

## RUBRIC Toolkit: Metadata Overview

Metadata is used:

- to describe and provide the location of a resource to assist with discovery
- to give information about its conditions of use
- to govern technical and display issues
- to contain long term preservation information
- to maintain version information
- to manage administrative aspects of the data

Much metadata is machine-generated by the repository, but it is important for repository managers to understand the extent to which metadata in their repositories can be customized for their own purposes. Refer to the section on Metadata and [Entering Metadata: Guides and Tools for Repositories](#) for more detailed information on the topic.

[Resource discovery](#)<sup>1</sup> and [IR interoperability](#)<sup>2</sup> are inseparable issues and both are tied to the effective use of metadata.

## Choosing the Right Metadata Standard

[Dublin Core](#)<sup>3</sup> is the simplest of metadata schema which ensures interoperability for search and retrieval among repositories and their harvesters. Whilst it is not the only available standard, it is considered to be the baseline metadata standard for the Open Archives Initiative Protocol for Metadata Harvesting. The [Open Archive Initiative \(OAI\)](#)<sup>4</sup> develops the standard to promote interoperability among repositories, and OAI repositories will normally be configured to generate Dublin Core records by default.

While the simple Dublin Core schema provides a basic level of interoperability, the Open Archive Initiative additionally encourages the use of more granular schema. A repository should include other metadata schema to describe their resources and conditions of access more fully for the benefit of their users. Examples of other metadata schema include MARC and an extended version of Dublin Core known as Qualified Dublin Core. Other metadata schema are discussed in the Metadata section.

[Metadata Schemes Points of Comparison](#)<sup>5</sup> is one of many useful articles that describes what to look for when comparing the range of standard schema available for specialised purposes.

[Choosing a Metadata Standard For Research Discovery](#)<sup>6</sup> by UKOLN includes a checklist for selecting the right standard for your purpose.

The basic criteria to use are:

- interoperability
- extensibility and growth

RUBRIC is supported by the Systemic Infrastructure Initiative as part of the Commonwealth Government's Backing Australia's Ability - An Innovative Action Plan for the Future (<http://backingaus.innovation.gov.au>)

- sustainability
- granularity
- ease of use and existing skills

Images and videos present additional preservation, sustainability and rights issues to other resource types. There are specialist metadata schema for such non-text materials and these are also discussed briefly in the comprehensive Metadata section.

## Metadata for Harvesting

The [ARROW Discovery Service](#)<sup>7</sup> has produced a [Harvesting Guide](#)<sup>8</sup> (currently being updated) that recommends different levels of metadata content for harvesting. The ARROW Discovery Service is an OAI compliant national harvester managed by the National Library of Australia.

It liaises with a range of international service providers to manage the harvesting conditions of Australian material, including:

- [Google](#)<sup>9</sup>
- [Google Scholar](#)<sup>10</sup>
- [Yahoo!](#)<sup>11</sup>
- [OAster](#)<sup>12</sup>
- [SCIRUS](#)<sup>13</sup> (Elsevier)
- [EdNA Online \(Australian\)](#)<sup>14</sup>
- [Libraries Australia](#)<sup>15</sup>

The ARROW Discovery Service's [Public Funding, Public Knowledge, Public Access](#)<sup>16</sup> explains the agreements with Google and the OAster service to secure higher rankings for its university repository harvested resources. OAster's agreement with Yahoo! and Google is summarized at <http://www.oaister.org/sru.html><sup>17</sup>.

IR managers may liaise independently with the same providers or they may choose to register with the ARROW Discovery Service which can negotiate on their behalf.

The following terms are the Dublin Core elements that are used by OAI service providers for harvesting metadata records. (each element is repeatable and optional). The terms in bold type are the most essential:

- **title**
- **subject** (*includes keywords and controlled vocabularies*)
- **description** (*includes abstract or other summary*)
- **type** (*e.g journal article, conference paper, thesis*)

RUBRIC is supported by the Systemic Infrastructure Initiative as part of the Commonwealth Government's Backing Australia's Ability - An Innovative Action Plan for the Future (<http://backingaus.innovation.gov.au>)

- source
- relation
- coverage
- **creator**
- publisher
- contributor
- rights
- **date**
- **identifier**
- language
- format

The comprehensive Metadata section explains that all of the above elements will apply to the document or article deposited in the repository with one exception. At least one identifier that is also a resolvable link (URI) will need to point to the repository's metadata page that describes and links to that resource. Otherwise a harvesting service will direct users directly to that resource and bypass the repository.

[The National Science Digital Library \(NSDL\)](#)<sup>18</sup> advises that best practice to ensure your IR is harvested is to register with [the official OAI Registry](#)<sup>19</sup>.

Connecting with the Harvesters explains who to contact and how to register.

## Australian Digital Theses

[Australian Digital Theses \(ADT\)](#)<sup>20</sup> metadata and harvesting requirements vary slightly, even if this material is co-located with other repository material. ADT will still expect to be able to harvest those theses separately from the rest of the repository archive.

The ADT harvester only processes a limited range of Dublin Core elements:

- dc.title
- dc.creator
- dc.subject
- dc.description
- dc.date
- dc.language
- dc.publisher
- dc.rights

dc.identifier

dc.type

If there are any other DC elements in an ADT record (e.g. dc.format), they will be ignored by the ADT harvester for normal processing and indexing purposes.

Open Archive searching requires that ADT records in a repository be grouped and harvested as a discrete set of items separately from the other records in the repository. [Sets Guidelines for Repository Implementers](#) on the [Open Archive Initiative](#) website provides technical details for constructing Sets.

It is recommended that the dc.type or dc.relation element be used for the SetName for ADT harvesting. Enter "Australasian Digital Thesis" as the value for dc.type or "Australasian Digital Thesis Program" as the value for dc.relation.

**Example:**

dc.type Australasian Digital Thesis

dc.relation Australasian Digital Thesis Program

For ADT to harvest a repository according to OAI-PMH (Open Archive Initiative Protocol for Metadata Harvesting) standards, repository managers will need to inform ADT of:

- the URL of your server
- the SetSpec of the ADT records to be harvested
- the SetName for the ADT records to be harvested (i.e. dc.type Australasian Digital Thesis)

**Examples:**

DSpace record:

- URL: <http://researchspace.auckland.ac.nz/dspace-oai/request>
- SetSpec: hdl\_2292\_2
- SetName: PhD Theses

VITAL record without Collections:

- URL: <http://repository.usq.edu.au/oaiprovider>
- SetSpec: Australasian Digital Thesis
- SetName: Australasian Digital Thesis

VITAL record with ADT items in a Collection:

- URL: <http://repository.usq.edu.au/oaiprovider>
- SetSpec: rubric:299
- SetName: Australasian Digital Thesis

## Further Guides to Metadata

Guides that have most relevance for repositories in higher education institutions:

- [Best Practices for Sharable Metadata](#)<sup>21</sup>
- [IFLA's Digital Libraries: Metadata Resources](#)<sup>22</sup>
- [UKOLN's Metadata](#)<sup>23</sup>
- DC Metadata Thesaurus for Repositories
- [IMS Best Practice Guide](#)<sup>24</sup>

Criteria for evaluating a metadata schema for a digital repository can be found at:

- [Choosing a Metadata Standard for Research Discovery](#)<sup>25</sup> (UKOLN 2006) this discusses an up to date checklist of guidelines to assist one in choosing a metadata schema for a repository.
- [Metadata Schemes Points of Comparison](#)<sup>26</sup>
- [Repository Librarian and the Next Crusade: The Search for a Common Standard for Digital Repository Metadata](#)<sup>27</sup> (Goldsmith B B & Knudson, F 2006) - a comparative evaluation of MARCXML, Dublin Core, ONIX, PRISM and MODS.

---

“RUBRIC Toolkit: Metadata Overview” produced July 2007



**Copyright**<sup>29</sup> 2007 **RUBRIC**<sup>30</sup>

- 1 <http://webdoc.sub.gwdg.de/edoc/aw/d-lib/dlib/december04/vandesompel/12vandesompel.html>
- 2 <http://www.emeraldinsight.com/Insight/viewPDF.jsp?Filename=html/Output/Published/EmeraldFullTextArticle/Pdf/0150250401.pdf>
- 3 <http://dublincore.org/documents/dces/>
- 4 <http://www.openarchives.org/pmh/>
- 5 <http://eprints.rclis.org/archive/00005544/01/comparingschemes.pdf>
- 6 <http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-63/html/>
- 7 <http://search.arrow.edu.au/>
- 8 <http://www.arrow.edu.au/docs/files/harvesting.pdf>
- 9 <http://www.google.com.au/>
- 10 <http://scholar.google.com/schhp?ie=UTF-8&oe=UTF-8&hl=en&tab=ws&q=>
- 11 <http://au.yahoo.com/>
- 12 <http://www.oaister.org/>
- 13 <http://www.scirus.com/>
- 14 <http://www.edna.edu.au/edna/go>
- 15 <http://www.nla.gov.au/librariesaustralia/>
- 16 <http://www.nla.gov.au/pub/gateways/archive/76/Campbell-ArrowProject.html>
- 17 <http://www.oaister.org/sru.html>
- 18 <http://oai-best.comm.nsd.l.org/cgi-bin/wiki.pl?RepositoryLifecycle>
- 19 <http://www.openarchives.org/Register/BrowseSites>
- 20 <http://adt.caul.edu.au/>
- 21 <http://oai-best.comm.nsd.l.org/cgi-bin/wiki.pl?PublicTOC>
- 22 <http://www.ifla.org/II/metadata.htm>
- 23 <http://www.ukoln.ac.uk/metadata/>
- 24 [http://www.imslobal.org/metadata/mdv1p3pd/imsmd\\_bestv1p3pd.html](http://www.imslobal.org/metadata/mdv1p3pd/imsmd_bestv1p3pd.html)
- 25 <http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-63/html/>
- 26 <http://eprints.rclis.org/archive/00005544/01/comparingschemes.pdf>
- 27 <http://www.dlib.org/dlib/september06/goldsmith/09goldsmith.html>
- 28 <http://creativecommons.org/licenses/by-sa/2.5/au/>
- 29 <http://creativecommons.org/licenses/by-sa/2.5/au/>
- 30 <http://www.rubric.edu.au/>