Developing the DataCite Metadata Schema

July 6, 2012

Elizabeth Newbold, The British Library
Outline

• What is DataCite and why does it need a metadata schema
• How we developed the schema
• Mandatory and optional properties
• Current work and future developments
What is DataCite?

- Established in 2009 (December)
- A member of the International DOI Foundation
- A Registration Agency for DOI names
- Not for profit organisation
- 16 full members from Europe, North America and Australia
- Members work with data centres in their own countries
- Provide a shared infrastructure for minting DOIs
Why develop a metadata schema?

• Part of the infrastructure for DOI registration – metadata is required for each object (requirement of the IDF)
• Need a metadata schema for clients to use in order to register DOIs
• Support citation and discovery of datasets
• There was not a schema we could use that fitted our needs
Aims of the metadata schema

“...a core list of metadata properties chosen for accurate identification of data for citation and retrieval purposes...”

• Support citation and discovery of datasets
• Must be suitable for all types of data
• Must work with all disciplines and subjects – domain agnostic
• Do this in the simplest manner
Development of the schema

- Metadata working group established in Autumn 2009 – group meets virtually
- Representatives from across DataCite partners and different backgrounds
- First public version released for comment/consultation in August 2010
- Consultation from August – September 2010 – dedicated mailing list for feedback
- Feedback from individuals and groups
- Comments reviewed and schema revised September – December 2010
• Version 2.0 of the metadata scheme released January 2011 major changes as a result of feedback
• Metadata supervisory appointed based at TIB in Germany
• Minor changes in 2011
  • Version 2.1 released April 2011
  • Version 2.2 released July 2011 – current version
• Working group continues to meet
The DataCite Metadata Schema

Descriptive metadata

Metadata is supplied by the data clients and added to the metadata store
• 5 mandatory properties
• 12 optional properties

Additionally
• 2 administrative metadata properties which are assigned by DataCite
The Metadata Schema – mandatory properties

• **Identifier (with type attribute)**
  • The identifier is a unique string that identifies a resource – the allowed value is the DOI

• **Creator (with type and name identifier attributes)**
  • The main researchers involved in producing the dataset

• **Title (with optional type attribute)**
  • A name by which the resource is known

• **Publisher**

• **PublicationYear**
  • The year in which the data was or will be made publicly available
The mandatory metadata elements are required for registration and minting DOIs but also provide the elements for the citation.

Recommend citation style:
Creator (Publication Year): Title. Publisher. Identifier

Rees, A.P.; (2012): Impact of phosphate and iron fertilisation on rates of nitrogen fixation in surface waters by stable-isotope mass spectrometry during the FEEP programme cruise PO311. British Oceanographic Data Centre (BODC), Natural Environment Research Council (NERC). http://dx.doi.org/10.5285/6B0962C8-574B-473B-A18A-1F9AC95AA7BF
Optional properties help describe the dataset by providing the opportunity for more detailed description

- Subject (with scheme attribute)
- Contributor (with type and name identifier attributes)
- Date (with type attribute)
- Language
- **ResourceType** (with description attribute)
- AlternativeIdentifier (with type attribute)
- RelatedIdentifier (with type and relation type attributes)
- Size
- Format
- Version
- Rights
- Description (with type attribute)
Allowed values and controlled lists

Some of the properties have controlled lists of allowed values

There are + and –

+ standardised
+ helps discovery
- deciding what to use
- maintaining and adding
On-going and future work

- Schema continues to be reviewed and developed based on feedback from data centres and users
- Revisions and new versions
  - small revisions e.g. clarity on some of the definitions for optional property
  - new version, major change e.g. addition of new property
- How do we work with other metadata schemas? e.g. DDI rich metadata
- Development of a formal DataCite Dublin core application profile (DC2AP)
DataCite Dublin Core Application Profile (DC2AP)

Objectives:

- Develop a formal DataCite Dublin core application profile (DC2AP) based on Dublin Core’s Singapore Framework Application Profiles
- Document a methodology for updating the profile on an ongoing basis and identify responsibilities

Rationale:

- Formalise the DataCite metadata schema
- Confirm interoperability with Dublin Core
- Develop a model for future updates, maintenance and reuse of the DataCite metadata schema
Current status

- Development of and RDF version of the DC2AP is underway
- Draft usage guidelines have been developed
- Draft syntax guidelines have been developed
- Working towards ensuring on-going maintaince of the DC2 AP and its alignment with the DataCite metadata schema
- Target release date to coincide with the next release of the DataCite schema
Further information

- DataCite metadata schema [http://schema.datacite.org](http://schema.datacite.org)
- IsCitedBy: A metadata scheme for DataCite [http://www.dlib.org/dlib/january11/starr/01starr.html](http://www.dlib.org/dlib/january11/starr/01starr.html)
- Dublin Core [http://dublincore.org/specifications](http://dublincore.org/specifications)
- Scholarly Works Application Profile [http://www.ukoln.c.uk/repositories/digirep/index/Eprints_Application_Profile](http://www.ukoln.c.uk/repositories/digirep/index/Eprints_Application_Profile)