

# Building a Disciplinary Metadata Standards Directory

Alex Ball<sup>1</sup>      Sean Chen<sup>2</sup>      Jane Greenberg<sup>3</sup>      Cristina Perez<sup>3</sup>  
Keith Jeffery<sup>4</sup>      Rebecca Koskela<sup>5</sup>

2014-02-26

## Abstract

The Research Data Alliance (RDA) Metadata Standards Directory Working Group (MSDWG) is building a directory of descriptive, discipline-specific metadata standards. The purpose of the directory is to promote the discovery, access and use of such standards, thereby improving the state of research data interoperability and reducing duplicative standards development work.

This work builds upon the UK Digital Curation Centre's Disciplinary Metadata Catalogue, a resource created with much the same aim in mind. The first stage of the MSDWG's work was to update and extend the information contained in the catalogue. In the current, second stage, a new platform is being developed in order to extend the functionality of the directory beyond that of the catalogue, and to make it easier to maintain and sustain. Future work will include making the directory more amenable to use by automated tools.

## Contents

1. Motivation.....	1
2. Purpose of the Working Group .....	2
3. DCC Disciplinary Metadata Catalogue .....	3
4. Phase 1 development .....	5
5. Phase 2 development .....	6
6. Longer term development .....	7

## 1 Motivation

There are many benefits to using standards.

<sup>1</sup>DCC/UKOLN Informatics, University of Bath, <sup>2</sup>School of Law, Duke University, <sup>3</sup>University of North Carolina, Chapel Hill, <sup>4</sup>EuroCRIS, <sup>5</sup>University of New Mexico.

- 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

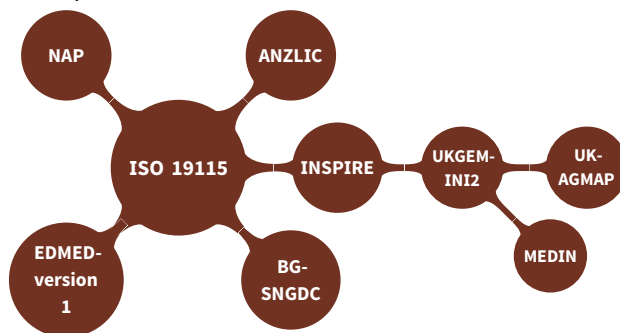
That's a perfectly general observation. What about metadata standards in particular?

¶ Metadata exists to serve a particular purpose. Some standards support discovery services such as search engines or directories. Some support preservation activities, others support packaging and transmission. Still others provide the contextual metadata needed to support a full range of administrative tasks, and clarify how a resource may be used. But perhaps the greatest variety exists among standards aimed at making data reusable in highly specific contexts, such as microarray experiments or materials testing.

¶ So someone coming to document their data for the first time needs to ask: Is there a standard that suits my purpose?

**Yes** Use it!

**Sort of...** Specialize!

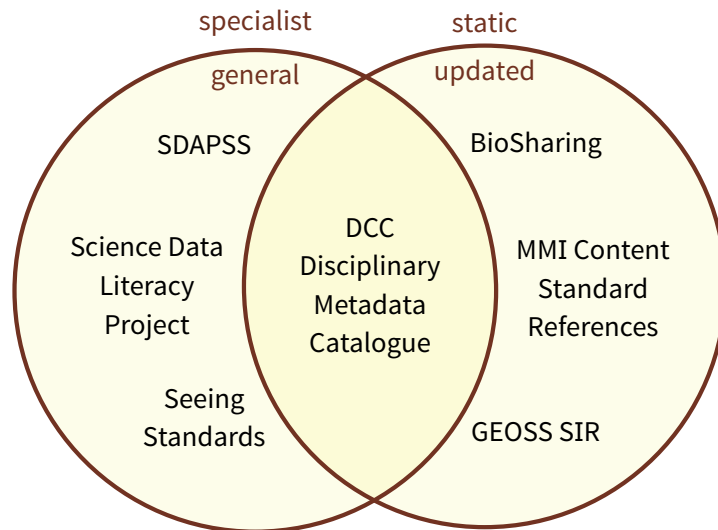


**No** New standard needed... *but even then there might be generic standards one could borrow from and specialize.*

But answering that question means having a good knowledge of all the metadata standards out there. Which brings me to the purpose of the Metadata Standards Directory Working Group.

## 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 plan for long-term growth and maintenance of the directory



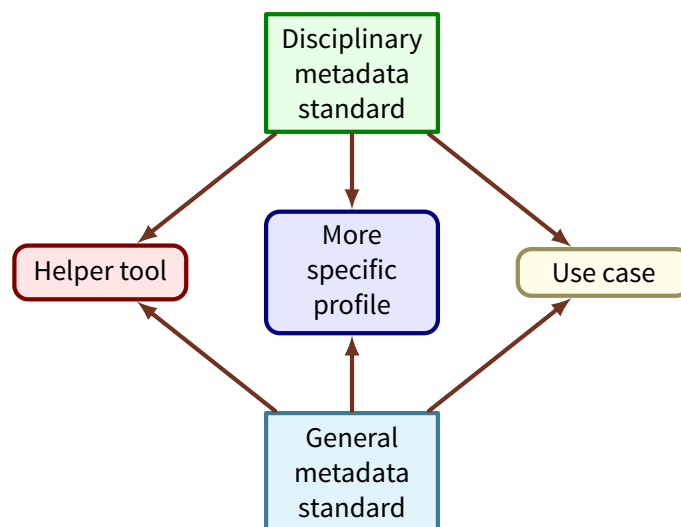
**Figure 1:** Venn diagram showing pre-existing lists of metadata standards.

Before forging ahead, the Working Group looked for lists and catalogues of metadata standards that others 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. This was fairly new: it was launched in January 2013, at about the same time as the Working Group was forming.

The catalogue was aimed primarily at institutional support staff, typically librarians or Research Office staff, who would be advising researchers on how they ought to document their data.

### 3 DCC Disciplinary Metadata Catalogue

It therefore contains the sort of information they would need to know (Figure 2).



**Figure 2:** Elements of the DCC Disciplinary Metadata Catalogue

As well as finding out what the disciplinary metadata standards were and their specifications,

- § they might also want to know about profiles that have been tailored to certain circumstances;
- § if there is no explicit standard for a discipline, the curator might want to know about broader standards that could be adapted;
- § for any given standard, it would be useful for the curator to know about tools that have been written for working with it;
- § and if there is a repository or data portal that is already working with the standard, they would be the obvious place to go to for advice on how researchers should be using it, or how the standard might be implemented locally.

SPASE Data Model

An information model for describing the elements of the heliophysics data environment, and a set of resource types which can be used to describe data along with its scientific context, source, provenance, content and location. It is designed to support a federated data system where data may reside at different locations and may be separated from the metadata which describes it. The preferred expression form is XML.

The Space Physics Archive Search and Extract (SPASE) effort is implemented by the SPASE Consortium which is composed of representatives of the international Heliophysics data community. The Current Release of the data model (2.2.2) was updated in October 2012.

Mappings	<a href="#">OAI</a>
Related Vocabularies	<a href="#">SPASE Dictionary</a>
Specification	<a href="http://www.spase-group.org/docs/schema/">http://www.spase-group.org/docs/schema/</a>
Standard's website	<a href="http://www.spase-group.org/data/">http://www.spase-group.org/data/</a>

**Extensions**

[IMPEX Data Model](#)  
A simulation extension to the [SPASE](#) data model.

**Tools**

[SPASE Metadata Editor](#)  
A web-based editor for generating [SPASE](#) descriptions.

[SPASE Tools](#)  
The [SPASE](#) website's list of tools for working with [SPASE](#) metadata and the [SPASE](#) framework.

**Use Cases**

[NSSDC SPASE Registry](#)  
The National Space Science Data Center's registry of [SPASE](#)-described space science mission data.

[SPASE Inside](#)  
The [SPASE](#) website's list of systems that use [SPASE](#) compliant metadata to enable search services.

← Description

← Key links/facts

← Links to extensions

← Links to tools

← Links to use cases

**Figure 3:** The catalogue entry for the SPASE Data Model

You can see this model reflected in the catalogue records themselves. Here (¶ Figure 3) is one of them. We have a description of the standard at the top, and then we have a table of relevant links and key facts. These can include

- mappings from that standard to other metadata standards;
- vocabularies that should or could be used with it;
- a statement about the currency of the standard;
- the organization that develops the standard;
- the specification for the standard; and
- the website or home page for the standard.

Below that we have the links to § profiles and extensions that reference the standard, § tools for working with it, and § repositories or portals that have implemented it. Unlike the top level standards, these have minimal records in the system: just enough to be able to generate these lists.

¶ So why not just use the DCC Catalogue?

- Not obvious how to contribute updates or recommend new entries
- Part of a larger site, so limited functionality and write access

- Hosted by single organization
- Scope not quite broad enough
  - Designed to support UK Higher Education
  - Excludes contextual/administrative metadata standards – *e.g. CERIF*
  - Excludes preservation metadata standards – *e.g. PREMIS*
  - Excludes structural metadata standards – *file formats and formatting conventions*
  - Excludes standards relating to non-tabular data: audio, video, interview transcripts, etc.

§ Good starting point, but more work needed.

Therefore the working group decided to collaborate with the DCC on turning their catalogue into the kind of directory that was needed. We're doing this in several phases.

## 4 Phase 1 development

In the first phase, we engaged the services of two students, Sean Chen and Cristina Perez, to run a survey soliciting suggestions for new entries (Figure 4). They used a Google Form, and once they had checked it with the working group chairs and some other interested parties, sent it out to various mailing lists.

Sent to lists:

- RDA-All
- RDA MSDWG
- RDA Metadata Interest Group
- EuroCRIS
- EPOS
- DC-SAM
- ASIS&T RDAP summit series
- DataONE
- ESIP
- STFC
- ...

**Metadata Directory Collection**

In cooperation with the UNIC SILS Metadata Research Center and the Research Data Alliance students at UNIC SILS are gathering information about metadata standards that apply to scientific data. The aim of this is achieve a short term goal of the Research Data Alliance's Metadata Standards Directory Working Group is:

Develop a prototype web-based directory (RDA Metadata Directory) listing metadata standards applicable to scientific data. The initial emphasis will be on widely-used and domain community- endorsed metadata standards and schemes with significant interpretation / re-use capability.

Information submitted to this project will be integrated with similar information from the Digital Curation Center to prototype and build a sustainable platform for sharing and exposing information about metadata standards.

If you have previously submitted information to our collection: check the Digital Curation Center's directory of standards for science data at: <http://www.dcc.ac.uk/resources/metadata-standards>

Please submit one standard per form. At the end of the questionnaire a link to another form submission is available.

\* Required

Standard Name \*

Description of the Metadata Standard  
May be a URL to about page or free text summary.

Metadata Standard URL \*  
URL for the standard's home page

Domain of the Metadata Standard \*  
Domains where the standard is in use

Biology  
 Earth Science  
 Physical Science

**Figure 4:** Extract from the survey form used to collect suggestions

¶ In the two weeks of the initial survey period (October 2013), we received 28 useful replies from specialists in Australia, Europe and North America, covering a range of disciplines (Figure 5). This led to us adding 14 standards, 4 extensions, 13 tools and 19 use cases to the catalogue, plus several updates. In fact this survey is still open and more responses are coming in. There's a link to it on this slide (and I'll repeat it at the end) if you feel moved to contribute a suggestion.



## 6 Longer term development

We have some longer-term goals which may spill over into a future working group.

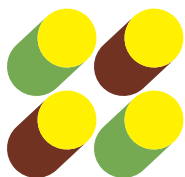
- Extend the directory to serve more use cases. *The catalogue was designed with a single use case in mind: a data curator browsing for metadata standards and resources that might be useful for a particular researcher or project. We will look at how useful the directory might be for:*
  - searching
  - querying and processing by automated tools
- ...representation of records in RDF?
- ...representation of the standards themselves in RDF? *This would enable the directory to be used by tools for building application profiles, for example.*
- Categorize metadata standards by primary application, e.g. discovery, preservation, reuse. *We're working on the use case analysis to make this possible right now.*

¶ So that's where we are. Please do come and talk to us about this work at the 3rd RDA Plenary in Dublin, Ireland next month, or drop by the website and join the group.

Alex Ball<sup>1</sup>, Sean Chen<sup>2</sup>, Jane Greenberg<sup>3</sup>, Cristina Perez<sup>3</sup>, Keith Jeffery<sup>4</sup>, Rebecca Koskela<sup>5</sup>. <sup>1</sup>DCC/UKOLN Informatics, University of Bath, <sup>2</sup>School of Law, Duke University, <sup>3</sup>University of North Carolina, Chapel Hill, <sup>4</sup>EuroCRIS, <sup>5</sup>University of New Mexico.



Except where otherwise stated, this work is licensed under the Creative Commons Attribution 4.0 International licence: <http://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>

DCC Disciplinary Metadata: <http://www.dcc.ac.uk/resources/metadata-standards>

Standards survey: <http://bit.ly/1fToaqd>