



Citation for published version:

Ball, AJ 2013, 'The DCC Disciplinary Metadata Catalogue', Paper presented at CAMP-4-DATA, 13th International Conference on Dublin Core and Metadata Applications, Lisbon, Portugal, 6/09/13.

Publication date:
2013

Document Version
Early version, also known as pre-print

[Link to publication](#)

Publisher Rights
CC BY

University of Bath

Alternative formats

If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

DCC Disciplinary Metadata

Alex Ball

2013-09-06

Abstract

In January 2013, the UK Digital Curation Centre launched a Disciplinary Metadata catalogue. This catalogue lists descriptive standards that aid in data re-use, alongside information on profiles, tools and examples of how these standards have been used by repositories.

The Disciplinary Metadata catalogue was first proposed to us (in May 2012) as a research project by Liz Bedford, who had previously worked on our Tools Catalogue.

She had noted that institutional data curators tend to concentrate on administrative and basic discovery metadata when working with research datasets. But much of the emphasis of the open data movement generally and funder requirements specifically is on data sharing. The data needs to be reusable. To make that happen, the data needs to be described and documented according to discipline-specific metadata standards. ¶ As Liz put it in her proposal,

“... disciplinary metadata standards... indicate the domain-specific information that will allow data to be *interpreted correctly* by others in the field. Since data curators *cannot become experts* in all of the subjects under research within their institutions, a particular need exists for guidance regarding disciplinary metadata standards.”

— Liz Bedford (emphasis added)

So Liz devised a resource that data curators could consult when working with researchers on how they ought to document their data for reuse. ¶ She considered what data curators would need to know and came up with this plan (see Figure 1).

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;

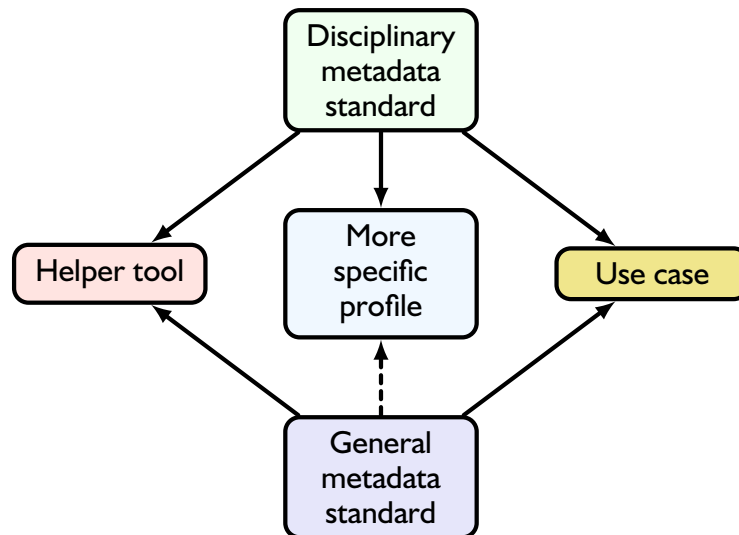


Figure 1: Elements of the DCC Disciplinary Metadata catalogue

§ 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.

Putting this catalogue together had the possibility of getting out of hand very quickly, so we made sure to scope it carefully. ¶

- Metadata standards, but not
 - languages/protocols – *structures in which information is conveyed*
 - taxonomies/vocabularies – *controlled terminology*
- Descriptive metadata, but not
 - administrative metadata – *e.g. CERIF*
 - preservation metadata – *e.g. PREMIS*
 - structural metadata – *file formats and formatting conventions*
- Active research data, but not
 - publications
 - learning objects
- Tabular data, but not
 - audio
 - video
 - narrative text – *such as interview transcripts*

The approach Liz took was straightforward and sensible. ¶

1. Literature review

- Ball, *Scientific Data Application Profile Scoping Study*¹
- Riley & Becker, *Seeing Standards: A Visualization of the Metadata Universe*²
- DCC Diffuse³
- UKOLN Application Profiles Support Project⁴

2. Repository review

- Databib⁵ – *Re3data wasn't around at this point*

3. Early thoughts about implementation

4. Choose taxonomy of disciplines (UK Higher Education Standards Agency Joint Academic Coding System version 3)

5. Assemble information about standards, profiles/extensions, tools and use cases, *and how they relate to one another*

6. Finalize implementation and enter information on website

7. Testing and release

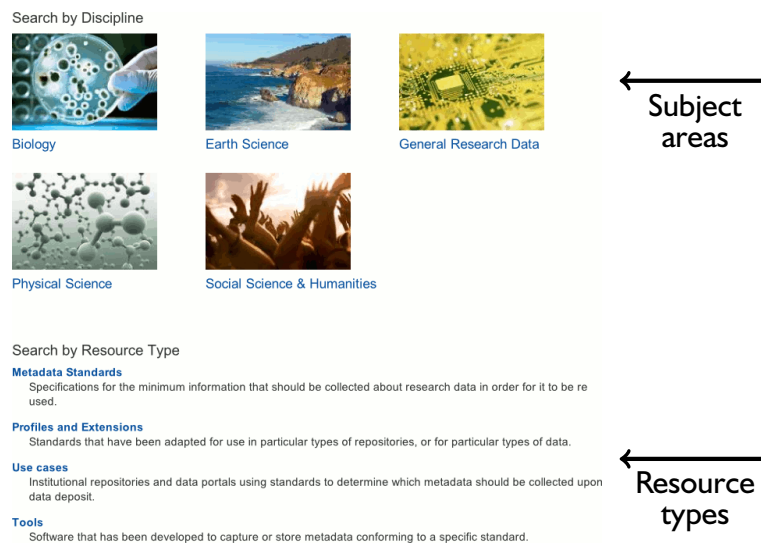


Figure 2: DCC Disciplinary Metadata catalogue

And so here is the final result (☞ see Figure 2). All the different resources have been sorted into four broad subject areas:

- Biology
- Earth Science
- Physical Science
- Social Science/Humanities

¹<http://www.ukoln.ac.uk/projects/sdapss/>

²<http://www.dlib.indiana.edu/~jenlrile/metadatamap/>

³<http://www.dcc.ac.uk/resources/standards/diffuse/>

⁴<http://www.ukoln.ac.uk/projects/ap/>

⁵<http://databib.org/>

Physical Science

Nuclear and Particle Physics Chemistry Physics Crystallography
Materials Science Solar physics Space science Astronomy Multi-disciplinary Biochemistry

Metadata Standards

AVM - Astronomy Visualization Metadata
A standard defining discovery metadata for fully rendered astronomical imagery.

CIF - Crystallographic Information Framework
An extensible standard file format and set of protocols for the exchange of crystallographic and related structured data.

CSMD-CCLRC Core Scientific Metadata Model
A study-data oriented model that captures high-level information about scientific studies and the data that they produce, primarily tailored for the physical sciences.

International Virtual Observatory Alliance Technical Specifications
A set of specifications, including metadata standards, that enable the integration of many astronomical archives into an international virtual observatory.

SPASE Data Model
An information model for describing the elements of the heliophysics data environment.

Extensions

eBank UK Metadata Application Profile
A Dublin Core Metadata Application Profile created for the eBank UK project, which provides access to the detailed results of scientific experiments in crystallography.

← Disciplines

← Relevant metadata standards

List of

- profiles/ extensions
- use cases
- tools

Figure 3: Browsing by discipline: Physical Science

There is also a fifth category, General Research Data, which is aimed at disciplines that have not yet settled on a metadata standard, or for discipline-agnostic applications.

Should you want to, you can also browse through the resources of each type in alphabetical order, without any disciplinary grouping. But most people browse the catalogue by going through one of these subject portals. If you do, ¶ you will be greeted by a page like Figure 3.

The page provides separate lists of metadata standards, profiles, use cases and tools relevant to the subject area. Each item in each list has been tagged with one or more disciplines, and you can filter the lists using the tag cloud of disciplines at the top of the page. So, for example, if you click on Astronomy, you filter out all the standards, profiles and so on that are not relevant to Astronomy. As is usual with tag clouds, the bigger the word, the more resources are associated with it.

I should point out that only the metadata standards have records in the catalogue. The links in the other three lists go straight out onto the Web to the respective home pages of the resources. But the links in the list of metadata standards do take you through to the catalogue records, and I'll show you one of those now. ¶

If we look at the record for a metadata standard (see Figure 4), you'll see 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.

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 service where data reside at different locations and may be separated from the metadata which describes it. The preferred expression form is JSON-LD.

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

Mappings	OAI
Related Vocabularies	SPASE Dictionary
Specification	http://www.spase-group.org/docs/schema/
Standard's website	http://www.spase-group.org/data/

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.

Figure 4: The catalogue entry for the SPASE Data Model

METADATA STANDARDS DIRECTORY (MASDIR) WORKING GROUP

V0.01	2013/04/29
V0.02	2013/08/01

Submitted by J. Greenberg, K. Jeffery, and R. Koskela
29 July 2013

not conducive to collective maintenance. In the UK, JISC and Digital Curation Centre recently launched the Disciplinary Metadata resource covering a variety of disciplines, and the RDA MSDIG has been evaluating this resource.³ The assessment to-date reveals that this is an important accomplishment to build upon. It is apparent that collaborative metadata registration systems have

Figure 5: Case statement for the RDA Metadata Standards Directory Working Group

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.

So, what next for our catalogue?

- Review periodically for currency
- Add entries in response to suggestions – *so if you have any, please let me know*
- Work with **Research Data Alliance Metadata Standards Directory Working Group** (see Figure 5)

Alex Ball. DCC/UKOLN Informatics, University of Bath. <http://alexball.me.uk/>



Except where otherwise stated, this work is licensed under Creative Commons Attribution 2.5 Scotland: <http://creativecommons.org/licenses/by/2.5/scotland/>



The DCC is funded by JISC.

For more information, please visit <http://www.dcc.ac.uk/>

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