

National Research Discovery Service

Metadata Guidelines version 1.0

April 2007



Document Control

Revision history

Revision	Date	Author	Reason for Change
1.0	5 Apr 2007	Gordon Paynter	Initial release based on discussions amongst NRDS Project participants.

Table of Contents

1. INTRODUCTION.....	4
1.1. PURPOSE OF THIS DOCUMENT.....	4
1.2. BACKGROUND	4
1.3. STATUS OF THIS DOCUMENT.....	4
1.4. LINKS TO OTHER DOCUMENTS.....	5
2. GENERAL PRINCIPLES AND ASSUMPTIONS.....	6
2.1. SCOPE OF INSTITUTIONAL REPOSITORIES	6
2.2. INTEROPERABILITY STANDARDS	6
2.3. METADATA TRANSFER.....	6
2.4. IMPORTANCE OF METADATA FIELDS.....	6
3. METADATA GUIDELINES FOR REPOSITORIES	7
3.1. IDENTIFIER	7
3.2. TITLE.....	8
3.3. AUTHOR (CREATOR).....	8
3.4. PUBLISHER	9
3.5. ABSTRACT (DESCRIPTION).....	10
3.6. DATE.....	11
3.7. SUBJECT	12
3.8. TYPE	13
3.9. THESIS.....	15
3.10. RIGHTS	16
3.11. CONTRIBUTOR	17
3.12. COVERAGE	18
3.13. LANGUAGE	18
3.14. RELATION.....	19
3.15. FORMAT	20
3.16. SOURCE	20
4. DERIVED METADATA USED IN THE DISCOVERY SERVICE	21
APPENDIX A: CONTROLLED SUBJECT METADATA OPTIONS.....	25
APPENDIX B: TYPE METADATA OPTIONS.....	27
OVERVIEW	27
COMPARISON OF EPRINTS, DSPACE AND PBRF TERMS	28
APPENDIX C: THESIS METADATA OPTIONS	31
APPENDIX D: NEW ZEALAND THESIS AND DISSERTATION TYPE VOCABULARY	32

1. Introduction

This document contains a set of metadata guidelines for institutional repositories that wish to contribute metadata to the National Research Discovery Service (NRDS).

1.1. Purpose of this document

This document sets out metadata guidelines for use in New Zealand Institutional Repositories that wish to contribute metadata to the NRDS. The metadata guidelines are based on the current and future metadata standards used at New Zealand research institutions in their institutional repositories.

Note that this document discusses metadata from a specialised point of view: how it can be used for research discovery in the NRDS. This is a narrow view of the use of metadata, and is not intended as a complete overview of metadata best practice.

Future versions of this document may take a more holistic view of best practice for institutional repository metadata.

1.2. Background

The goals of the NDRS Project include building a national research discovery service; and promoting best practice, consistency and the use of standards in New Zealand's research repositories. Both of these goals are dependent on a network of institutional repositories with high-quality metadata.

The NRDS Project Planning Meeting held at the National Library on 30 January 2007 included a lengthy discussion of metadata standards for New Zealand institutional repositories as they pertain to the proposed discovery service. The participants agreed:

1. To work towards a common set of metadata guidelines for use in their repositories;
2. That a set of baseline standards will be developed based on unqualified Dublin Core metadata, and an enhanced specification will also be developed using qualified Dublin Core;
3. On the best practices to use for several metadata fields (in most but not all cases a consensus was reached);
4. That the objective of the process was a set of metadata guidelines and a qualified Dublin Core metadata schema appropriate for use in OAI-PMH harvests.

In subsequent discussions by email, nzir-l subscribers recommended best practices for other metadata fields, and a further four drafts were prepared.

The NRDS Project Metadata Meeting was held at the National Library on 05 March 2007 to discuss finalise the third draft. The participants agreed:

1. That representatives of the repositories may produce future versions of the document with a more general focus.

1.3. Status of this document

This document is the initial release. It is anticipated it may need to be further refined during the initial implementation project.

1.4. Links to other documents

This document should be read in conjunction with the following NRDS Project documentation:

1. National Research Discovery Service Project Plan
[nl_cims-#218885-National Research Discovery Service Project Plan](#)
2. NRDS Project Planning Meeting Agenda
[nl_cims-#226394-NRDS Project: Planning Meeting Agenda \(2007-01-30\)](#)
3. NRDS Project: Planning Meeting Minutes
[nl_cims-#227946-NRDS Planning Meeting Minutes - 30 January 2007](#)
4. NRDS Project: Discovery System Requirements:
[nl_cims-#226236-NRDS Project: Discovery System Requirements](#)

2. General principles and assumptions

2.1. Scope of institutional repositories

Institutions are advised to develop policy around the scope of their repository, and to consider issues such as the types of document that will be included and whether to allow staff to archive research outputs that were produced at other institutions.

The NRDS has no requirements around scope issues, and will harvest all metadata records exposed by the source repositories. However, it does make the following recommendations.

1. Each institution will use its own definition of a “research output” for inclusion in the NRDS. We recommend that research outputs be limited to the types listed in the PBRF List of Research Output Types, and specifically that Learning Objects be excluded.
2. Each institution will set its own policy about whether metadata-only records should be added to the repository when the resource is not available in a digital form. We recommend that these resources are not included because users are likely to expect all resources in NRDS to be available in digital form.
3. Each institution will set its own policy about whether to supply digital objects that are in a repository in digital form, but which are not freely accessible due to copyrights, embargoes or other rights restrictions. We recommend that these resources are included in the repository and harvested by NRDS.

2.2. Interoperability standards

1. International standards and best practices will be followed wherever possible to improve metadata re-use in global repositories.
2. Local practices and schemes will only be used where no international options exist or where they can be mapped to international options.

2.3. Metadata transfer

The NRDS assumes that:

1. As a minimum requirement, all repositories will be capable of exporting their metadata using OAI-PMH with the unqualified Dublin Core metadata schema (oai_dc).
2. Some projects will also provide enhanced metadata using qualified Dublin Core.
3. UTF-8 encoding is required (because it is required by OAI-PMH).
4. Free text fields, such as Title and Abstract, may be translated into multiple languages. In these cases, the language of each field will be indicated in the XML encoding.
5. When multiple metadata values are supplied, the appropriate field will be repeated (as opposed to transferring multiple values in one field).

2.4. Importance of metadata fields

The levels of importance used in this document are:

- **Mandatory:** The field is required. If not supplied the record cannot be used by the NRDS.
- **Important:** The field offers significant functionality to users who are discovering a resource. It is highly recommended.
- **Medium:** The field offers some functionality to users for resource discovery, though the resource will still be discoverable if it is omitted. It is recommended.
- **Low:** The field has no special use in the NRDS, but will be imported and used in descriptions and re-exports. It is entirely optional.

3. Metadata guidelines for repositories

3.1. Identifier

DC definition (Identifier): An unambiguous reference to the resource within a given context.¹

DC definition (Bibliographic Citation): A bibliographic reference for the resource.

Each metadata record must provide a unique HTTP URL that will serve as the primary link for the resource, and will be used to refer searchers to the metadata record at the originating repository. Multiple identifiers, including multiple HTTP URLs, may also be provided (for example, if a document is stored in multiple formats or in several parts). The NRDS will select one URL to be the primary link (as described in Section 4).

Name	Importance in NRDS	QDC Encoding	Encoding Scheme	Format	Notes
URL	Mandatory	Identifier	URI	URL.	Each metadata record must at least include a unique HTTP URL that will be the Primary Link for the resource, used to refer users back to the source repository.
Handle ²	Low	Identifier	URI	Handle format (e.g. hdl:123.456).	Handle format identifiers may be converted to equivalent HTTP URLs.
Citation	Low	Identifier.BibliographicCitation	-	Free text citation or OpenURL syntax	NRDS will generate citations for display, but institutions may wish to override these with a preferred format. HTML should be formatted as for the Abstract field.
Catalogue Record	Low	Identifier	-	Free text.	May be institutional or national.
ISBN	Low	Identifier	URI	URN format (eg. urn:isbn:1234-5678).	
ISSN	Low	Identifier	URI	URN format (eg. urn:issn:1234-5678).	

¹ <http://dublincore.org/documents/dcmi-terms/> (Accessed 2007-02-12.)

² <http://www.handle.net/> (Accessed 2007-02-13.)

3.2. Title

DC definition (Title): A name given to the resource.

DC definition (Alternative): Any form of the title used as a substitute or alternative to the formal title of the resource.

Name	Importance in NRDS	QDC Encoding	Encoding Scheme	Format	Notes
Title	Mandatory	Title	-	Free text.	The full title of the document at item level, including any subtitle.
Alternative Title	Low	Title.Alternative	-	Free text.	Not currently used in NZ repositories, but may be in the future.

3.3. Author (Creator)

DC definition: An entity primarily responsible for making the resource.

Name	Importance in NRDS	QDC Encoding	Encoding Scheme	Format	Notes
Author	Mandatory (Repeat)	Creator	-	Surname, Firstnames/Initials.	Every document must have at least one author. This is the preferred format as it is easy to sort and provides more information. This format is currently used in most NZ institutions.
				As the name appears in the	This format is currently used at AUT.

				document.	
		Contributor.Author ³	-		DSpace may supply author metadata in this field.

3.4. Publisher

DC definition: An entity responsible for making the resource available.

This field stores both Institutional and “traditional” (external) publisher information.

Institution metadata captures the Institution (and Department) responsible for producing and distributing the research output. Each institution may add refinements identifying the faculty, school or department that published the document. The Publisher field is appropriate for Institution metadata when the institution has re-published the document by making it available through the institutional repository.

External publisher metadata applies to documents that have been formally published, for example in a journal or conference proceedings.

Name	Importance in NRDS	QDC Encoding	Encoding Scheme	Format	Notes
Institution	Important	Publisher	-	Institution.	The institution is used to promote the source repositories.
Institution and Department	Important	Publisher	-	Institution. Organisational Unit. Organisation Subunit.	The institution augmented with an arbitrary number of institutional subunits (Faculty, School, Department, etc) using a full stop as a delimiter. May also be captured in Thesis metadata (e.g. Thesis.Grantor).
External Publisher	Medium	Publisher	-	Free text.	Identifies an external publisher if one exists, such as a journal publisher or conference proceedings publisher.

³ Author is an agent MARC Relator for Contributor – see “MARC Relator terms and Dublin Core” <http://dublincore.org/usage/documents/relators/> (Accessed 2007-03-13)

The Institution is based on the originating repository. It is usually the institutional affiliation of the primary Author, but may not be. If a document has Authors at multiple institutions it may appear in more than one institutional repository. In this case, no other special action is required, though it may mean a resource is described more than once in the NRDS. It is recommended the names of all the institutions are included (in separate Publisher fields) in each repository.

3.5. Abstract (Description)

DC definition (Description): An account of the resource.

DC definition (Abstract): A summary of the content of the resource.

DC definition (TableOfContents): A list of subunits of the content of the resource.

Name	Importance in NRDS	QDC Encoding	Encoding Scheme	Format	Notes
Abstract	Important	Description	-	Free text.	<p>If no abstract is available, omit this field (do not use a placeholder like “No abstract”).</p> <p>Plain text is the preferred format. (A blank line can be used to indicate a paragraph break.)</p> <p>If formatting is required, HTML may be used. We recommend restricting to these elements (and attributes): P, A (HREF), I, BR. HTML entities (like &#257 for ā) may also be used.</p> <p>Some harvesters and some web browsers have trouble processing advanced HTML elements, HTML entities, and other markup (such as MathML). Caution and testing are advised to avoid display issues.</p>
		Description.Abstract	-		
Table of Contents	Low	Description.TableOf Contents	-	Free text	A textual separator should be used between items in the table in case the field is displayed as one continuous block of text.

3.6. Date

DC definition (Date): A point or period of time associated with an event in the lifecycle of the resource.

DC definition (DateAccepted): Date of acceptance of the resource (e.g. of thesis by university department, of article by journal, etc.).

DC definition (DateSubmitted): Date of submission of the resource (e.g. thesis, articles, etc.).

DC definition (DateIssued): Date of formal issuance (e.g., publication) of the resource.

DC definition (DateValid): Date (often a range) of validity of a resource.

The date field has many uses. The most important for the NRDS is the date that will be used in a citation (the date of publication or award).

Name	Importance in NRDS	QDC Encoding	Encoding Scheme	Format	Notes
Date	Low	Date	W3CDTF	W3CDTF profile of ISO 8601 ⁴ (eg: yyyy, yyyy-mm, yyyy-mm-dd, yyyy-mm-ddThh:mm+12:00, or yyyy-mm-ddThh:mmZ)	
Date accepted	Important	Date.DateAccepted	W3CDTF		
Date copyrighted	Low	Date.DateCopyrighted	W3CDTF		
Date created	Low	Date.Created	W3CDTF		
Date submitted	Low	Date.DateSubmitted	W3CDTF		
Date issued	Important	Date.Issued	W3CDTF		
Date modified	Low	Date.Modified	W3CDTF		
Date valid	Important	Date.Valid	W3CDTF		

⁴ <http://www.w3.org/TR/NOTE-datetime>

3.7. Subject

DC definition: The topic of the resource.

The Subject field contains controlled or uncontrolled descriptors. See Appendix A for a description of the possible vocabulary options.

The value of the Subject field as a discovery tool increases with the level of consistency among repositories. Highly consistency subject metadata is very useful for search and browse; inconsistent metadata is less useful. As many of the participants at the Planning Meeting indicated they will use subject metadata based on Marsden Fund “Fields of Research” Classification Codes”,⁵ these are recommended. If/when an Australasian version of these classification codes becomes available, this may replace the Marsden encoding scheme.

Where it is possible to separately identify temporal and spatial subjects, they should appear in Coverage.

Name	Importance in NRDS	QDC Encoding	Encoding Scheme	Format	Notes
Uncontrolled keywords	Medium	Subject	-	Free text	The author usually supplies uncontrolled keywords.
Marsden	Important	Subject	MarsdenClassification	Code followed by Descriptor text	Recommended. Repositories may use the full set of descriptors, or only the 23 top-level descriptors, or only the top two levels of descriptors. Both the code and descriptor should be transmitted even if the code is hidden in the user interface.
			MarsdenClassification	Descriptor text then code in brackets	
			MarsdenClassification	Descriptor text only	
			MarsdenClassification	Code only.	
MSH/NUH	Medium	Subject	TBD		New Zealand specific.
NZGPN	Medium	Subject	TBD		New Zealand specific.
PBRF Subject Areas	Low	Subject	TBD		Discouraged (subject to change).

⁵ <http://marsden.rsnz.org/downloads/2007Classification.pdf> (Accessed 2007-03-13)

Name	Importance in NRDS	QDC Encoding	Encoding Scheme	Format	Notes
ASRC	Low-Medium	Subject	TBD		Discouraged (Marsden is almost identical, so should be used instead).
OECD Research Codes			TBD		
LCC	Low-Medium	Subject	LCC	LCC.	Most of these schemes are useful for searching (assuming descriptor text is provided, not just a code like QA76.9). However, they are complex to assign, and if library resources are going to be dedicated to cataloguing to a one of these schemes, consider using Marsden instead for consistency. If authors are going to perform assignment, Marsden will be simpler.
LCSH		Subject	LCSH	LCSH.	
DDC		Subject	DDC	DDC.	
UDC		Subject	UDC	UDC.	
MESH		Subject	MESH	MESH.	
Digital Commons subject codes		Subject	TBD	Free text.	

3.8. Type

DC definition: The nature or genre of the resource.

See Appendix B for a description of the different metadata options.

For the purpose of the NRDS, the primary use of Type is to describe the type of research output (i.e. whether it is a journal article, conference proceeding, thesis, etc).

The Eprints Type Vocabulary (in conjunction with DCMI Type Vocabulary) is recommended because it is a standard, has well-defined terms, and is designed for use with scholarly publications. Institutions may use terms from the PBRF and DSpace Types instead; these will be crosswalked to the Eprints Type Vocabulary using equivalences outlined in Appendix B.

Name	Importance in NRDS	QDC Encoding	Encoding Scheme	Format	Notes
DCMI Type Vocabulary	Low (when used in isolation)	Type	DCMIType	Controlled vocabulary as text tokens (eg. StillImage)	12 terms. Very high level, most research outputs would appear simply as "Text". Too broad to be useful for a repository of research outputs on its own.
			URI	Controlled vocabulary as URI (eg. http://purl.org/dc/dcmitype/StillImage)	
Eprints Type Vocabulary	Important	Type	EPrintsType	Controlled vocabulary as text tokens (eg. ScholarlyText)	Recommended. 15 terms, all refinements of the DCMI Type "Text". Used with the 12 top-level DCMI terms to produce a set of 27 terms. Designed for describing scholarly publications.
			URI	Controlled vocabulary as URI (eg. http://purl.org/eprint/type/ScholarlyText)	
PBRF List of Research Output Types	Important (will be mapped to Eprints)	Type	TBD	Controlled vocabulary.	24 top-level terms, plus refinements of Conference Contribution. Lacks some outputs, such as Dataset, Image, Learning Object, Map, and Recording. Can be mapped to Eprints Types.
DSpace Default Type	Important (will be mapped to Eprints)	Type	TBD	Controlled vocabulary.	22 terms. Lacks terms for conference contributions. Can be mapped to Eprints Types.

3.9. Thesis

Thesis metadata is additional descriptive metadata that applies only to of theses and dissertations (the terms are used interchangeably). This field may include the type, level, name, and grantor of the thesis. Appendix C describes Thesis metadata options in more detail.

It is recommended that institutions maintain sufficient metadata that they can contribute to the Australasian Digital Theses Project.⁶ It is therefore recommended (in order of decreasing preference) that:

- Institutions that wish to supply detailed thesis metadata should use the ETD-MS format, and ensure that the Thesis.Degree.Level metadata uses the New Zealand Thesis Type vocabulary outlined in Appendix D.
- Institutions that wish to supply minimal thesis metadata should use the New Zealand Thesis Type vocabulary because it is easy to implement in either qualified or unqualified DC (using the Type field), provides a consistent way to distinguish thesis levels (the most important aspect of thesis metadata for the purpose of limiting search), and is sufficient for conversion and export to ADT.

Name	Importance in NRDS	QDC Encoding	Encoding Scheme	Format	Notes
ETD-MS	Important	Thesis.Degree.Name	-	Controlled vocabulary (institution-specific).	Recommended. The NZ Thesis Type vocabulary in Appendix D is recommended for Thesis.Degree.Level. The ETD-MS standard recommends the text “Electronic Thesis or Dissertation” in Type field but this is redundant and should be omitted.
		Thesis.Degree.Level	NZThesis Type	Controlled vocabulary as per Appendix D.	
		Thesis.Degree.Discipline	-	Free text; usually the name of a program or department.	
		Thesis.Degree.Grantor	TBD	Controlled list of institution names.	
NZ Thesis Level	Important	Type	NZThesis Type	Controlled vocabulary as per Appendix D.	Recommended
			-	Free text term matching description from Appendix D.	

⁶ <http://adt.caul.edu.au/> (Accessed 2007-02-13.)

Name	Importance in NRDS	QDC Encoding	Encoding Scheme	Format	Notes
ADT Thesis Type	Important	Type	ADTThesisType	Controlled vocabulary.	Optional. 3 Terms: Masters Thesis, PhD Doctorate, Professional Doctorate.

Institutions that are unable to supply thesis metadata may use Type metadata that can be crosswalked to the NZ Thesis Type terms. For example, the PBRF Output Type terms for theses are ‘Awarded Doctoral Thesis’ and ‘Awarded Research Masters Thesis’, which can be mapped to the NZ Thesis Type types “Doctoral” and “Research Masters” respectively.

3.10. Rights

DC definition (Rights): Information about rights held in and over the resource.

DC definition (AccessRights): Information about who can access the resource or an indication of its security status.

The Rights field has a relatively low priority for the NRDS because the discovery service deals only with the metadata records, not the resources themselves. However, researchers may use access rights information to evaluate the ease with which a resource can be obtained.

The EPrints AccessRights Vocabulary⁷ is recommended for this purpose because it is a standard, has well-defined terms, and is designed for use with scholarly publications. It is optional, and it is generally assumed that resources in the contributing repositories will be open-access. Repositories with “near-line” content (e. g. digitise-on-demand) may use “Restricted Access”.

Name	Importance in NRDS	QDC Encoding	Encoding Scheme	Format	Notes
Rights	Low	Rights	-	Plain text. Citation.	
			URI	URL. ISSN. ISBN.	
Access Rights	Medium	Rights.AccessRights	EPrintsAccessRights	Controlled vocabulary as text tokens (eg. OpenAccess)	Recommended.

⁷ http://www.ukoln.ac.uk/repositories/digirep/index/Eprints_AccessRights_Vocabulary_Encoding_Scheme (Accessed 2007-03-13)

Name	Importance in NRDS	QDC Encoding	Encoding Scheme	Format	Notes
			EPrintsAccessRights	Controlled vocabulary as URI (eg. http://purl.org/eprint/accessRights/OpenAccess)	This field is optional, and by default resources are assumed to be OpenAccess. If a resource is not open access, it is recommended a short, plain-text statement is included describing the rights restrictions.
			-	Free text.	A short, plain-text statement explaining access restrictions of Restricted Access and Closed Access resources.

3.11. Contributor

DC definition: An entity responsible for making contributions to the resource.

A contributor to the document who is not the Author. See the Author guidelines for use of Contributor.Author.

Name	Importance in NRDS	QDC Encoding	Encoding Scheme	Format	Notes
Contributor	Low	Contributor	-	As for Author.	
Editor	Low	Contributor	-		
Advisor	Low	Contributor	-		

3.12. Coverage

DC definition (Coverage): The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant.

DC definition (Spatial): Spatial characteristics of the intellectual content of the resource.

DC definition (Temporal): Temporal characteristics of the intellectual content of the resource.

Name	Importance in NRDS	QDC Encoding	Encoding Scheme	Format	Notes
Coverage	Low	Coverage	-	Free text.	
Coverage	Low	Coverage.Spatial	-	Free text.	
Coverage	Low	Coverage.Temporal	-	Free text.	

3.13. Language

DC definition: A language of the resource.

Name	Importance in NRDS	QDC Encoding	Encoding Scheme	Format	Notes
Language	Low	Language	RFC3066	RFC 3066 ⁸ codes (language code plus optional hyphen and country code).	RFC 3066 is recommended as it supersedes the other two standards.
		Language	ISO639-2	ISO 639-2 ⁹ codes (two or three letters).	
		Language	RFC1766	RFC 1766 ¹⁰ codes.	
		Language	-	Text descriptions (e.g. "French").	

⁸ <http://www.ietf.org/rfc/rfc3066.txt> (Accessed 2007-02-13.)

⁹ <http://lcweb.loc.gov/standards/iso639-2/langhome.html> (Accessed 2007-02-143.)

¹⁰ <http://www.ietf.org/rfc/rfc1766.txt> (Accessed 2007-02-13.)

3.14. Relation

DC definition: A related resource.

Name	Importance in NRDS	QDC Encoding	Encoding Scheme	Format	Notes
Relation	Low	Relation	URI	URL. Citation. ISSN. or ISBN.	
		Relation.hasFormat			
		Relation.hasPart			
		Relation.hasVersion			
		Relation.isFormatOf			
		Relation.isPartOf	URI	URN format (eg. urn:issn:1234-5678, urn:isbn:1234-5678)	The Journal (or Book) that an article (or chapter) is published in.
		Relation.isReferencedBy		Free text.	The bibliographic Identifier of the corresponding Catalogue record.
		Relation.isReplacedBy			
		Relation.isRequiredBy			
		Relation.isVersionOf	URI	URL.	Working papers and preprints may use to reference published equivalents.
		Relation.References			
		Relation.Replaces			
		Relation.Requires			

3.15. Format

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

Name	Importance in NRDS	QDC Encoding	Encoding Scheme	Format	Notes
Format	Low	Format	IMT	Free text.	Not directly used because NRDS does not harvest resources, only metadata.

3.16. Source

DC definition: The resource from which the described resource is derived.

Name	Importance in NRDS	QDC Encoding	Encoding Scheme	Format	Notes
Source	Low	Source	URI	URL. Citation. ISSN. or ISBN.	

3.17. Derived metadata used in the discovery service

NRDS will transform parts of the supplied metadata to improve its findability and usability, however it will not augment the metadata with additional information.

The following metadata fields will be derived by the NRDS from the metadata provided by the repositories.

Name	Importance in NRDS	Source metadata field	Format	Notes
Primary Link	Mandatory (Single)	Identifier	HTTP URL.	<p>The Primary Link is the HTTP URL of the metadata record in the repository, and is used to refer searchers back to the originating repository.</p> <p>If a single HTTP URL is supplied, it is used as the primary Link.</p> <p>If multiple HTTP URLs are used, then the primary is selected in this order:</p> <ul style="list-style-type: none"> • The handle.net URL. • The URL at the source repository (shortest first). • The URL ending in .htm or .html (shortest first). • The shortest URL
Clean Title	Mandatory (Single)	Title	Free text.	<p>The title used to browse by Title.</p> <p>HTML will be stripped as for Clean Abstract.</p> <p>If multiple titles are available, the longest will be selected.</p>
Author	Mandatory	Creator	Surname, Firstnames/Initials.	The name used to browse by Author.

Name	Importance in NRDS	Source metadata field	Format	Notes
	(Repeated)	Contributor.Author	As the name appears in the document.	The system will attempt to transform names from “appears in the document” format into “surname, firstnames” format.
Clean Abstract	Mandatory (Single)	Abstract	Plain text.	A plain-text description of the resource used in search result display. HTML other than these elements (and attributes) will be stripped: P, A (HREF), I, BR. HTML entities (like ā for ā) will be retained. MathML markup will be ignored. The field will be truncated to 200 chars.
Institution	Mandatory (Single)	Publisher	Institution.	The Institution is used to provide a link to the source institution. If the Institution is not supplied in the metadata record, the NRDS will assign an institution based on the originating repository. The institution name will be placed in AC.Rights ¹¹
Institutional Structure	Optional (Single)	Publisher	Institution. Organisational Unit. Organisation Subunit.	If the institution is augmented with an arbitrary number of institutional subunits (Faculty, School, Department, etc) then NRDS will allow users to browse the organisation structure.

¹¹ <http://metadata.net/admin/draft-iannella-admin-01.txt> (Access 2007-03-13)

Name	Importance in NRDS	Source metadata field	Format	Notes
Primary date	Mandatory (Single)	Date	YYYY	<p>The date used to browse by Date, to limit searches, and (possibly) in citations..</p> <p>It will be derived from the available Date metadata as follows:</p> <p>For unqualified DC, the earliest date will be chosen.</p> <p>For qualified DC, a date will be selected in this order:</p> <ul style="list-style-type: none"> • Date issued • Date accepted • Date valid • The earliest date field • Metadata record creation date
Marsden Field of Research	High (Repeated)	Subject	Marsden code	<p>The Subject used to browse by Subject, to limit searches, and to limit notifications.</p> <p>Marsden Subject codes will be verified and copied to this field.</p> <p>PBRF Subject terms will be crosswalked to Marsden codes then added to this field.</p> <p>Unqualified Subject metadata will be matched against Marsden terms, and matches added. Unmatched terms will be matched against PBRF codes and crosswalked to this field.</p>

Name	Importance in NRDS	Source metadata field	Format	Notes
Eprints Type Vocabulary	High (Single)	Type	Controlled vocabulary	The Type used to limit searches, limit notifications. The combined Eprints and DCMI terms be used. Eprints and DCMI Type terms will be added to this field.
Thesis Type	High (Single)	Thesis.Degree.Level Type	Controlled vocabulary	The Thesis Level used to limit searches. The NZ Thesis Type value is repeated in the Type field for simple DC exporting (Thesis.Degree is excluded from simple DC).
Citation	High (Single)	Identifier (BibliographicCitation)	Free text	The Citation is used to summarise a resource to a user. HTML elements will be stripped as for Abstract. If no Citation is provided, a citation will be constructed based on other fields.
Language	High (Single)	Language	RFC 3066 two-letter codes.	The Language used to limit searches. Defaults to 'en' if not supplied. Crosswalked from (first) Language metadata field.
Eprints AccessRights	Medium	Rights (AccessRights)	Eprints AccessRights controlled vocabulary as text or URI.	Defaults to 'OpenAccess' if not supplied.

Appendix A: Controlled subject metadata options.

Scheme	Maintainer	Assigned by	Size	Examples	Used by	Notes
Australian Standard Research Classification (ASRC)	Australian Research Council	Author / Librarian	~1,000	300202 Plant Nutrition		Based on OECD Research Codes Schema supplied n Fedora /DSpace (?).
Marsden Fund Fields of Research Classification Codes ¹² (Marsden)	The Marsden Fund / Royal Society of New Zealand	Author / Librarian	~1,000	27000 Biological Sciences 300202 Plant Nutrition	Coda CPIT/OARiNZ Massey UoA UoC (?) VUW (?)	Based on OECD Research Codes. Repositories may use the full set of descriptors, or only the 23 top-level descriptors, or only the top two levels of descriptors. Weak in social sciences. Coda will use top-level. CPIT/OARiNZ using top-level Marsden (“actually ASRC but it is the same thing at the OECD level”). Massey building DSpace schema. UoA providing as “Field of Research”, probably limited to top 23 terms. Author-supplied, library will add if missing.
Performance-Based Research Fund (PBRF) Subject Areas ¹³	TEC	Researcher	42	Accounting and finance Visual arts and crafts	VUW (?)	Consider splitting “compound” descriptors (e.g. “Accounting and Finance” becomes “Accounting” and also “Finance”).
Maori Subject Headings Nga Upoko Tukutuku (MSH/NUT)	LIANZA, Te Te Rōpū Whakahaui, National Library	Librarian / Cataloguer	~1,000	Pakitara Pakiwaitara Pāngarau		Note frequent use of macrons. Explicitly “appropriate for use at a public library level”.
New Zealand Geographic Place Names (NZGPN)	Land Information New Zealand		~40,000			

¹² <http://marsden.rsnz.org/downloads/2007Classification.pdf>

¹³ http://www.tec.govt.nz/downloads/a2z_publications/pbrf2006-guidelines.pdf (page 65)

Scheme	Maintainer	Assigned by	Size	Examples	Used by	Notes
Library of Congress Subject Headings (LCSH)	Library of Congress	Librarian / Cataloguer	200,000 +	Zoology Forestry and community		Used by University of Auckland in the Library Catalogue but not in ResearchSpace.
Library of Congress Classifications (LCC)	Library of Congress	Librarian / Cataloguer	100,000 +	HF5601 Accounting		Used by Otago Business School.
Dewey Decimal Classification (DDC)	Online Computer Library Center	Librarian / Cataloguer	?	330.94 European economy		
Universal Dewey Classification (UDC)	Universal Decimal Classification Consortium	Librarian / Cataloguer	?	59+636 Zoology and animal breeding		
Medical Subject Headings (MESH)	US National Library of Medicine	Librarian / Cataloguer	?	C06.301 Digestive System Neoplasms		
OECD Research Codes ¹⁴ / CRS Purpose Codes	Organisation for Economic Co-operation and Development	Author / Librarian	~1,000	32172 transport equipment industry		ASRC and Marsden are apparently both based on these codes.
Digital Commons Subject Codes	ProQuest (?)			HEALTH SCIENCES, MEDICINE AND SURGERY (0564) HEALTH SCIENCES, HEALTH CARE MANAGEMENT (0769)		Used by Digital Commons theses collections based on Digital Dissertations Abstract Classification Scheme. UoA imported from Digital Commons but will not continue in ResearchSpace@Auckland

¹⁴ http://www.oecd.org/document/21/0,2340,en_2825_495602_1914325_1_1_1_1,00.html

Appendix B: Type metadata options.

Overview

Scheme	Maintainer	Assigned by	Size	Examples	Used by	Notes
DCMI Type Vocabulary ¹⁵	Dublin Core Metadata Initiative	Author	12	Text Dataset Image	VUW	12 terms: Collection, Dataset, Event, Image, InteractiveResource, MovingImage, PhysicalObject, Service, Software, Sound, StillImage, Text.
Eprints Type Vocabulary Encoding Scheme ¹⁶	JISC Digital Repositories Programme	Author	15	Book Conference Item Conference Paper		<p>15 terms: Scholarly Text, Book, Book Item, Book Review, Conference Item, Conference Paper, Conference Poster, Journal Item, Journal Article, News Item, Patent, Report, Submitted Journal Article, Thesis or Dissertation, Working or Discussion Paper.</p> <p>Formal definitions are supplied.</p> <p>The list is a refinement of the DCMI term "Text", and is used with the DCMI Types list to produce a combined list of 27 descriptors.</p> <p>The Eprints authors are also members of the DCMI, so there is the potential parts of it may be promoted for adoption by DCMI. Also, the Eprints project is aiming to embed it in future releases of IR applications (DSpace, Eprints, Fedora, Digital Commons) in the same way that DC is native to these applications currently.</p>

¹⁵ <http://dublincore.org/documents/dcmi-type-vocabulary/> (Accessed 2007-02-12.)

¹⁶ http://www.ukoln.ac.uk/repositories/digirep/index/Eprints_Type_Vocabulary_Encoding_Scheme (Accessed 2007-02-13)

Scheme	Maintainer	Assigned by	Size	Examples	Used by	Notes
PBRF List of Research Output Types ¹⁷	TEC	Author	24+	Authored Book Journal Article	UoC VUW (?)	<p>24 top-level terms, plus refinements of Conference Contribution: Artefact/Object/Craftwork; Authored Book; Awarded Doctoral Thesis; Awarded Research Masters Thesis; Chapter in Book; Commissioned Report for External Body; Composition; Conference Contribution (abstract - full conference paper - conference paper in published proceedings - poster presentation - oral presentation – other); Confidential Report for External Body; Discussion Paper; Design Output; Edited Book; Exhibition; Film/Video; Intellectual Property (eg. patent, trademark); Journal Article; Monograph; Oral Presentation; Performance; Scholarly Edition; Software; Technical Report; Working Paper; Other Form of Assessable Output.</p> <p>Likely to change every few years when PBRF is revised.</p> <p>Lacks some important categories that are included in DSpace, such as Dataset, Image, Learning Object, Map, and Recording.</p> <p>See mapping to DCMI and Eprints terms below.</p>
DSpace Default Types	DSpace Project	Author	22	Book chapter Dataset Learning Object	UoA	<p>22 Terms: Animation; Article; Book; Book chapter; Dataset; Learning Object; Image; Image, 3-D; Map; Musical Score; Plan or blueprint; Preprint; Presentation; Recording, acoustical; Recording, musical; Recording, oral; Software; Technical Report; Thesis; Video; Working Paper; Other.</p> <p>Not limited to research output types.</p> <p>See mapping to DCMI and Eprints terms below.</p>

Comparison of Eprints, DSpace and PBRF Terms

The following table illustrates the relationship between the DCMI, Eprints, PBRF and DSpace types. It should be read as follows:

- *High-level DCMI Type*: This column contains the DCMI Type fields.
- *Refinements from Eprints Type Vocabulary*: This column contains the Eprints Type Vocabulary terms that refine the DCMI Type “Text” descriptor. The ► symbol is used to show subclass relationships within the Eprints Type vocabulary.
- *Equivalent PBRF Terms*: This column arranges the PBRF Terms against the single matching term from DCMI terms and Eprints. This column can be used to crosswalk from PBRF to DCMI/Eprints terms (but not vice-versa).

¹⁷ http://www.tec.govt.nz/downloads/a2z_publications/pbrf2006-guidelines.pdf (page 43).

- *Equivalent DSpace Terms*: This column arranges the DSpace Default Terms against the single matching term from DCMI and Eprints. This column can be used to crosswalk from DSpace terms to DCMI/Eprints terms (but not vice-versa).

High-level DCMI Type	Refinements from Eprints Type Vocabulary	Equivalent PBRF Terms	Equivalent DSpace Terms
Text	Scholarly Text		
	▶ Book	Authored Book Edited Book Monograph Scholarly Edition	Book
	▶ Book Item	Chapter in Book	Book chapter
	▶ Book Review		
	▶ Conference Item	Conference Contribution Conference Contribution – abstract Conference Contribution – oral presentation Conference Contribution – other	Presentation
	▶▶ Conference Paper	Conference Contribution – full conference paper Conference Contribution – conference paper in published proceedings	
	▶▶ Conference Poster	Conference Contribution – poster presentation	
	▶ Journal Item		
	▶▶ Journal Article	Journal Article	Article
	▶ News Item		
	▶ Patent	Intellectual Property (eg. patent, trademark)	
	▶ Report	Commissioned Report for External Body Confidential Report for External Body Technical Report	Technical Report
	▶ Submitted Journal Article		Preprint
	▶ Thesis or Dissertation	Awarded Doctoral Thesis Awarded Research Masters Thesis	Thesis

High-level DCMI Type	Refinements from Eprints Type Vocabulary	Equivalent PBRF Terms	Equivalent DSpace Terms
	► Working or Discussion Paper	Discussion Paper Working Paper	Working Paper
Collection			
Dataset			Dataset
Event		Exhibition Oral Presentation Performance	
Image			
InteractiveResource			
MovingImage		Film/Video	Animation Video
PhysicalObject		Artefact/Object/Craftwork	
Service			
Software		Software	Software
Sound			Recording, acoustical Recording, musical Recording, oral
StillImage			Image Image, 3-D Map Musical Score Plan or blueprint
Unclassified / Unclassifiable		Composition Design Output Other Form of Assessable Output	Learning Object Other

Appendix C: Thesis metadata options

Scheme	Maintainer	Assigned by	Size	Examples	Currently used by	Notes
Australian Digital Thesis Type ¹⁸ (ADT Thesis Type)	ADT Project	Library	3	Masters Thesis PhD Doctorate Professional Doctorate		3 Terms: Masters Thesis, PhD Doctorate, Professional Doctorate. Note that undergraduate theses are not included in ADT. Suggests a thesis.degree field. Not clear if this is in use or proposed.
Robert Gordon University Electronic Theses ¹⁹ (RGU-ETD Type)	Robert Gordon University	Library		Thesis. Doctoral. Ph D.		Specifies DC Type field. Format is "Type. Qualification level. Qualification name." where Type = Thesis or dissertation (default on system); Level = Diploma, Masters, Doctoral, Postdoctoral, etc (controlled look up list); Name = Specific degree (MPhil ,PhD, DPhil etc). (vocabulary controlled by individual institution, separated by full stops.)
Interoperability Metadata Standard for Electronic Theses and Dissertations ²⁰ (ETD-MS).	Networked Digital Library of Theses and Dissertations (NDLTD)	Library		<degree> <name>PHD</name> <level>doctoral</level> <discipline>Forestry</discipline> <grantor>Virginia Polytechnic Institute and State University</grantor> </degree>	UoA UoC	Well-documented, including MARC crosswalk. Recommends the text "Electronic Thesis or Dissertation" and also DCMI Type in Type field. Adds Elements for Thesis.Degree.Name, Thesis.Degree.Level, Thesis.Degree.Discipline, and Thesis.Degree.Grantor.

¹⁸ <http://adt.caul.edu.au/standards/metadata/> (Accessed 2007-02-13.)

¹⁹ <http://www2.rgu.ac.uk/library/guidelines/metadata.html>

²⁰ <http://www.ndltd.org/standards/metadata/current.html>

Appendix D: New Zealand Thesis and Dissertation Type Vocabulary

Institutions that supply thesis metadata should use the vocabulary below to indicate the academic level of a thesis or dissertation

The ► symbol indicates a narrower term. The abbreviation UF (“used for”) indicates a term that may be used, but for which an equivalent preferred term exists and is recommended. For example, the term “PhD” may be used, but the equivalent term “PhD Doctorate” is preferred.

Term	Token	Description	Currently Used by
Post-doctoral	Postdoctoral		VUW
Doctoral	Doctoral	Work submitted towards a Doctorate or equivalent degree.	UoC, UoA, VUW
UF: Awarded Doctoral Thesis			PBRF, coda
► PhD Doctorate	PhdDoctorate		ADT
UF: PhD			UoO
► Professional Doctorate	ProfessionalDoctorate		ADT
Masters	Masters	Work submitted towards a Masters or equivalent post-graduate degree.	UoC, UoA, VUW
UF: Masters Thesis			ADT
► Research Masters	ResearchMasters		UoO
UF: Awarded Research Masters Thesis			PBRF, coda
► Coursework Masters	CourseworkMasters		UoO
Undergraduate	Undergraduate	Work submitted towards an undergraduate degree.	UoA
► Bachelors	Bachelors		VUW
► Honours	Honours		UoO

NRDS will recognise all the tokens and terms above. When parsing terms, it will strip punctuation and (i.e. “Master’s” will be recognised as “Masters”) and allow for variation in case (i.e. “masters” will be recognised as “Masters”).