

Summary of the activities of the Metadata Standards Directory Working Group

Alex Ball

University of Bath

DC-2015: Current Developments in Metadata for Research Data,
3 September 2015

Abstract

The Research Data Alliance Metadata Standards Directory Working Group concluded its activities at the organization's Fifth Plenary in March 2015. This presentation explains the goal of this project, and reports on the design, implementation, and future plans for this tool.

1 Motivation

I don't need to tell anyone at a Dublin Core conference that Standards are a Good Thing BUT *they are only a Good Thing* if everyone uses the same standards:

If there are Not enough standards → everyone does their own thing
but if there are Too many *incompatible* standards → things get stuck in silos

Either way, you end up with duplicated work, and missed opportunities for tools because the developer effort is too diffuse. The ideal of having one standard for a given purpose – that everyone knows about, contributes to and uses – is very hard to achieve, which is why Dublin Core advocates an application profile approach. In a nutshell, this means asking yourself §: Is there a standard that suits my purpose?

- If there is, that's great, use it.
- If there's one that almost fits the bill, still use it, but adapt it to suit your needs, perhaps by being more strict about what goes in certain fields.
- If there isn't, that's the point at which to consider writing your own scheme, though there might be generic standards one could borrow from and specialize.

Apologies if this sounds obvious, but the point is, you can't answer that question without a good knowledge of all the metadata standards out there. And if you don't have that knowledge, where can you go to find it?

2 Purpose of the Working Group

This is where the Research Data Alliance's Metadata Standards Directory Working Group comes in.

1. *The purpose of the group was to Develop an RDA Metadata Standards Directory listing standards relevant for research data that would be*
 - Comprehensive, *covering all disciplines and most generic applications, and*
 - Easy for anyone to contribute to or update
2. *The group also decided to Define and develop use cases for research metadata, to help with organizing standards within the directory, and to*
3. Develop a plan for long-term growth and maintenance of the directory

¶ *The Stakeholders for this directory include*

- researchers
- data managers
- data scientists
- research support staff
- tool developers
- repositories
- agencies
- societies

...in fact most of the Research Data Alliance's growing community.

We have engaged with them in various ways...

- *Two years ago at the Dublin Core Conference in Lisbon we held CAMP-4-DATA: This event attracted 26 participants from 15 countries and covered a range of scientific metadata issues, from infrastructure models and frameworks, through usage and tracking of metadata, to common fields to associate with persistent IDs.*
- *I gave a presentation at the International Digital Curation Conference 2014, in San Francisco*
- *And at about the same time, my colleague Jane Greenberg presented our work at the RDA-EU Working Group Core Meeting, 2014, in Garching, near Munich*
- *We also made use of our Working Group mailing list, which is accessible from the RDA website*

3 Activity

Before forging ahead with creating its own directory, the Working Group stayed true to its principles and looked for lists and catalogues of metadata standards that others

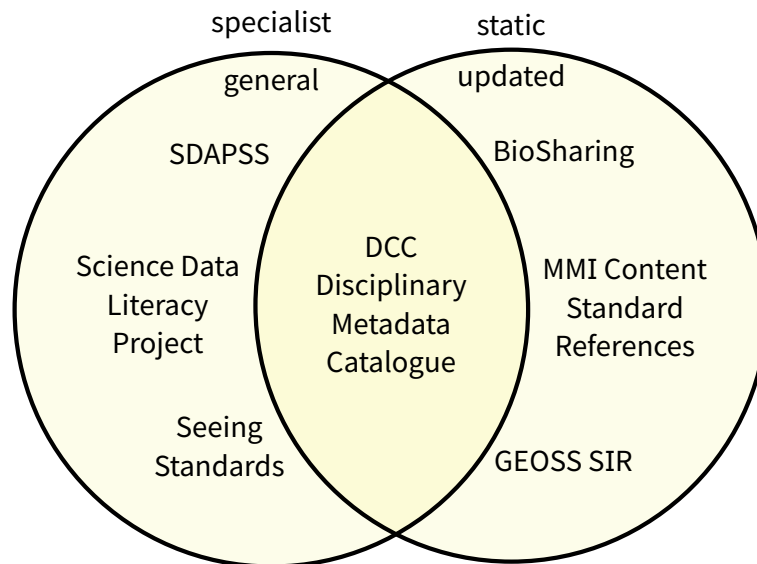


Figure 1: Venn diagram showing pre-existing lists of metadata standards.

had already assembled (¶ Figure 1). What stuck out was there was already a resource that was approximately what the group had in mind: the DCC Disciplinary Metadata Catalogue.

But it wasn't a perfect fit. ¶

Why? DCC directory was

- tightly focused on metadata-for-reuse *rather than contextual or discovery metadata*;
- aimed at *the UK Higher Education Sector*;
- *and it was* mostly the work of two people, *meaning it was incomplete* in some disciplines.

So the first step was to solve these problems by expanding the coverage of the directory. How?

- *We ran a Survey to enhance the directory: you can still view it at <http://bit.ly/1fToaqd> but we don't advise you use it any more as there are better ways to contribute; I'll come back to that later.*
 - *For the full story, I recommend looking up Cristina Perez' dissertation [C. I. Perez, *The RDA's Metadata Standards Directory: Information gathering* (Master's dissertation, University of North Carolina, Chapel Hill, 2013).]*
- *but in a nutshell, By late 2013, we had added:*
 - 14 new standards,
 - 4 new profiles/extensions, *and*
 - 13 new tools for working with metadata *as a result of the survey.*

¶ The next step was to develop a set of use cases.

At the RDA Plenary 3 (Dublin) we asked the group to come up with use cases for the directory itself, to help us figure out how to develop it further. They came up with the use cases for 5 stakeholder groups:

- Tool developers, *which gave us some very good ideas for follow-on work*
- Data custodians
- Journal editors, funders
- Librarians
- *and a fifth group that took a different approach, and looked at whether new services might be possible if the directory provided persistent identifiers and URLs for standards.*

Then, at the RDA Plenary 4 (Amsterdam) we asked the group about use cases for metadata standards in particular disciplines, to help us think about how to organize the entries in the directory. The use cases we received back were from:

- Geospatial engineering
- Materials science synchrotron data
- Humanities and Social Sciences
- Cultural heritage

These latter use cases have also been fed into the RDA Metadata Interest Group's collection of use cases.

¶ Since it was unclear how the community could help maintain the DCC directory, the last step was to set up a version of the directory to which anyone can contribute easily, and which is easy to sustain. So we've set up a directory on similar lines to the DCC one, but running on GitHub Pages (Figure 2).¹

- Each standard gets its own page, giving a brief summary of what it does and what it is used for.
- There are links to the standard's home page and specification for getting further information
- There are also links to vocabularies, ontologies or taxonomies usually used by the standards
- There is a classification into broad subject areas and specific disciplines; they are given as links so you can see other standards relevant to those areas.

¹<http://rd-alliance.github.io/metadata-directory/>

<p>Metadata</p> <p>RDA Metadata Directory</p> <hr/> <p>Edit this page</p> <hr/> <p>Getting Started</p> <ul style="list-style-type: none"> View the standards View the extensions View the tools View the use cases Browse by subject areas <hr/> <p>Adding standards</p> <ul style="list-style-type: none"> Adding extensions Adding tools Adding use cases <hr/> <p> github @twitter linkedin facebook </p>	<h2 style="margin: 0;">MIDAS-Heritage</h2> <p>A British cultural heritage standard for recording information on buildings, archaeological sites, shipwrecks, parks and gardens, battlefields, areas of interest and artefacts.</p> <p>Sponsored by the Forum on Information Standards in Heritage, MIDAS Version 1.1 was released in October 2012.</p> <p>Summary Edit</p> <p>Standard Website http://www.english-heritage.org.uk/publications/midas-heritage/</p> <p>Specification http://www.english-heritage.org.uk/content/publications/publicationsNew/guidelines-standards/midas-heritage/midas-heritage-2012-v1_1.pdf</p> <p>Related Vocabularies</p> <p>INSCRIPTION</p> <p>Subjects</p> <ul style="list-style-type: none"> Arts and Humanities Social and Behavioral Sciences <p>Disciplines</p> <ul style="list-style-type: none"> Archaeology Architecture Building Conservation Heritage Studies Historical and Philosophical Studies History by Area <p>Extensions Add</p> <p>CARARE metadata schema Edit</p> <p>An application profile of the MIDAS Heritage standard intended for delivering metadata to the CARARE service environment about an organisation's online collections, monument inventory database and digital objects.</p>
--	---

Figure 2: A sample metadata standard record from the RDA Metadata Standards Directory.

- Alongside the standard are lists of associated profiles that use it, tools for working with it, and repositories or archives that use it. These give an idea of the level of adoption and provide pointers to the researcher on how to use the standard ‘for real’.
- Pages are generated from simple easy-to-edit text files.
- *There’s an international team of Maintainers who can edit the pages directly.*
- *But more importantly, Anyone can fork the whole directory, make changes on their copy, and submit a pull request to the maintainers. If they are not happy doing that, we also have an issue tracker where people can suggest new entries, updates, or other improvements.*

The whole thing is version controlled (so any errors can easily be rolled back) and it is a simple matter to turn a forked copy into a working website should anything happen to the original. I managed to get it up and running on my laptop in about twenty minutes. But both the DCC and DataONE have committed to keeping the directory available.

¶ We’re pleased to report that the system is working. To date we’ve had ten new entries added through GitHub and several updates and corrections (see Figure 3). These have all been fed back into the original DCC directory as well. I’m afraid I don’t have usage figures for the site on GitHub, but I can tell you that this year, the DCC directory received 22 000 visits, accounting for 7% of the traffic to the DCC website. To give you a sense of what that means, it is the third most popular section of the site, after the events section with 14% of the traffic, and the How-to Guides section with 10% of the traffic (see Figure 4).

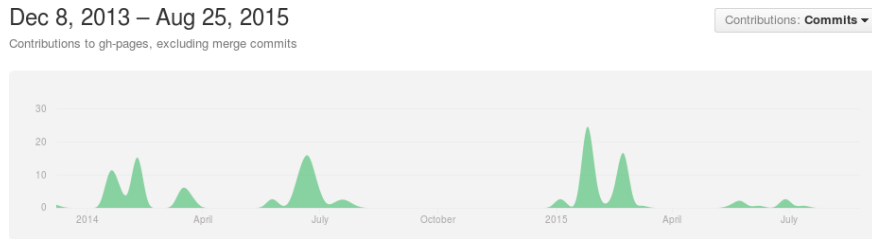


Figure 3: Usage statistics showing commits to GitHub since December 2013.

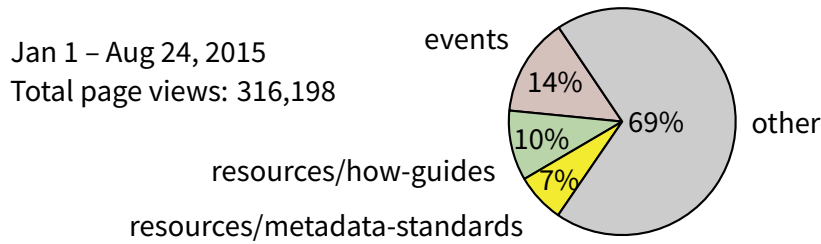


Figure 4: Usage statistics showing the proportion of page views attributable to sections of the DCC website since January 2015.

¶ The group wrapped up in March this year, but the work goes on. As well as continuing to expand and update the directory, we have ambitious plans to take it even further with a (Proposed) *Metadata Standards Catalog Working Group* (MSCWG)

- *The idea is to* Develop the Metadata Standards Directory into a Metadata Standards Catalog *by taking the entries and functionality of the catalogue and...*
- *Allowing* records to be added, searched and retrieved by API, *not just browsed by humans, though of course that will still be possible.*
- *We want to* Provide representations of records in machine-understandable form, *probably Linked Open Data but we'll see what works best.*
- *We want to* Develop, with community *participation and consensus*, recommendations of which standard(s) should be used for which purpose,
- *and* Provide information on elements defined by each standard, and the purpose they serve; *this will assist users to create application profiles or modular bundles of canonical metadata elements for specific purposes.*

For more information about the Metadata Standards Catalog Working Group, see the group page on the RDA website, where you can view the case statement and join the mailing list: <https://rd-alliance.org/groups/metadata-standards-catalog-working-group.html>

4 Acknowledgements

All that remains is for me to acknowledge the many contributors to this work:²

Fellow MSDWG co-chairs

- Jane Greenberg, (MRC)
- Keith Jeffery
- Rebecca Koskela, DataONE

DCC Disciplinary Metadata Catalogue

- Liz Bedford, DCC

Survey and GitHub work

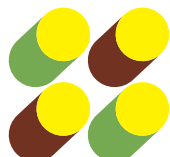
- Sean Chen, (MRC)
- Cristina Perez, (MRC)
- Kate Anne Alderete, DataONE
- Adrian Ogletree, (MRC)

I would like to thank the Working Group members who suggested directory entries, provided use cases, and helped to steer the work.

Finally, thank you for listening.



Except where otherwise stated, this work is licensed under the Creative Commons Attribution 4.0 International licence: <https://creativecommons.org/licenses/by/4.0/>



The Metadata Standards Directory Working Group is part of the Research Data Alliance, which is supported by the European Commission, the US Government and the Australian Government.

For more information, please visit <https://rd-alliance.org/working-groups/metadata-standards-directory-working-group.html>

²(MRC) = Drexel University Metadata Research Center; DCC = UK Digital Curation Centre.