RUBRIC
Regional Universities Building Research Infrastructure Collaboratively

Entering Metadata: Guides and Tools for Repositories

Table of Contents

Introduction to Metadata Training

Institutional Repository editors and library cataloguers will find the following metadata crosswalk and data entry guidelines useful as an introduction to applying metadata and to understanding key principles underlying the way metadata is treated in repositories.

Key differences: Library vs Repository

The different ways libraries and repositories collect and expose work affect the way metadata is managed. For example:

- Libraries collect works that are compilations of articles by different authors and treat the collected work as the main item. Repositories archive each authored article as the primary item and the host publication is treated as a related item.
- A library collection draws information in to the institution for the use of the institution. A repository collection is established to expose its resources openly to the entire world and to promote the work of an institution.

AACR2 is the common data entry standard for libraries. Its limitations in the face of rapid technological developments have led to the development of the Resource Description and Access (RDA).

It is even more necessary for repository metadata management to radically review the rationales for past standards and be prepared to revise or replace them if the repositories present a whole new set of rationales.

Crosswalks between MARC, MODS and Dublin Core are included in the following guidelines. These crosswalks are intended to:

- assist technical staff with spreadsheets mapping data to Dublin Core
- assist those working with metadata entry and quality control to understand the underlying principles
- demonstrate how DC metadata is the key to making repositories compliant with Open Archive standards and internationally accessible

Online Metadata Tools

Crosswalks

The Library of Congress MARC Standards website[1] provides access to complete crosswalk or mapping guides across MARC, MODS and DublinCore, including:

- MARCXML Conversion to MODS and Dublin Core Stylesheets[2]
- MODS Conversion to MARC and Dublin Core Stylesheets[3]

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)

DCMI resources

The Dublin Core Metadata Initiative (DCMI) [4]has a tools and software page of resources for:

- creating metadata templates
- changing metadata templates
- automatic extraction and production of metadata
- conversion between metadata formats

Schema Conversion

The MarcEdit tool[5] automatically converts schema.

Repository Ingest Tools: Survey and ARROW practice

CAIRO[6] (Complex Archive Ingest for Repository Objects), funded by JISC[7]'s Repositories and Preservation program, has conducted a study comparing a wide range of metadata and other extraction tools used in repositories with related common open source licenses:

CAIRO tools survey: a survey of tools applicable to the preparation of digital archives for ingest into a preservation repository[8] (21 May 2007)

Within the Australian ARROW[9] community (using the VITAL repository) there are members who have customized the VALET ingest tool for their repository. These can be contacted through the community.

Data Entry Guidelines

The following guidelines provide main headings for repository metadata requirements, including Dublin Core. These guidelines are generic (most are Dublin Core terms) and can be easily related to the requirements of specific repositories.

Headings which do not state “DC term” are not Dublin Core elements but are still integral parts of repository metadata.

In some repositories, the creation of Dublin Core fields will be a default part of the repository software. The following guidelines are sufficiently comprehensive on data entry principles that apply across all repositories as well as mapping metadata to Dublin Core where that is required as part of the initial repository configuration.

Beneath each main heading there is:

- a scope note explaining the definition and limitations of the term
- a note on related terms closely associated with the main term.
- notes are entered under MARC, MODS and data entry best practice and standards headings. These are critical parts of the guidelines, explaining:

- expected practice in repositories
- exceptions
- differences from practices normally found in a traditional library.

All Dublin Core elements are repeatable, with one exception: the ARROW Discovery Service harvester recommends that there be only one resource type in the Simple Dublin Core field.

It is not necessary to display all fields in the portal display of a repository record. For example, a list of RFCD codes may be hidden from the main record display but exposed for a browse list to be indexed for searching. The extent and ease of configurability in repositories will vary.

When deciding how much metadata to enter, consider the following points:

- records should not be cluttered with unused information
- not all metadata needs to be displayed: some may be useful for searching, for authentication of the integrity of records and archived resources or for audit purposes
- it is better to add a little more rather than a little less metadata because granularity of data entry is a strength and potentially facilitates its use and value into the future

Abbreviations used:

DC term	Dublin Core term
SN	Scope note
RT	Related term

The Detailed Description of MODS elements[10] page provides more detailed explanations for the use of the MODS elements.

title (DC term)

SN	A name given to the deposited resource, not the parent publication of the resource. Thus the title of a deposited book chapter or journal article will be mapped to this DC element, and not the title of the book or journal in which they appear.
RT	Relation

Crosswalks to Dublin Core

MARC	MARC: 245 $a $b $p $n
	MARC: 246 $a $b $p $n
Notes	Other MARC title fields do not apply to the title of the deposited resource so only the above fields should be mapped to the dc.title
MODS	MODS: <titleInfo><title>
Notes	MODS allows <titleInfo> subelements to be parsed: <nonSort>, <title>, <subTitle>, <partNumber>, <partName>

MODS subelements should be concatenated in Dublin Core, separated by a space or other form of punctuation. Enter the nonfiling text (“The”, “A”, “An”) in the <nonSort> element.

Data entry: best practice and standards

Notes

The main title should be the title of the resource at the time it is published. Other variant titles (e.g. a preprint title) can be added as alternative titles or as notes. If the title is not published enter the title as it appears on the resource.

Enter titles in full, including initial articles.

Use normal punctuation as would be used in an academic citation.

If there is no punctuation separating the main title from the subtitle (e.g. the two are separated only by a line space), use a colon to separate the two parts.

If there is no title provided with the resource, supply one.

Rationales

A resource can be known by multiple titles. Different versions (e.g. preprints and postprints) can contain title variations. Running titles, acronyms within titles, advertised titles etc can vary from the title on the resource.

The resource is most likely to be known and recognized by the publication title.

In some repositories the title will be used in a citation in the form in which it is entered in the record.

Title is an essential element required by harvesters.

creator (DC term)

SN

The person or persons responsible for the intellectual content of the deposited resource, not its presentation.

RT

contributor
affiliation (not a dc element)
role (not a dc element)

Crosswalks to Dublin Core

MARC

MARC: 100 $a $q
MARC: 700 $a $q

Notes

In a repository, archiving the scholarly output of an institution all creators will be personal names, so do not use MARC 110, 111, 710 or 711 tags. Each deposited resource will represent the work of a personal author or authors, even if it is published as part of a compilation under a corporate authorship.

Repository records will have entries for editors of conference publications,

supervisors of theses, and names of submitters of resources to the repository who are not authors of those resources. These names should be stored in a MARC 720 tag, with $e to indicate their relationship to the resource (submitter, editor, supervisor, etc.)

Some names in the 100 and 700 MARC tags that are mapped to the dc.creator will also use a $e relator subfield to indicate their role (e.g. submitter). This is for for internal administrative or authentication purposes and should not be mapped to a Simple Dublin Core element.

MODS MODS: <name type="personal"><namePart>

Notes MODS allows <name> subelements to be parsed: <namePart>, <displayForm>, <affiliation>, <role>, <description> MODS subelements should be concatenated in Dublin Core, separated by a space or other form of punctuation.

Data entry: best practice and standards

Data entry Enter the names of multiple authors of a resource in the same order in which they appear on the resource, even if this results in the name of an author not belonging to the repository's institution being entered first.

Maintain an authority file of personal names entered in the repository and always enter the same author with the same name format.

The nature of this authority file will depend on the staff and time resources available in the institution. Use of normal library authority standards such as LC name authorities, National Library authorities, AACR2 standards for forms of foreign names, etc. is discouraged in repository authority list creation. An authority list may be compiled from the forms of names appearing in the institution's formal staff directory or even from the form of the name when it is first encountered by a repository editor.

MODS also allows for a name variation to be nested with the standard form of a name with its <displayForm> element.

Rationales The order in which names appear on some multi-authored articles can have significance.

Some authors will have different forms of their name appearing across different publications, and a repository author index should contain one entry for each name.

There are no "see" or "see also" functions in most repositories at present and standardized name authorities such as those of the Library of Congress can sometimes be obscure without this functionality.

Even though the repository record will have a standardized form of an author's name, the form of name as it appears on the resource will still be displayed for users on the resource. If the difference between the formats of name is significant an explanatory note can be added to the record.

Affiliation

SN	An institution to which the author is associated. Typically this will be the university of the submitting author. This is not a Dublin Core element but is an important identification of the author in repositories.
RT	creator contributor

Do not crosswalk to Dublin Core

MARC	MARC: 100 $u MARC: 700 $u
Notes	$u in MARC can be an affiliation or address (e.g. email address) of a name. In repositories, however, always use the institution to which the author belongs. $u is not repeatable in the same MARC field. (Some repositories map affiliation to become part of the name in dc.creator. There are trade-offs to be made when and wherever it is mapped in DC. Keeping Dublin Core Simple[11] explains the problem of mapping this to DC)
MODS	MODS: <name> <affiliation>
Notes	MODS allows <name> subelements to be parsed: <namePart>, <displayForm>, <affiliation>, <role>, <description> MODS subelements should be concatenated in Dublin Core, separated by a space or other form of punctuation.

Data entry: best practice and standards

Note	Institutions may opt to enter values only for authors from their own institution. Be aware, however, that an author may have belonged to another institution at the time the resource was created and published. “Affiliation” may be used for in-house and display purposes only. It should not be mapped to Simple Dublin Core for harvesting.
Rationale	There is no scope in Simple Dublin Core for the “affiliation” of the author. It is important for institutions to clarify which authors belong to their own institutions.

Role

SN	A term that describes the relationship between the name and the resource. This is not a Dublin Core element but DC does use Relator terms[12] in Qualified Dublin Core. These do not apply to Simple Dublin Core.

RT creator
contributor

Do not crosswalk to Dublin Core

MARC MARC: 100 $e
MARC: 700 $e

Notes $e in MARC can be an affiliation or address (e.g. email address) of a name. In repositories, however, always use the institution to which the author belongs.

$e is repeatable in the same MARC field.

MODS MODS: <name> <role>

Notes MODS puts all names in a repeated<name> with type of contribution indicated in <role>. It does not make the explicit distinction between creator and contributor in terms of primary vs. secondary roles. An application may wish to designate use of Creator or Contributor for all MODS names or use the role value to determine which DC element is used.

MODS allows <name> subelements to be parsed: <namePart>, <displayForm>, <affiliation>, <role>, <description> MODS subelements should be concatenated in Dublin Core, separated by a space or other form of punctuation.

Data entry: best practice and standards

Notes Institutions may opt to enter values only for authors from their own institution. Be aware, however, that an author may have belonged to another institution at the time the resource was created and published.

"Role" may be used for in-house search and display purposes only. It should not be mapped to Simple Dublin Core for harvesting.

Rationale It is important for audit purposes that the "submitter" (role term) of a resource be recorded. It may be desirable in certain cases to display an "editor" (role term) of a conference paper, or a "supervisor" (role term) of a thesis that is a deposited resource.

contributor (DC term)

SN Use the "contributor" term for persons responsible for making contributions to the resource (e.g. thesis supervisors, editors) but who are not also responsible for creating the resource.

Contributor is used as the default for "author" in some repositories, eg DSpace. However, the Dublin Core "creator" term is repeatable for multiple authors and by DCMI definition is intended for persons responsible for the creation of the resource, so "creator", not "contributor" should be used for authors.

RT	creator affiliation role

Crosswalks to Dublin Core

MARC	MARC: 720 $a $q $e
Notes	Repository records will have entries for editors of conference publications, supervisors of theses, and names of submitters of resources to the repository who are not authors of those resources. These names should be stored in a MARC 720 tag, with $e to indicate their relationship to the resource (submitter, editor, supervisor, etc.). Do not enter these in a MARC 700 tag in order to avoid confusing them with creators when mapped to DC.
MODS	MODS: <name><namePart>
Notes	MODS puts all names in a repeated <name> with type of contribution indicated in <role>. It does not make the explicit distinction between creator and contributor in terms of primary vs. secondary roles.

Data entry: best practice and standards

Note	Be careful about implications of entering all contributors to a resource, and of making them available to DC mapping. It is not compulsory to have a dc.contributor field and it may be omitted altogether with some resource types. Enter the names of multiple contributors of a resource in the same order in which they appear on the resource.
Rationale	In the case of supervisors of theses, for example, supervisors change throughout the creation of a thesis, belong to different institutions, have different statuses and degrees of involvement with the author. Relationships among any of these can also sometimes be a sensitive issue. The order in which names appear on a resource can have significance.

date (DC term)

SN	The Dublin Core definition of the “date” term is “a date associated with the life cycle of the resource”. This means that a resource can have multiple dates associated with it (e.g. date of creation, of submission to a publisher, of publication, of accession to repository, of a subsequent modification of the repository record).
RT	coverage

Crosswalks to Dublin Core

MARC MARC: 008/00-05
MARC: 008/07-14

MARC: 260 $c

Notes Only crosswalk to Dublin Core dates that are important for discovery.

The subfield $c can be repeated for multiple dates. However MARC does not allow for the potential range of dates associated with a resource in a repository. Accession to repository and subsequent modification dates, for example, should ideally be covered automatically as part of the versioning metadata updates. In some cases it may be necessary or desirable for authentication and audit records to maintain date notes in a MARC 5XX tag.

MARC 260 $c is normally used for date of publication.

MODS MODS: <originInfo><dateIssued>
MODS: <originInfo><dateCreated>
MODS: <originInfo><dateCaptured>
MODS: <originInfo><dateOther>

Notes Only crosswalk to Dublin Core dates that are important for discovery.

Record Creation dates should be machine generated:
<mods:recordCreationDate
encoding="iso8601">20030331</mods:recordCreationDate>

Data entry: best practice and standards

Notes Dates should be formatted in the Dublin Core according to the W3C encoding rules for dates and times[13]. Accordingly be careful not to map any accompanying letters with a date (e.g. “ca.”) to the dc.date element.

Dates required for versioning, authentication and reporting purposes in repositories:

a. **Date of issue**:

the date the resource was issued or published, or if not intended for publication, the date the resource was completed or made public. For purposes of reporting this date should be entered in full (year-month-date in a standardized format).

b. **Date of submission**:

Use the date the first draft of the resource was submitted for publication. Often the version of the resource that is submitted for the repository is the preprint. But even if it is not, use the earlier date as the indication of when the essential substance of the work was first produced.

c. **Date entered in the repository**:

This date is for internal and/or display purposes and need not be mapped to Dublin Core. It should be machine generated.

d. **Date modified in the repository**:

This date is for internal and/or display purposes and need not be mapped

to Dublin Core. This date should be machine generated.

If a work was “first” published on two different dates (e.g. online and print versions being published in different years) use the date the work first appeared in public domain regardless of format.

In case of unpublished theses, use the date the degree was awarded rather than the date the thesis was completed.

To locate the date of publication:

- the year of publication must be stated within or on the resource
- if a journal article or conference publication (not other types of resources) is web-based or in digital format and no year of publication is stated within or on the resource, consult a letter from a journal editor or conference organiser for the year of publication. (a letter from an editor or conference organiser cannot override a year of publication stated within the resource.)
- if no date exists within or on a conference publication then use the date the conference was held as the year of publication
- do not rely on copyright dates or ‘date last updated’ that appear on web pages to indicate the date of the publication of the resource
- if a resource is known to have been published after the publication date (not created date) contained within or on the resource, use the printed publication date

Rationale Intention of the Date of Issue is to indicate when the article was first published, or if not published, first made public.

description (DC term)

SN A description is an account of the content of the resource. The Dublin Core definition includes abstracts, tables of contents, or any free text summary or account of the content.

RT subject
coverage

Crosswalks to Dublin Core

MARC MARC 520 $a (Either leave the MARC indicators ## or modify according to the type of summary: e.g. 3# for abstract)

MARC 5XX $a

Notes Any number of 5XX tags can in theory be mapped to a dc.description element. Thus if one chose to use, say, an inhouse 599 tag to describe the peer-review status of a work, then this peer review value could be mapped from that 599 tag to a dc.description element.

It is best practice to always populate a MARC 520 tag to map to a

dc.description element.

MODS MODS: <abstract>
MODS: <note>
MODS: <tableOfContents>

Notes

Data entry: best practice and standards

Notes If no abstract is available enter the summary. If no abstract or summary accompanies the resource enter descriptive sentences from the introduction or conclusion or title or table of contents, or briefly summarize in own words. If no other alternative repeat the main portion of the title. Do not leave this field blank. Precede entries that are not abstracts with a bracketed indicator of the nature or source of the entry (e.g. [Conclusion]:)

Rationale Even though “abstract” is the default heading for a description in some repositories, the data is typically mapped to dc.description which is broader in scope that strict “abstracts”. This is one of the most useful fields for both users and harvesters. (Some harvesters, e.g. OAIster, even have display problems if this field is not populated in a record.)

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)

subject (DC term)

SN The topic or content of the resource. Subjects will include keywords, keyword phrases and controlled vocabularies.

RT abstract
coverage

Crosswalks to Dublin Core

MARC MARC: 650 $a $b $x $y $z $2 (controlled vocabularies)

MARC: 653 $a $a $a . . . . (keywords)

Also: 600, 610, 611, 630, 651

Notes In Australian and New Zealand repositories the 650 tag will include the **RFCD or Marsden code** (both number and descriptive label). A repository may be configured so as to hide this from the main portal display page of a record if desired, while still retaining it for browse-indexing and search purposes.

In other subject entries it is best practice to include the **$2 subfield** to indicate the source of any controlled vocabularies used. The repository is to be potentially accessible to a wider community than traditional libraries so it will not always be obvious to service providers how to interpret a value in this field unless it is explained in the $2 subfield (e.g., $2 LCSH).

Keywords and keyword phrases will be entered in the 653 tag, each contained within a separate $a subfield within the one 653 tag.

MODS MODS: <subject>
<topic>
<name>
<occupation>

MODS: <classification>

Notes MODS also designates the authority for the controlled vocabulary: <mods:subject authority="lcsh"><mods:topic>

Data entry: best practice and standards

Notes Multiple RFCD codes may be entered.

It is not best practice to use RFCD or Marsden codes for the general search function.

It is not best practice to rely entirely on keywords for subject entries. Editors should monitor keywords used and add additional ones if appropriate to the repository record. This will not affect the keywords as chosen by the author on the resource, but may be advisable for more effective search and retrieval purposes.

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)

When mapping keywords for display, configure them so that each keyword or keyword phrase is separated by a semicolon.

Maintain a common standard regarding capitalization or noncapitalization of keywords within the repository.

Rationales Cross disciplinary research makes multiple RFCD codes obligatory.

RFCD and Marsden codes are designed for government administration, research and reporting purposes, and not for topic searching. They are useful among academics within the same research fields and and who report their research to the same national jurisdiction. These academics know and use the codes more than other users of the repository.

Keywords can sometimes be chosen according to the transient fashion of the day and suffer from limited long-term value. Some authors also choose keywords that have very narrow applicability within their specialist field with the result that a broader topic more useful for search and recovery purposes can be omitted altogether.

Semicolons are becoming the standard practice in repositories such as EPrints and DSpace since they potentially cause less confusion for users who often see commas used to separate multiple parts of a single name or topic name.

A common standard of format within the repository enhances its professional image.

coverage (DC term)

SN DCMI definition: The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant.

DCMI comment: Spatial topic may be a named place or a location specified by its geographic coordinates. Temporal period may be a named period, date, or date range. A jurisdiction may be a named administrative entity or a geographic place to which the resource applies. Recommended best practice is to use a controlled vocabulary such as the Thesaurus of Geographic Names [TGN[14]]. Where appropriate, named places or time periods can be used in preference to numeric identifiers such as sets of coordinates or date ranges.

RT description
subject

Crosswalks to Dublin Core

MARC MARC: 033 $a (Formatted date/time and/or coded place of creation, capture, or broadcast associated with an event.)
MARC: 043 $a (Geographic area codes associated with an item.)

MARC: 513 $b (period covered by a report)
MARC: 522 $a (geographic coverage)

MARC: 650 $y $z (chronological and geographic subdivisions)
MARC: 651 $y $z

MARC: 752 $a (hierarchical place name)

Notes Recommended best practice by DCMI is to use controlled vocabularies for this value. Hence the MARC 033 (date = dc.coverage.temporal) and MARC 043 (geographic area code = dc.coverage.spatial) – although both temporal and spatial coverage in Simple Dublin Core sit in the (repeatable) dc.coverage element.

Not all MARC fields need to be mapped to DC. Tag 651 can be mapped to both dc.subject and dc.coverage if other MARC tags listed above are not used.

MODS MODS: <subject>
<geographic>
<temporal>
<hierarchicalGeographic>
<cartographic>

Data entry: best practice and standards

Notes It is best practice to ensure temporal and spatial values are mapped to dc.coverage elements.

Rationale Populating the Dublin Core “coverage” field eliminates the risk of a harvester supplying default values.

language (DC term)

SN The language of the intellectual content of the resource.

RT --

Crosswalks to Dublin Core

MARC MARC/35-37

MARC 041 $a

Notes Note, it is not best practice to use the MARC 546 language note tag to map to dc.language. This MARC tag is a free text note. Coded values for languages are entered in 041 and/or the fixed 008/35-37 fields. A three letter MARC code for language is used in the 008/35-37 field but this can be mapped to a default DC language value en-aus.

MODS MODS: <language>

Notes e.g. <mods:language authority="rfc3066">en</mods:language>

Data entry: best practice and standards

Notes	The recommended best practice by DCMI is to use a controlled vocabulary such as RFC 3066 which, in conjunction with ISO 639, defines two- and three-letter primary language tags with optional subtags. So Australian English is represented as en-aus, which is made up from en (RFC3066) and aus (ISO 639).
Rationale	OAI harvesters typically search for standard codes to indicate language.

publisher (DC term)

SN	DC definition: An entity responsible for making the resource available. A publisher can be the author's institution or a commercial publisher.
RT	Place of publication

Crosswalks to Dublin Core

MARC	MARC: 260 $b MARC: 773 $d
Notes	773 $d is the subelement of the host item of the resource, and includes the place, publisher, date of publication as a text string in a single subfield (e.g. $d [Berlin], Elsevier, 2007)
MODS	MODS: <originInfo><publisher>
Notes	--

Data entry: best practice and standards

Notes	Enter the publisher in full, except for Pty and Ltd. Where there is a hierarchy in the publishing organization enter the broadest umbrella institution first followed by successive narrower institutions. In case of unpublished works (e.g. theses) that emanate from the institution, enter the institution as the value in the publisher field.
Rationale	The Dublin Core definition of publisher, and therefore the DC expectation in this element, is any entity that is responsible for making the resource available. Hence a university can appear as “publisher” of a thesis even though the thesis is not strictly “published” in a commercial sense.

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)

Place of publication

SN	The geographic location of the publisher of a resource.
RT	publisher

Do not crosswalk to Dublin Core

MARC	MARC: 260 $a MARC: 773 $d
Notes	773 $d is the subelement of the host item of the resource, and includes the place, publisher, date of publication as a text string in a single subfield (e.g. $d [Berlin], Elsevier, 2007)
MODS	MODS: <originInfo> <place> <placeTerm type="text"> MODS: <originInfo> <publisher>
Notes	The latter MODS entry is the equivalent of the MARC 773 $d host item subelement.

Data entry: best practice and standards

Notes	Enter the place of publication value.
Rationale	The place of publication may be expected for bibliographic citation purposes. Some repositories collate data including place of publication to generate standard bibliographic citations of the resource.

rights (DC term)

SN	DC definition: Information about rights held in and over the resource.
RT	--

Crosswalks to Dublin Core

MARC	MARC: 506 $a (restrictions on access note) MARC: 540 $a $u (terms governing use and reproduction)
Notes	The 540 tag will normally have the official copyright statement from the copyright owner and/or publisher of the resource. Where a link to an online copyright statement is required use the $u subfield for the URI. The 506 tag will explain in further detail any access limitations on the resource.
MODS	MODS: <accessCondition>
Notes	MODS combines the 2 MARC tags values into the single <accessCondition>

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)

element, and can optionally distinguish between them by “type”:

<accessCondition type=“restrictionOnAccess”>
<accessCondition type=“useAndReproduction”>

Data entry: best practice and standards

Notes	DCMI Glossary comment[15]: Typically a Rights element will contain a rights management statement for the resource, or reference a service providing such information. Rights information often encompasses Intellectual Property Rights (IPR), Copyright, and various Property Rights. If the rights element is absent, no assumptions can be made about the status of these and other rights with respect to the resource.
Rationale	Publishers will often require a link to their page or a standard copyright statement. These can be pasted into this field

relation (DC term)

SN	Dublin Core relation is a related resource. For a conference paper or journal article or book chapter this would mean the conference publication, the journal title and the book title respectively. Conference names and series titles are also “relations” of resources.
RT	title

Crosswalks to Dublin Core

MARC	MARC: 440 $a MARC: 490 $a MARC: 530 $a MARC: 534 $a MARC: 710 $a MARC: 711 $a MARC: 773 $t MARC: 787 $t MARC: 830 $a MARC: 856 $u
Notes	Do not use 710 $a for a creator's affiliation. Creator affiliations are covered by $u in the 100 and 700 tags. The host item entry (MARC 773) can contain **complete bibliographic information** for the host item of the resource: **$t title** (e.g. journal title)

$d place, publisher, date of publication (e.g. [Berlin], Elsevier, 2007)
$g relationship information (e.g. Vol. 17, no. 98 (Feb. 2007), p. 78-159)
$n note (e.g. peer reviewed)
$x ISSN
$z ISBN

The MARC 787 tag is a nonspecific relationship entry and may contain other types of data apart from a host item for a resource. Check this field in the case of a batch upload.

Be careful to distinguish bibliographic data pertaining to the resource from data pertaining to its related title. An ISSN applies to a journal publication and belongs in the 773 tag, not in the 022 tag that would refer to the main title entry for the journal article.

Use the MARC 856 $u field and subfield, with $q for the format (e.g. application/PDF) for the offsite DOI or other offsite URI to the article.

MODS MODS: <relatedItem>

Notes <relatedItem> data is parsed into subelements in MODS (any MODS element may be used). For example, if giving a reference to a resource fully described in MODS relatedItem, one could use:
<relatedItem><identifier>
and/or title of a resource:
<relatedItem><titleInfo><title>

Data entry: best practice and standards

Notes Do not use this for links to other instances of the same resource, or to links or bibliographic references to the same resource.

"Relation" also strictly includes other versions of the resource (e.g. a preprint of a published version) but there is no scope at this stage to monitor this level of relationship in repositories.

Rationale Relation is defined by DCMI to mean a related resource. Do not use for another instance of, or link to, the same resource.

identifier (DC term)

SN DCMI definition[16]: An unambiguous reference to the resource within a given context. Recommended best practice is to identify the resource by means of a string or number conforming to a formal identification system. Examples of formal identification systems include the Uniform Resource Identifier (URI) (including the Uniform Resource Locator (URL), the Digital Object Identifier (DOI) and the International Standard Book Number (ISBN).

DCMI guidelines[17]: This element can also be used for local identifiers (e.g. ID numbers or call numbers) assigned by the Creator of the resource to apply to a particular item. It should not be used for identification of the metadata record itself.

RT --

Crosswalks to Dublin Core

MARC

MARC: 013 $a

MARC: 020 $a
MARC: 022 $a
MARC: 024 $a

MARC: 773 $x $z

MARC: 852 $u

Notes Only crosswalk identifiers of the resource to Dublin Core, not the identifiers of the host publishing title. The MARC fields in italics above will typically represent identifiers of host items only and should not be mapped to Dublin Core.

Note for OAI harvesting: It is expected that a service provider (harvester) will direct users initially to the resource's metadata page in the repository. This is indicated in the MARC 852 field. (From there users can navigate to the full text of the resource.) In order for service providers to direct users to the repository's metadata page for the resource, the identifier in OAI data provider's Dublin Core record must be the identifier of the metadata page itself, not the full text of the resource. This identifier will normally be machine generated. All other DC values in this DC record will relate to the resource, not the metadata page with the link to the resource.

MARC 020 $a should only be used when the resource itself is identified by an ISBN (e.g. a book). Otherwise enter ISBN details in the host item subfield 773 $z.

MODS

MODS: <identifier>

MODS: <location> <URL>

Notes The identifier type (e.g. <identifier> *with* type="doi") should be retained and associated with the identifier value.
e.g. <mods:identifier type="uri">http://palmm.fcla.edu/feol/[18]
</mods:identifier>

Follow standards for entry of identifiers. With ISSN and ISBN follow a single standard for data entry: e.g. hyphen in an ISSN and all digits and letters in an ISBN together without spaces or dashes and capitalized.

Local identifiers will be machine generated:
<mods:recordIdentifier>12345</mods:recordIdentifier></mods:recordInfo>

Data entry: best practice and standards

Notes Do not map local identifiers (e.g. PIDS – persistent identifiers – a local unique repository identifier) to Dublin Core

Rationales	Local identifiers do not have significance outside the repository institution.

source (DC term)

SN	DCMI definition: The resource from which the described resource is derived.
RT	--

Crosswalks to Dublin Core

MARC	--
Notes	Do not crosswalk manual data entries to the Dublin Core element “source”.

MODS	--
Notes	Do not crosswalk manual data entries to the Dublin Core element “source”.

Data entry: best practice and standards

Notes	This value may be machine generated to indicate the name of the file from which the metadata is generated.
Rationales	DSpace and EPrints recommend that it not be used for data entry. ARROW Discovery Service does not scan for “dc.source”.

type (DC term)

SN	DCMI definition: The nature or genre of the content of the resource. DCMI recommended best practice is to use a controlled vocabulary such as the DCMI Type Vocabulary [DCMITYPE[19]]. This thesaurus however is limited for repository purposes. Repository supports have created other controlled lists. To describe the file format, physical medium, or dimensions of the resource, use the Format element.
RT	--

Crosswalks to Dublin Core

MARC	MARC: 655 $a $2
Notes	Use a controlled thesaurus. Indicate the source of the thesaurus (e.g. $2 LCSH)

MODS	MODS: <typeOfResource>

MODS: <genre>

Notes Use separate instances of Type for each MODS element value.

If converting MODS typeOfResource values to Dublin Core Resource Type values, see conversion details below. If MODS <genre> contains authority="dct", that may be used in dc:type and typeOfResource dropped.

Conversion of MODS typeOfResource values to DC Resource Type vocabulary

MODS typeofResource	**DC Type value**
typeOfResource collection="yes"	Collection (use in addition to specific value below)
software and mods:genre="database"	Dataset
cartographic material	Image
multimedia	InteractiveResource
moving image	MovingImage
three-dimensional object	PhysicalObject
software and mods:genre="online system or service"	Service
sound recording, sound recording-musical, sound recording-nonmusical	Sound
still image	StillImage
software	Software
text, notated music	Text

Data entry: best practice and standards

Notes

Do not repeat this element. (all other elements may be repeated.)

RUBRIC is involved with MACAR[20] in looking into the possibility of reaching guidelines towards national standards for a resource type vocabulary in repositories. A final decision has not yet been made on this list (5th October 2007), but one is expected before the end of the year.

In the meantime, the ARROW Discovery Service harvesting guide[21] states as a rule that a type value “must be one of the ARROW list of recognized types”.

These are:

```
arc project report

article

book

book chapter

collection
```

```
conference paper

email

reading list

multi-media object

research dataset

research paper

rich media (non-text)

still image

technical report

thesis

working/discussion paper
```

The guide also states that if other type values are to be used then the Discovery Service administrator should be contacted. Many repositories, for example, do use variant terms from the ones above. (see the "Current thesauri for resource types" below.)

There are more types that are being considered by MACAR, such as software, musical compositions, datasets and others, and some of the above terms may change. Before the end of the year a more complete "standard" list should be available.

Rationale

Although the ARROW Discovery Service harvesting guide[22] says multiple "type" elements may be supported (presumably for a single document), not all repository solutions currently support multiple resource types for the one document. Hence multiple resource types attached to one document in one repository could run into preservation difficulties if there comes a time for a future migration of data.

It also needs to be kept in mind that the ARROW Discovery Service harvesting guide is currently under review and is expected to be revised soon. The work of MACAR may influence its revision.

Unfortunately there has been little consistency among the many thesauri in use, and some confuse formats or types of resources with genres of resources. Beware confusion between types (genres) and formats (MIME types).

Current thesauri for resource types (See Appendix B for a more complete set of comparisons)

ARROW-VITAL types:	**DSpace 1.4 default types:**	**Fez 1.1 default types:**	**Eprints default types:**
			Article
book	Animation	Book	Book Section
book chapter	Article	Book Chapter	Book
conference paper	Book	Conference Paper	Monograph
image	Book chapter	Conference Poster	Conference/Workshop

journal article	Dataset	Conference Proceedings	Thesis
thesis	Learning Object	Dept Technical Report	Patent
working paper	Image	Generic Document	Other
	Image, 3-D	Journal Article	
	Map	Magazine Article	
	Musical Score	Seminar Paper	
	Plan or blueprint	Software	
	Preprint	Thesis	
	Presentation	Working Paper	
	Recording, acoustical		
	Recording, musical		
	Recording, oral		
	Software		
	Technical Report		
	Thesis		
	Video		
	Working Paper		
	Other		

format (DC term)

SN DCMI definition: The file format, physical medium, or dimensions of the resource.

DCMI comment: Examples of dimensions include size and duration. Recommended best practice is to use a controlled vocabulary such as the list of Internet Media Types [MIME].

DCMI reference: http://www.iana.org/assignments/media-types/[23]

In repositories the file format metadata value will typically be machine generated.

RT --

Data entry: best practice and standards

MARC 856 $q (electronic file format type: e.g. $q application/pdf)

Other possible format crosswalks:

245 $h (medium)

300 $a (physical description)
306 $a (playing time)

533 $e (physical description of a reproduction)

Notes In repositories the file format metadata value of the archived resource will typically be machine generated.

The above crosswalks are for additional format values that a repository policy may choose to add to a record.

MODS	MODS: <physicalDescription>
	MODS: <internetMediaType>
	MODS: <extent>
	MODS: <form>
Notes	Use separate instances of Format for each MODS element value.

Data entry: best practice and standards

Notes	Format information of a resource has a different function in a repository from what it has in traditional libraries. Repository format is principally for machine information. MARC non-electronic format information is principally descriptive data for the user.
Rationale	In repositories the file format metadata value of the archived resource will typically be machine generated.

Index of the MARC fields referred to in the above Guidelines

	MARC field	**MARC subfield**	**DC element**
Language	008	eng	language
Patent control	013	a	identifier
ISBN	020	a	identifier
ISSN	022	a	
Identifier	024	a	identifier
Date of an event	033	a	coverage
Language	041	a	language
Geographic area code	043	a	coverage
Personal name	100	a	creator
Role	100	e	
Affiliation	100	u	
Title	245	a, b, p, n	title
Alternate title	246	a, b	title
Edition	250	a	
Publisher, place	260	a	
Publisher	260	b	publisher
Date	260	c	date
Format	300	a	
Title of series	440	a	relation

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)

Title of series	490	a	relation
General note	500	a	description
Dissertation	502	a	description
Bibliography	504	a	description
Contents, formatted	505	a	description
Access restrictions	506	a, u	rights
Credits	508	a	contributor
Participants/performer	511	a	description
Report, period covered	513	b	coverage
Summary, abstract	520	a	description
Audience	521	a	description
Geographic coverage	522	a	coverage
Citation	524	a	identifier
Copyright	540	a, u	rights
Language note	546	a	language
ARROW: RQF note	591		
ARROW: RQF note	592		
ARROW: RQF note	593		
Subject	600, 610, 611, 630, 650, 651	a, x	subject
Subject	650, 651	y, z	coverage
Subject: uncontrolled	653	a	subject
Personal name	700, 720	a	creator, contributor
Role	700, 720	e	
Affiliation	700, 720	u	
Corporate name	710	a	contributor
Conference	711	a, c, d, n,	contributor
Conference note	711	g	description
Hierarchical place	752	a	coverage
Host item	773	t	relation
Title of series	830	a	relation
Non-repository URI	852	u	relation
URI	856	u	identifier

File Naming for Repositories

This document offers guidelines for deciding why and how to apply consistent procedure to the file names in a repository. It assumes a basic knowledge of filenaming structure, such as

the necessity for the three or four digit extension indicating document type (".pdf", ".doc", ".jpeg").

This document does not prescribe a single convention for filenames. Much will depend on the nature of the content and different files deposited in the repository. For example, multimedia or image files without single titles, unpublished preprints, institutional working papers may be better handled with different naming methods. Workflows will also impinge on filenaming as decisions about editing names will embrace the input of submitting authors balanced against editorial policies and practices, and scheduling and staffing constraints balanced against the need for naming conventions being a small but vital part of a larger collaborative environment.

Preliminary considerations

Prior planning of filenaming conventions to be used can reduce risks of necessity to change file names down the road, avoiding broken links and bookmarks.

Relevance to repository clients:

- Filenames are not always visible in main portal display of all repositories but they are visible in the address box of the browser and as the name of the file that is downloaded by the user.
- If the filename assigned by the depositor is visible to users when the file is downloaded, the filename will become the reference for the user to locate, store and cite the file. For these reasons RUBRIC considers it best practice to apply a meaningful filename convention.

Relevance to administrative staff:

- Will administrative staff at any time need to retrieve a copy of the file again even after it is uploaded into a repository? As a backup copy? For future reference in event of questions about integrity of the uploaded file? As an 'at hand' copy in event of later discovery of uploading errors or wrong files attached to records? These are among the reasons files are retrieved by those entering the data into repositories: a useful convention in assigning filenames is therefore again best practice.
- Consistently and well named files can assist by making them easy to locate for maintenance and updating.

Filenaming conventions do's and don'ts and considerations:

“Remember that you are not only storing information but also providing a collaborative environment for others.” (http://www.alexport.net[24])

DO:

- **factor in self-submission.** If one is intending to encourage academics to submit their archive their own works in the repository then consider how a file naming convention may impact on them and their willingness to adapt. Or will the library be willing to change the names of submitted files? Perhaps some faculties or centres may wish to adopt their own file naming schemes. Decisions about establishing a filenaming scheme will need to consider these issues.

- **consider the use of an author's name**, especially as the first characters of the filename, is a logical choice for files authored by the one person. However many documents have multiple authors within the institution. Does one want 2 different conventions: one for documents with a sole author and another for documents with multiple authors? Or is it best to avoid author names altogether and focus on document type, title, date?
- **keep it relatively short**. Adding authors' names and initials, detailed dates (20060810), item types (e.g. prp for preprint, ja for journal article, etc.) may be informative but consider the usefulness of such information once the file is uploaded and archived. Too much information could make a filename too unwieldy to be worth the effort. Theoretically a filename can have 128 or more characters depending on the operating system used. In practice, however, many software systems break long filenames. And the longer the filename the more likely there will be errors in both creating and retrieving it.

DON'T:

- **use punctuation characters.** Some characters such as / : * ? " < are reserved by the operating system's shell. Spaces should be avoided or replaced with hyphens or underscores particularly with images. (Underscores, however, may sometimes be confused for spaces.)
- **use upper case characters**. Best practice is not to use upper case characters since upper and lower case characters are not treated uniformly across all operating systems. Consistency makes them easier to work with, too.
- **use spaces**. Some web browsers will discard anything after a space.
- **use a repository's PID** (e.g. rubric156, arrowdev1421, vital1043) since this will lose its meaning and possibly cause confusion in the event of a future migration to another repository.

Considerations:

- **use of abbreviations**: before deciding on institutional or other abbreviations consider how long-term is their potential relevance and obvious meaning. Also avoid anything difficult to spell.
- **a brief yet meaningful** filename may increase its potential for impact and recall by users, both public and administrative. Is the filename clear and mnemonic? There will be less likelihood of making mistakes in assigning such a filename. Consider a filename consisting of just a few title keywords.
- **use of numbers**. Avoid the use of numbers unless they clearly represent, say, a date or a publication issue number. If using a date then ensure that it will always be entered in a consistent format.
- **there should only be one "."** in the entire filename, and it should be right before the file extension.

The most short and meaningful file naming scheme may require a minimum of record keeping offline. Hence if a filing scheme is to have a unique number for each year this may require a simple and easy to manage offline list of used numbers for each year. Or it may be even simpler and useful for users if the filename consisted of a few keywords of the title of the document.

Examples of filenaming conventions

Some specific filenaming conventions are described on the following websites. Keep in mind that these sites have different purposes to many university repositories, so should only be

seen as a guide to the principles of filenaming in other contexts.

File Naming Conventions for Digitally-stored Images[25]

TASI (Technical Advisory Service for Images)[26]

Ontolog-forum[27]

The Alexport[28]

Other filename examples -- from university repositories:

Example 1

Filename:
[author/s lastnames] + [year] + [journal title initials] + [doc type]. [extension]
csergogoldgeier05EUSApaper.pdf

Original document:
Csergo, Zsuzsa and Goldgeier, James M. (2005) "The European Union, the post-communist world, and the shaping of national agendas" . In European Union Studies Association (EUSA) > Biennial Conference > 2005 (9th), March 31-April 2, 2005, pages 31.

Example 2

Filename:
[abbreviated title] + [author/s lastnames] + [year]. [extension]
The_genetic_contribution_-_Ball_et_al._2000.pdf

Original document: The genetic contribution of single male immigrants to small, inbred populations: a laboratory study using Drosophila melanogaster by Stephen J. Ball, Mark Adams, Hugh P. Possingham & Michael A. Keller. -- Published in Heredity (2000)

Example 3

Filename:
[author/s lastnames] + [journal title initials] + [year]. [extension]
Eichhorst-Rhein-EUSA-2005.pdf

Original document:
Eichhorst, Werner and Rhein, Thomas. (2005) "The European Employment Strategy and Welfare State Reform: The Case of Increased Labour Market Participation of Older Workers" . In European Union Studies Association (EUSA) > Biennial Conference > 2005 (9th), March 31-April 2, 2005, pages 33, Austin, Texas.

Example 4

Filename:
[author initials] + [journal title initials] + [journal issue]. [extension]
bw_olc_17_3_02.pdf

Original document:
Customer relationship management â€“ your biggest brother? By Belinda Weaver, published

in Online Currents Vol.17, no. 3, April 2002.

Example 5

Filename:
[journal title initials] + [journal issue]. [extension]
anjh07.pdf

Original document:
For Australian Newspaper History Group Newsletter June

Example 6

Filename:
[author lastname] + [journal title initials] + [journal issue]. [extension]
jull_ia_3_01.pdf

Original document:
Nunavit: the still small voice of indigenous governance by Jull, published in Indigenous Affairs 3/01

Example 7

Filename:
[department] + [dept publication initials] + [journal issue] + [journal date]. [extension]
econ_dp_341_jul05.pdf

Original document:
Laurenceson, James and Qin, Fengming (2005) China's Exchange Rate Policy: The Case Against Abandoning The Dollar Peg Discussion Paper No. 341, School of Economics, The University of Queensland.

Example 8

Filename:
[author initials] + [jounal title initials] + [journal issue] + [journal year]. [extension]
wh_ki_s78_05.pdf

Original document:
A Chronic Disease Outreach Program for Aboriginal Communities Wendy E. Hoy, Srinivas Kondalsamy-Chennakesavan, Joanne Scheppingen, Suresh Sharma, and Ivor Katz. Kidney International (2005) 68, S76â€“S82

Example 9

Filename:
[abbreviated title] + [author/s lastnames] + [year]. [extension]
The_genetic_contribution_-_Ball_et_al._2000.pdf

Original document:
The genetic contribution of single male immigrants to small, inbred populations: a laboratory study using Drosophila melanogaster by Stephen J. Ball, Mark Adams, Hugh P. Possingham & Michael A. Keller. -- Published in Heredity (2000)

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)

Example 10

Filename:
[publication initials] + [journal issue]. [extension]
MITJPSPGC_Rpt133.pdf

Original document:
This is for an MIT Report, No.133, published by the Joint Program on the Science and Policy of Global Change. (Upper case is not recommended best practice.)

Example 11

Filename:
[author last name] + [author last name] + [year]. [extension]
Woodside_Summers_Morgan_2006.pdf

Original document:
Woodside, Frances and Summers, Jane and Johnson Morgan, Melissa (2006) 'Sponsorship of fast moving consumer goods â€“ does packaging endorsement contribute to brand attitude? a conceptual paper.' In: Fullerton, Sam and Moore, David, (ed.). International Business Trends: Contemporary Readings, Academy of Business Administration, pp. 213-221. (International Conference of the Academy of Business Administration, 16 - 20 Aug 2006, Munich, Germany.)

Example 12

Filename:
[title + [document type]. [extension]
Henryk_Grossman_and_the_recovery_of_Marxism_preprint.pdf

Original document:
Title: Henryk Grossman and the recovery of Marxism
Authors: Kuhn, Rick
Date Created: 2004

Conclusion

- Does a file naming scheme matter? (Who will use the filenames and why?)
- Consider criteria for long term sustainability, uniqueness and usefulness. This requires thoughtful planning.
- Don't forget appearance, image and mnemonics, they may be among the most noticed characters in the repository (but will they be noticed?)

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)

APPENDIX A

Templates

(Revised and updated 21 December 2007)

MARC book template

MODS book template

MARC book chapter template

MODS book chapter template

MARC conference paper template

MODS conference paper template

MARC creativework template

MODS creativework template

MARC graduate project template

MODS graduate project template

MARC herbarium template

MODS herbarium template

MARC image template

MODS image template

MARC journal article template

MODS journal article template

MARC report template

MODS report template

MARC thesis template

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)

MODS thesis template

MARC working paper template

MODS working paper template

APPENDIX B

Metadata tools and mapping spreadsheets

Resource and MIME Types Thesauri

Resource or item type thesauri are still evolving, and there are different controlled lists used by the different repository solutions.

The metadata types spreadsheet contains

- a link to the list of the registered **MIME types** recognized by Dublin Core
- and the item **resource types** used by
  - Dublin Core
  - VITAL
  - Fez
  - DSpace default
  - EPrints 2.1 and 3.0
  - MARC and MODS
  - Common citation tools (EndNote, RefWorks, Zotero, CiteULike)
  - RAE and PBRF
  - Various Australian university and research departments
- and a **table of comparisons** between these different thesauri

Metadata Spreadsheet

The metadata spreadsheet displays

- a **mapping** overview across
  - MARC
  - Dublin Core (Simple and Qualified)
  - VITAL, DSpace, Eprints applications of DC
  - Picture Australia application of DC

    - MODS
    - ETD (including ADT theses)
    - IEEE LOM metadata
    - ARK handle metadata
- **Harvesting** relevance of the respective metadata fields, with particular reference to
    - National Library of Australia's ARROW Discovery Service
    - New Zealand's National Research Discovery Service
- Notes on the **formatting of data entry** in specific fields
- Notes for **stylesheet editing** to facilitate correct portal display
- References to specific data entry **standards and exceptions**
- **New Zealand**'s National Research Discovery Service metadata entry guidelines
- **XML Namespace,** MARC Leader and 008 record data
- Links to DCMI listed **metadata tools**
- **Use of the Data Filters** to focus one's search across specific metadata and discovery elements.

Resource types metadata spreadsheets

The book metadata spreadsheet narrows the metadata registry information to its application for the book resource type.

The book chapter metadata spreadsheet narrows the metadata registry information to its application for the book chapter resource type.

The conference paper metadata spreadsheet narrows the metadata registry information to its application for the conference paper resource type.

The journal article metadata spreadsheet narrows the metadata registry information to its application for the journal article resource type.

The thesis metadata spreadsheet narrows the metadata registry information to its application for the thesis resource type.

The working paper metadata spreadsheet narrows the metadata registry information to its application for the working paper resource type.

The nontext metadata spreadsheet narrows the metadata registry information to its application for **image** and **video** resource types. These are still in very early days under development.

Dublin Core Application Profile for Scholarly Works

The Dublin Core Application Profile for Scholarly Works is a model based on a 2007 Ariadne article by Allinson, Johnston and Powell[29]. It employs both the FRBR and DCMI Abstract models. It is a proposal for further development of metadata descriptions of

scholarly works. It does not relate to repositories as they are functioning at present, it's a planning document hopefully for extending their functions in the near future.

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)

APPENDIX C

Repository display configuration worksheet

The repository portal configuration spreadsheet can be used in conjunction with technical staff who will assist with the configuration of the repository. It's purpose is to assist with deciding and communicating what data is to be displayed in the repository portal and what descriptive labels are to be attached to that data.

List of References

Allinson J & Johnston P & Powell, A 2007, 'A Dublin Core Application Profile for Scholarly Works', Ariadne Issue 50, viewed 22 May 2007, <http://www.ariadne.ac.uk/issue50/allinson-et-al/>

Cairo Project Team 2007, 'Cairo tools survey', The University of Manchester, viewed 22 May 2007, <http://cairo.paradigm.ac.uk/projectdocs/cairo_tools_listing_pv1.pdf>

Dublin Core 2006, 'Dublin Core Metadata Initiative', DCMI, viewed 15 November 2005 <http://dublincore.org/documents/dcmi-terms/>

Iana 2006, 'Internet Assigned Numbers Authority', viewed 21 May 2007, <http://www.iana.org/assignments/media-types/>

JISC 2007, Complex Archive Ingest for Repository Objects (CAIRO)', viewed 6 May 2007, <http://www.jisc.ac.uk/whatwedo/programmes/programme_rep_pres/cairo.aspx>

MarcEdit 2007, 'MarcEdit5 + Linux (and Mac) ', Oregon State University, viewed 6 May 2007, <http://oregonstate.edu/~reeset/marcedit/html/index.php>

MODS 2007, 'Metadata Object Description Schema', The Library of Congress, viewed 12 September 2006, <http://www.loc.gov/standards/mods/>

Network Development and MARC Standards Office, Library of Congress, Naming Conventions for Digital Resources, viewed 09 August 2006 <http://www.loc.gov/marc/naming.html>

Newman C 1997, 'Date and Time Formats', W3C, viewed 16 November 2006, <http://www.w3.org/TR/NOTE-datetime>

Publication of Archival Library & Museum Materials 2005, 'Florida Environments Online', State University of Florida, viewed 21 May 2007, <http://palmm.fcla.edu/feol/>

Sollins, J & Masinter, L 1994, 'Functional Requirements for Uniform Resource Names', viewed 9 August 2006, <http://ietf.org/rfc/rfc1737.txt>

TASI 2007, 'Technical Advisory Service for Images', JISC, viewed 21 May 2007, <http://tasi.ac.uk/advice/creating/filenaming.html>

The Alexport 2003, Shiman Associates Inc., viewed 10 August 2006 <http://www.alexport.net./>

The Getty 2000, 'Getty Thesaurus of Geographic Names Online', The J. Paul Getty Trust,

viewed 21 May 2007,
<http://www.getty.edu/research/conducting_research/vocabularies/tgn/index.html>

Tutorial: Creating an Interactive 3D Product Using VRML, viewed 10 August 2006,
<http://www.virtualrealms.com.au/vrml/tute01/before_you_begin.htm#naming>

U.S. Department of Energy 2005, File Naming Conventions and Directory Structure, viewed 09 August 2006 <http://www1.eere.energy.gov/communicationstandards/web/naming.html>

Wikipedia, ISO 9660, viewed 09 August 2006 <http://en.wikipedia.org/wiki/ISO_9660>

Wilburn G & M 2006, 'File Naming Conventions for Digitally-stored Images', Northern Journey, viewed 21 May 2007,
<http://www.northernjourney.com/photo/articles/filenaming.html>

Yim P 2004, 'File Naming Convention and Simple Versioning', ontolog-forum, viewed 16 November 2006, <http://ontolog.cim3.net/forum/ontolog-forum/2004-02/msg00029.html>

Acknowledgements

Godfrey, N 2007 RUBRIC Central, Metadata Specialist - http://metalogger.wordpress.com[30]

"RUBRIC Toolkit: Entering Metadata Guidelines" produced May 2007, revised December 2007

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)

1 http://www.loc.gov/marc/
2 http://www.loc.gov/standards/marcxml/
3 http://www.loc.gov/standards/mods//
4 http://dublincore.org/tools/
5 http://oregonstate.edu/~reeset/marcedit/html/index.php
6 http://www.jisc.ac.uk/whatwedo/programmes/programme_rep_pres/cairo.aspx
7 http://www.jisc.ac.uk/aboutus/about_jisc.aspx
8 http://cairo.paradigm.ac.uk/projectdocs/cairo_tools_listing_pv1.pdf
9 http://arrow.edu.au/about/
10 http://www.loc.gov/standards/mods/v3/mods-userguide-elements.html
11 http://www.dlib.org/dlib/january01/lagoze/01lagoze.html
12 http://dublincore.org/usage/documents/relators/
13 http://www.w3.org/TR/NOTE-datetime
14 http://www.getty.edu/research/conducting_research/vocabularies/tgn/index.html
15 http://dublincore.org/documents/usageguide/glossary.shtml
16 http://dublincore.org/documents/usageguide/elements.shtml#identifier
17 http://dublincore.org/documents/usageguide/elements.shtml#identifier
18 http://palmm.fcla.edu/feol/
19 http://dublincore.org/documents/dcmi-type-vocabulary/
20 http://www.arrow.edu.au/macar
21 http://arrow.edu.au/docs/files/harvesting.pdf
22 http://arrow.edu.au/docs/files/harvesting.pdf
23 http://www.iana.org/assignments/media-types/
24 http://www.alexport.net/
25 http://www.northernjourney.com/photo/articles/filenaming.html
26 http://tasi.ac.uk/advice/creating/filenaming.html
27 http://ontolog.cim3.net/forum/ontolog-forum/2004-02/msg00029.html
28 http://www.alexport.net/
29 http://www.ariadne.ac.uk/issue50/allinson-et-al/
30 http://metalogger.wordpress.com/
31 http://creativecommons.org/licenses/by-sa/2.5/au/
32 http://www.rubric.edu.au/