Metadata Schema 4.4

ANNE RAUGH, UNIVERSITY OF MARYLAND

Metadata Working Group

Madeleine de Smale, TU Delft (Co-chair)

Isabel Bernal, CSIC (Co-chair)

Jan Ashton, British Library

Marleen Burger, TIB

Robin Dasler, DataCite Product Manager

Martin Fenner, DataCite Technical Director

Samantha Folger, ETH Zurich

Ted Habermann, Metadata Game Changers

Violets Ilik, Columbia University

Anne Raugh, University of Maryland

Andreas la Roi, ETH Zurich

Liza Zolly, USGS

Schema Additions: *resourceTypeGeneral*

A number of new *resourceTypeGeneral* values will be added to the controlled list:

Journal
JournalArticle
Book
BookChapter
ConferenceProceeding
ConferencePaper
Standard
Report
Dissertation
Preprint
DataManagementPlan
PeerReview
ComputationalNotebook

This subproperty is required to be present and represents a more general type than the value of the *resourceType* property, which is free text.

*There will be a change in guidance in how to use the **resourceType** property for text.*

Examples: *resourceTypeGeneral*

resourceTypeGeneral:	*BookChapter*
resourceType:	Invited Review
resourceTypeGeneral:	*ConferencePaper*
resourceType:	Abstract of Poster
resourceTypeGeneral:	*ComputationalNotebook*
resourceType:	Jupyter Notebook
resourceTypeGeneral:	*Report*
resourceType:	Data Maturity Assessment

Examples: *resourceTypeGeneral*

```
<resourceType resourceTypeGeneral="BookChapter">Invited review</resourceType>

<resourceType resourceTypeGeneral="ConferencePaper">Abstract of Poster
   </resourceType>

<resourceType resourceTypeGeneral="ComputationalNotebook">Jupyter Notebook
   </resourceType>

<resourceType resourceTypeGeneral="Report">Data Maturity Assessment
   </resourceType>
```

Schema Additions: *subject*

A new subproperty, *classification*, will be added to *subject.* The *classification* subproperty will hold the code used for the *subject* term within a scheme that does not use URI codes for each term or concept.

A *subject* property, then, should have either a *valueURI* subproperty or a *classification* subproperty, but not both.

Example: *subject* with *valueURI*

The Unified Astronomy Thesaurus assigns URIs to each term. To include the subject term “Short period comets” (concept number 1452):

subject:	Short period comets
subjectScheme:	Unified Astronomy Thesaurus
schemeURI:	http://astrothesaurus.org
valueURI:	http://astrothesaurus.org/uat/1452

The same example in XML would look like this:

```
<subject subjectScheme=“Unified Astronomy Thesaurus”
         schemeURI=“http://astrothesaurus.org”
         valueURI=“http://astrothesaurus.org/uat/1452”>
   Short period comets</subject>
```

Example: *subject* with *classification*

The *Journal of Economic Literature* (JEL) system assigns codes but not URIs to its terms. To include the subject term “Consumer Protection” (concept D18):

subject:	Consumer Protection
subjectScheme:	Journal of Economic Literature
schemeURI:	https://www.aeaweb.org/econlit/jelCodes.php
classification:	D18

The same example in XML would look like this:

```
<subject subjectScheme=“Journal of Economic Literature”
         schemeURI=“https://www.aeaweb.org/econlit/jelCodes.php”
         classification=“D18”>
   Consumer Protection</subject>
```

Still in Discussion: Series Information

Specifically, series information for the resource itself, to describe the context in which it was published. Examples:

- The resource is an article. It is published in a journal which has its own identifier (an ISSN, for example). The *series information* in this case includes the name and identifier of the journal, volume number, and the pages containing the article.
- The resource is a chapter in a book. The book has its own identifier. In addition, the book is one of a series of book which also has a series identifier. The *series information* includes the name of the book, the chapter number of the resource within the book, the name of the book series, and a volume number of the book within that series.

We expect to finalize details on a design in November. There will be a new major property added to the end of the existing schema to hold this information.

Mock-up: Series Information

These are the subproperties we are planning to include to describe the containing publication:

Title
Type
Digital Identifier
Volume
Issue
Chapter Number
Article Number
Page Range (comprising the resource)

Still in Discussion: Full References

Users have requested that the full reference information (author, title, date, etc.) be given in a structured format (for example, as XML elements), in addition to the digital identifiers for *relatedIdentifiers.*

- Incorporating this directly into the existing *relatedIdentifier* property would not be backward-compatible.
- Additional reference information cannot be required, also in consideration of backward-compatibility.

This information coincides with much of what is desired for *series information*, so a general solution has been designed that we believe can accommodate both uses.

We also need to make sure that full information can be supplied for references to data sets and software.

Still in Discussion: Non-digital References

Users have requested a way to cite a work that has no digital identifier. Two alternate methods have been designed:

- Allow full references as free-form text in the value of *relatedIdentifier*.
- Allow the new *series information* property to be used without a digital identifier.

When the *series information* design is completed, the appropriate option above will be selected.

Target Release Date

January 2021

Comments Welcome

DataCite Metadata Discussion Group:

groups.google.com/g/datacite-metadata

DataCite Roadmap:

datacite.org/roadmap.html

See you in the break-out room...