

R7.3 - Updating DOI Metadata

Introduction

The DataCite Metadata Schema provides a list of core metadata elements that are used for resource citation and retrieval purposes, along with recommended use instructions. The resource that is being identified can be a dataset or any other type of research data. In addition to the five required metadata elements, the DataCite metadata group recommends that we also include Description and ResourceType when we create a DOI. These updates will improve searching of the Metadata Store (see <http://search.datacite.org/ui>). Beyond these improvements, Thomsen Reuters (TR) is cooperating with DataCite and will index all data that is in the DataCite Metadata Store that has both the Description and ResourceType elements. For TR to index all data in RUcore, we only have to add the Description to the DOI metadata and extend the ResourceType to include all of the controlled vocabulary items in the DataCite metadata v3.0 (p. 14). Note that TR uses a very broad definition of “research data” and so almost all of RUcore content will be indexed including photographs and maps and will be accessible in TR’s commercial product – Web of Knowledge. Indexing by TR provides another significant access point for RUcore resources.

In a recent CISC meeting, it was agreed that we would not block any of our resources from being indexed or harvested in the DataCite Metadata store. Although some of our material might not be considered “research data”, the criteria for making this decision is not obvious and it was the judgment of CISC that we not spend time on trying to make these decisions. To implement the DOI metadata update, we need to proceed as outlined in the following sections.

Mapping of Description and ResourceType

The mapping for Description and ResourceType is shown in the following table (from MDWG). At this point, we do not have any mappings for Model, Workflow, or Other. It is possible that our work with research data may require the Model type at some future point.

Mapping for ResourceType and Description

DataCite	WMS (DCMI)
Audiovisual	MovingImage
Collection	Collection
Dataset	Dataset
Event	Event
Image	StillImage
InteractiveResource	InteractiveResource
Model	N/A
PhysicalObject	PhysicalObject
Service	Service
Software	Software
Sound	Sound
Text	Text
Workflow	N/A
Other	N/A
descriptionType=abstract	Abstract (use any abstract, regardless of the type attribute).

Implementation Going Forward

Given the modularity of RUCore, the change to update the mappings as described above need only be made in the XSLT for DOI metadata. Since the DOI API uses the XSLT, other applications that create or modify DOIs do not need to be changed. It is proposed the XSLT update be completed in R7.3. Going forward from R7.3, all DOIs will have the correct metadata. We will need to update the existing objects that have outdated metadata and also assign DOIs retroactively to all objects that currently have HDLs.

Update of Objects that already have DOIs

For the objects in RUCore presently, the great majority have HDLs and these will be updated as discussed below. However, since the introduction of DOIs in R7.2, there are objects that use the earlier

metadata and will need to be updated. It is recommended that we first update the metadata in the current DOIs before proceeding with the retroactive assignment. (A mini-spec and script will need to be developed and reviewed).

Retroactive DOI Assignment

A draft mini-specification for retroactive assignment of DOIs has been reviewed in software architecture. The “rundoι.php” script (part of dlr/EDIT) can be run on the command line for batch operations and implements all of the methods in the DOI API. To perform the retroactive assignment, a list of Fedora PIDs for objects on production that do not have a DOI will need to be generated. The shell script, as outlined in the specification, can then be run on mss3. [Note, we might want to break up the process into smaller batches. Also, from the earlier review, a “dryrun” option will need to be added.] This script will replace the HDL in the metadata with the DOI. Note that the HDL (embedded in a URL) will continue to resolve properly.

rcj/MDWG – 01/10/2014