six very broad subject areas to the UNESCO Thesaurus. The latter is an international standard and already in linked data form so the new Metadata Standards Catalogue will be semantic web friendly. Given that we encountered some usability issues with the UNESCO headings, we plan to investigate this further and potentially revise the display back to the JACS approach.² Alex Ball has offered to add support for Boolean 'OR' in the API, so we could define our own top level headings.

Various other suggested improvements were made:

- Link out to the RDA directory to enable users to find out more about the standard, its use cases and tools.
- Increase the size of display of “Your selected standards” so it isn’t overlooked
- Make the comments box more obvious/useful to encourage researchers to fill it in e.g renaming it ‘additional information and comments’?
- Retrieve (or at least visualize) the metadata fields
- Retain (rather than reset) the dropdown selectors once a standard is chosen
- Allow multiple comment fields to address each standard separately
- Do a similar integration to the metadata standards list with re3data.
- Make recommendations based on selections e.g. the choice of a repository triggers the selection of persistent identifiers or relevant metadata standards and tools. Or knowing the institution and discipline provides filtered lists of repositories and standards.

The top three changes are very minor display issues so will be undertaken as we migrate this feature to the DMPRoadmap codebase. The next three have been deemed to be out of scope or undesirable from a usability standpoint. The final two comments align with our future directions with this work, so while out of scope for this project, are a good indication of what we hope integrations like this will eventually allow.

Dissemination Activities / Publications

We have spoken about the RDA Europe Collaboration Award at many events. The main ones are listed below, together with the context in which the presentation was made. At each of these we specifically gathered user requirements, provided progress updates or demonstrated outputs. We also referenced this Collaboration Award and the associated activities in many other talks as part of our ongoing outreach and dissemination e.g. an Active DMP presentation given at CERN.

#	Event and activity	Date	Publication
1	“Connecting DMPonline with the Metadata Standards Directory” poster ³ at IDCC, Amsterdam, to gather user requirements	23 Feb 2015	Poster

² See for example: <http://rd-alliance.github.io/metadata-directory/standards> where the six following subject headings used are: Arts & Humanities; Engineering; Life Sciences; Physical Sciences & Maths; Social & Behavioural Sciences; General Research Data.

³http://www.dcc.ac.uk/sites/default/files/documents/IDCC16/DMPonline%20IDCC16%20poster_v5.pdf

RDA Europe 3 - Collaboration Project Final Report

2	Progress update at the RDA UK event ⁴ as part of the Metadata session showing how RDA outputs are being used	2 Nov 2016	Presentation
3	“A postcard from the future: Tools and services from a perfect DMP world” workshop at IDCC, Edinburgh ⁵ to gather requirements for machine-actionable DMPs	20 Feb 2017	Workshop
4	Demonstration of project results in Active DMP IG session at RDA Plenary in Montreal ⁶	20 Sept 2017	Presentation
5	Announcement and subsequent discussion on the DMPonline user group listserv related to testing the new functionality	Oct/Nov 2017	Emails
6	Roll-up roll-up, get yer DMP update here ⁷ providing a public notice and call for testers	20 Oct 2017	Blog post

It is worth noting that during the period of the project, the Digital Curation Centre joined up with the California Digital Library to converge on a single codebase from which to operate both of our services. As the MSD integration will be part of the joint DMPRoadmap codebase, the potential uptake is much higher than could be envisaged at the proposal stage. The functionality will be deployed in each tool, reaching over 30,000 users in the UK, USA and Europe, and we have utilised this option in the generic DMP template used within DMPonline to profile how others can implement it. Other instances deploying our codebase will also benefit from these additions. A current list of these is available at: <https://github.com/DMPRoadmap/roadmap/wiki/Local-installations-inventory>

The DCC also now provides the infrastructure and hosting for the RDA Metadata Standards Directory. We were involved in constructing the initial catalogue and moved it to a broader community initiative via the Research Data Alliance WG structures. The current WG has been using GitHub but was looking for a host over Summer/Autumn 2017, and since we were already actively doing work on the Directory as part of this project, DCC agreed to use our existing developer effort to do this. It is now available at: <http://rdamsc.dcc.ac.uk>

Summary & Conclusions

Although the project ran for longer than first envisaged due to staff changes, codebase switches in the DCC/CDL join-up, and delays defining a contract with the Max Planck Institute, the activities and outcomes have been delivered in full and our goal to improve tool usability by integrating external registries into DMPonline has been validated. The outputs not only act as a proof-of-concept but are suitably robust and accepted by the user community to be part of active services. We have deployed this on DMPonline and it will be available in DMPTool when they migrate to the joint DMPRoadmap codebase in early 2018. This work ties into a much bigger stream of activity for us related to machine-actionable DMPs. It has kick-started an approach to integrate other related databases and we look forward to applying this model with FAIRsharing and others in the near future.

⁴ <https://www.rd-alliance.org/rda-workshop-jisc-2-november-2016-birmingham-uk>

⁵ <http://www.dcc.ac.uk/events/workshops/postcard-future-tools-and-services-perfect-dmp-world>

⁶ <https://www.rd-alliance.org/rda10-active-dmp-domain-repositories-joint-session> slideset 7

⁷ <http://www.dcc.ac.uk/blog/roll-roll-get-yer-dmp-update-here>