

Summary of the activities of the Metadata Standards Directory Working Group

Alex Ball

DCC/UKOLN, University of Bath

RDA Plenary 5, 9–11 March 2015

Abstract

The Metadata Standards Directory Working Group concluded its activities at the Research Data Alliance Fifth Plenary. This document represents the superset of the three presentations that gave a final summary report of the group's activities:

- Joint session of the Metadata IG, Metadata Standards Directory WG, Data in Context IG, and Research Data Provenance IG (9 March, 14:00)
- Outputs Plenary (9 March, 17:30)
- Metadata Plenary (10 March, 12:15)

1 Motivation

There are many benefits to using standards.

Standard protocols → different systems can communicate

Standard file formats → different software can process the same files

Standard metadata → different systems can *process/preserve/search/use/combine* data

§ BUT this only works if everyone uses the same standards:

Not enough standards → everyone does their own thing

Too many *incompatible* standards → things get stuck in silos

So the ideal is to have one standard for a given purpose, that everyone knows about, contributes to and uses. But how do we get to that utopia?

¶ What needs to happen is that anyone coming to document their data for the first time asks themselves this:

Is there a standard that suits my purpose?

Yes Use it!

Sort of... Specialize! *Figure 1 shows how various bodies have spun their own profile of ISO 19115, for example.*

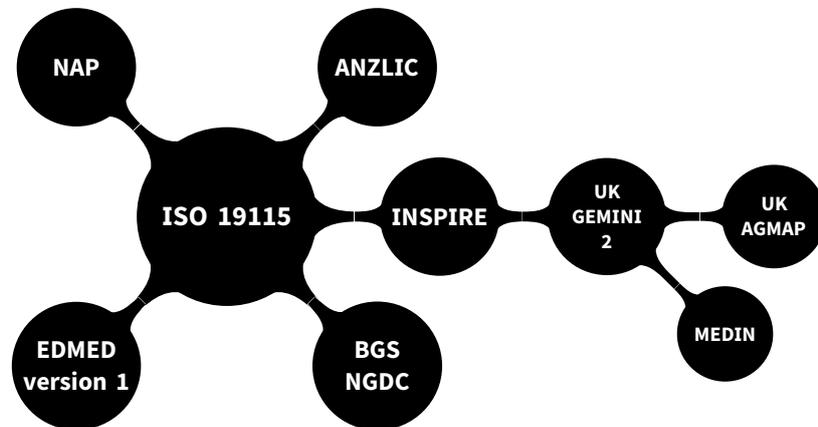


Figure 1: Some profiles of the ISO 19115 standard.

No New standard needed... *though there might be generic standards one could borrow from and specialize.*

But you can't answer that question without a good knowledge of all the metadata standards out there.

2 Purpose of the Working Group

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

3 Activity

Before forging ahead, the Working Group looked for lists and catalogues of metadata standards that others had already assembled (¶ Figure 2). What stuck out was there was already a resource that was approximately what the group had in mind: the DCC Disciplinary Metadata Catalogue.

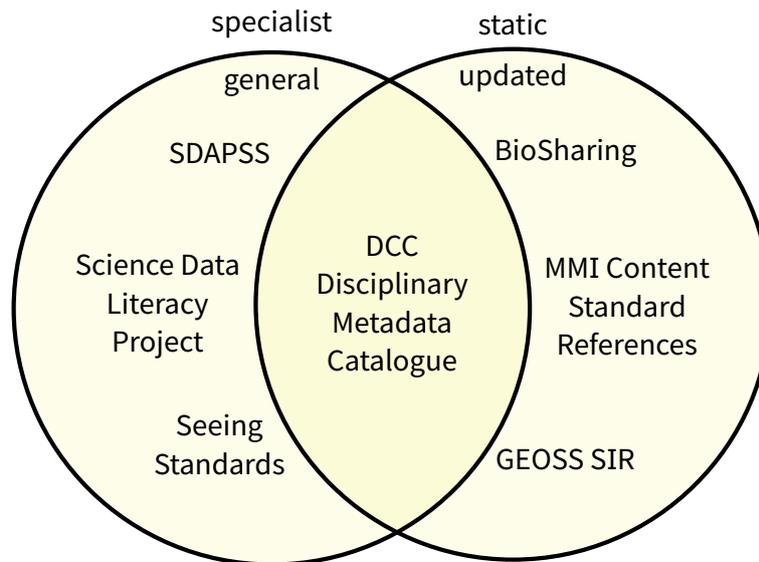


Figure 2: Venn diagram showing pre-existing lists of metadata standards.

But it wasn't a perfect fit. ¶

Why? DCC directory was

- tightly focused on metadata-for-reuse;
- aimed at UK HE;
- mostly the work of two people, so incomplete in some disciplines.

But the first step was to solve these problems by expanding the coverage of the directory.

How?

- 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.*
- By late 2013, added:
 - 14 new standards
 - 4 new profiles/extensions
 - 13 new tools for working with metadata

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

At 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:

- Tool developers *gave us some very good ideas for follow-on work*
- Data custodians
- Journal editors, funders
- PURLs for standards

- Librarians

Then, at 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:

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

These latter use cases have also been fed into the 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 3).¹

Figure 3: A sample metadata standard record from the RDA Metadata Standards Directory.

- 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

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

- 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.
- 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. We are already receiving pull requests through GitHub, and those contributions are being fed back into the DCC directory as well. We had three in February relating to the Protein Data Bank, which also allowed us to add a use case we’d known about for a while, and this week we received a correction to our DDI entry.

¶ The group wraps up at this Plenary, but the work will go 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 Catalogue Working Group* (MSCWG)

- *The idea is to Develop the Metadata Standards Directory into a Metadata Standards Catalogue by taking the entries and functionality of the catalogue and...*
- *Allow records to be added, searched and retrieved by API, not just browsed by humans, though of course that will still be possible.*
- *Provide representations of records in machine-understandable form, probably Linked Open Data but we’ll see what works best.*
- *Develop, with community participation and consensus, recommendations of which standard(s) should be used for which purpose.*
- *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.*

4 Acknowledgements

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⟩

Thank you to Working Group members who suggested directory entries, provided use cases, and helped to steer the work.

NB. ⟨MRC⟩ = Drexel University Metadata Research Center; DCC = UK Digital Curation Centre.



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>