The DCC Disciplinary Metadata Catalogue

Alex Ball
DCC/UKOLN Informatics,
University of Bath, UK
a.ball@ukoln.ac.uk

Abstract

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

Keywords: metadata; datasets; data reuse; application profiles; standards

1 Motivation

The Digital Curation Centre (DCC) is a collaboration between the Universities of Edinburgh, Bath and Glasgow. It provides assistance and expertise to the UK Higher Education sector on research data management and other digital curation matters. One of the main vehicles for this is the DCC website, through which the DCC disseminates its published guidance. The website also hosts directories of useful external resources.

One such directory is of digital curation tools. This directory was unsatisfactory in many ways, so the DCC engaged the services of research consultant Liz Bedford to produce a more structured and selective version, which would be more intuitive to use and easier to maintain. This she did, and the new Tools Catalogue was launched in May 2012.

On completion of the work, Liz proposed a research project to create a second catalogue, this time of disciplinary metadata standards. In her proposal, she noted that institutional data curators tend to concentrate on administrative and basic discovery metadata when working with research datasets. In order for a dataset to be useful to other researchers, however, it needs to be documented in a detailed manner according to the needs of the research discipline:

> …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.
>
> — Bedford (2012)

2 Scope

Liz devised a resource that data curators could consult when working with researchers on how they ought to document their data for reuse. She proposed that there were five main types of resource that a data curator would need to know about; these are illustrated in Figure 1.

- They would need to know about the metadata standards in common use in the discipline in question, and where to find their specifications.
- They might also want to know about profiles that have been tailored to certain sub-disciplines or local contexts.
- 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.

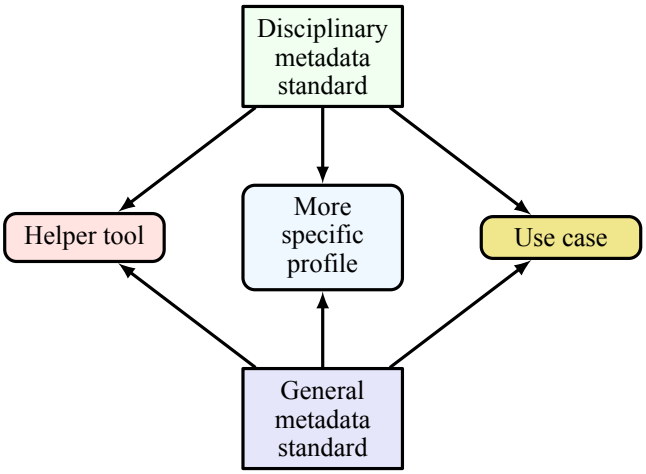

FIG. 1: Elements of the DCC Disciplinary Metadata catalogue

- If there is a repository or data portal that is already working with the standard (a 'use case'), it is likely to be a good source of advice on how researchers should be using the standard, or how it might be implemented locally.

In order to making compiling the catalogue a realistic proposition given the effort available, the project was given a tight scope.

- The catalogue would concentrate on metadata standards that identify key pieces of information about a dataset, rather than standards that deal only with how that information (or the dataset itself) should be structured, conveyed or serialized. It would also avoid going into detail about controlled vocabularies and taxonomies, though it would note if a standard used a particular external vocabulary.
- The catalogue would concentrate on standards for detailed, descriptive metadata rather than administrative, contextual, preservation or structural metadata.
- The catalogue would concentrate on metadata standards that describe active research data, rather than publications or learning objects. More specifically, it would concentrate on standards relating to tabular data rather than, say, audio files, videos, or narrative text (such as interview transcripts).

3 Methodology

The approach taken to compiling the catalogue may be summarized as follows.

1. A literature review was conducted to compile a list of candidate standards and profiles. Among the key resources reviewed were Ball (2009), Riley and Becker (2010), 'DCC DIFFUSE Standards Frameworks' (n.d.) and 'Application Profiles Support Project' (n.d.).
2. A review was conducted of the repositories listed in 'Databib' (n.d.), to identify those using particular standards and profiles.
3. A plan for how to implement the catalogue was sketched out.
4. A taxonomy of disciplines was chosen; as the DCC's primary target audience is the UK Higher Education sector, the taxonomy used by that sector – JACS (HESA & UCAS, n.d.) – was a strong contender and soon selected.
5. The list of standards, profiles, tools and use cases was classified by discipline and completed with descriptive information, inter-relationships and other links.

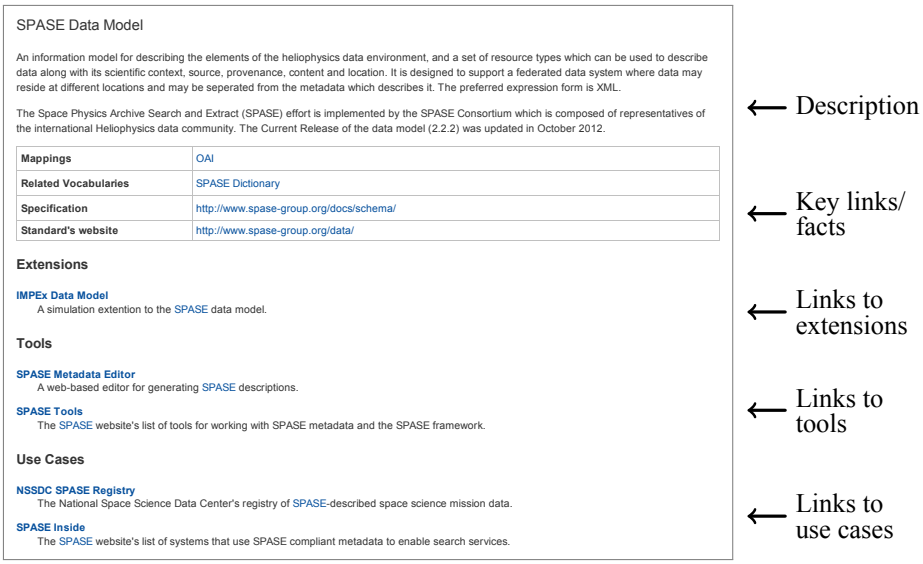

FIG. 2: The catalogue entry for the SPASE Data Model

6. The implementation was finalized, and the information entered on the website.
7. The catalogue was tested, and finally published in January 2013 ('Disciplinary Metadata', n.d.).

4 Implementation

The catalogue is implemented as a set of Drupal nodes. Three types of nodes were defined to support the catalogue:

- **Disciplinary Metadata Standard.** This is used for disciplinary or general metadata standards, and collects the most information out of the three types. These nodes each render as a page; an example is shown in Figure 2.
- **Disciplinary Metadata Standard Implementation.** This is used for profiles, tools and use cases. The information collected is limited to a description, a link to the home page for the resource, and classification information. These nodes are used to build lists but do not render as a page.
- **Disciplinary Metadata Subject Area.** This is used for grouping the resources under broad subject headings. The nodes are rendered as pages – an example is shown in Figure 3 – but do not themselves collect much information. The list of Subject Area nodes is used as a controlled vocabulary within the other two types.

A page for a metadata standard (see Figure 2) displays a description of the standard followed by a table of key links and 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 ('sponsor');
- the specification for the standard; and
- the website or home page for the standard.

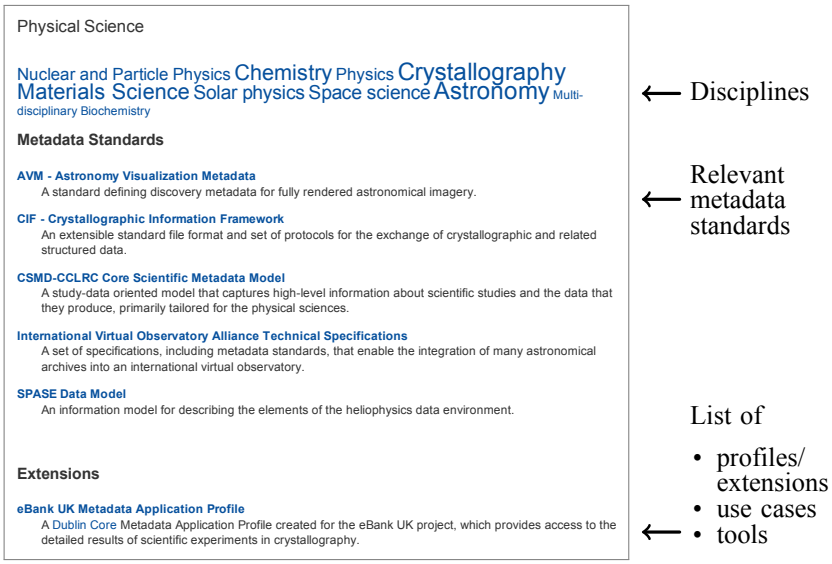

FIG. 3: The subject area page for Physical Science

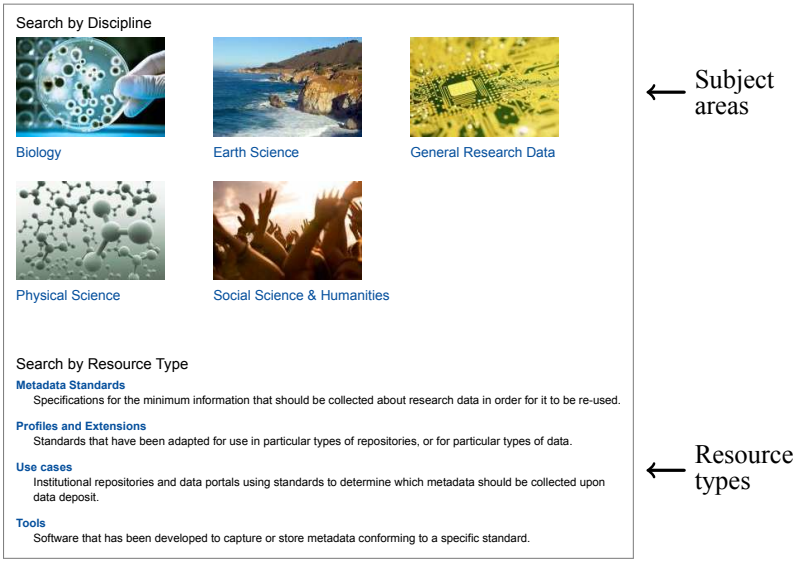

FIG. 4: DCC Disciplinary Metadata catalogue front page

Beneath that are the lists of profiles ('extensions'), tools and use cases that have been tagged as related to that standard. The links shown lead directly to the resource home pages.

A page for a subject area (see Figure 3) simply lists all the metadata standards, profiles, tools and use cases that have been tagged as belonging to that subject area. The links to standards lead to the catalogue pages on the DCC website, while (as before) the links in the other lists are external. The tag cloud at the top of the page allows the lists to be filtered by discipline. As is typical for tag clouds, a larger typeface indicates that more resources are associated with that discipline.

Figure 4 shows the front page for the catalogue. It shows the four subject areas currently in use: Biology, Earth Science, Physical Science, and Social Science & Humanities. Discipline-agnostic resources have been added to the 'General Research Data' subject area, for the benefit of those working in a discipline without its own metadata standard or in a discipline-agnostic context. The page also provides the facility, should it be needed, to browse all the resources alphabetically without any subject-based filtering.

5 Next steps

The DCC reviews the catalogue entries periodically to keep them up to date. It also extends an open invitation for anyone to suggest standards to include; indeed, at the time of writing five standards and their associated tools, profiles and use cases have been added in response to public suggestion.

Taking this a step further, the DCC is working with the Research Data Alliance (RDA) Metadata Standards Directory Working Group on broadening out responsibility for maintaining the catalogue, giving it more global utility and putting it on a more sustainable basis. As the case statement for the working group puts it:

> 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. The assessment to-date reveals that this is an important accomplishment to build upon. — Greenberg, Jeffery and Koskela (2013)

6 Acknowledgements

The Digital Curation Centre is supported by Jisc.

References

Ball, A. (2009, June 3). *Scientific data application profile scoping study*. University of Bath. Retrieved 27 September 2013, from http://www.ukoln.ac.uk/projects/sdapss/papers/ball2009sda-v11.pdf

Bedford, L. (2012, May). DCC disciplinary metadata evaluation. Project plan. Version 2. Digital Curation Centre.

Greenberg, J., Jeffery, K. & Koskela, R. (2013, August 1). *Metadata standards directory (MASDIR) working group*. Research Data Alliance. Retrieved 1 October 2013, from https://rd-alliance.org/working-groups/metadata-standards-directory-masdir-working-group.html

HESA & UCAS. (n.d.). Joint Academic Coding System (JACS) version 3.0. Retrieved 1 October 2013, from Higher Education Statistics Agency website: http://www.hesa.ac.uk/content/view/1776/649/

Application Profiles Support Project. (n.d.). Retrieved 1 October 2013, from UKOLN website: http://www.ukoln.ac.uk/projects/ap/

Databib. (n.d.). Retrieved 1 October 2013, from http://databib.org/

DCC DIFFUSE Standards Frameworks. (n.d.). Retrieved 1 October 2013, from Digital Curation Centre website: http://www.dcc.ac.uk/resources/standards/diffuse/

Disciplinary Metadata. (n.d.). Retrieved 1 October 2013, from Digital Curation Centre website: http://www.dcc.ac.uk/resources/metadata-standards

Riley, J. & Becker, D. (2010). Seeing standards: A visualization of the metadata universe. Retrieved 1 October 2013, from http://www.dlib.indiana.edu/~jenlrile/metadatamap/