

Additional Metadata Guidelines for Contributors to NZResearch.org.nz

To support Metadata best practice requirements for New Zealand Repositories

March 2012



Document Control

Revision history

Revision	Date	Author	Reason for Change
1.0	July 2011	Compiled by Leonie Hayes following guidance from the NZ Repository Managers	Additional Metadata Guidelines for NZ Repositories are optional and not required for the SSI service. Therefore they can be wider in scope but still inform best practice for NZ repositories. The original Metadata Guidelines have been split into separate documents.
1.1	March 2012	Emerson Vandy	Edits to reflect transition to production service delivery on Shared Search Infrastructure (SSI).

Table of Contents

1. INTRODUCTION.....	4
1.1. PURPOSE OF THIS DOCUMENT.....	4
2. ADDITIONAL METADATA GUIDELINES FOR NZ REPOSITORIES.....	5
2.1. IDENTIFIER.....	5
2.2. TITLE.....	5
2.3. PUBLISHER.....	6
2.4. OTHER DESCRIPTION FIELDS.....	7
2.5. DATE.....	7
2.6. SUBJECT.....	8
2.6.1. <i>Controlled subject metadata information.....</i>	<i>10</i>
2.7. TYPE.....	11
2.8. CONTRIBUTOR.....	12
2.9. LANGUAGE.....	12
2.10. FORMAT.....	13

1. Introduction

This support document is based on earlier iterations of the guidelines used at New Zealand research institutions in their institutional repositories.

These support metadata guidelines are optional and are not comprehensive. They **may** be used to **help** inform metadata practices.

1.1. Purpose of this document

The purpose of the Additional Metadata Guidelines for NZ Repositories is to provide additional assistance to New Zealand repository managers beyond the recommended Metadata required for the Shared Search Infrastructure (SSI) service. The documents referenced are:

- Metadata Guidelines for Contributors to NZResearch.org.nz
- SSI Aggregator Guidelines for Harvesting

2. Additional Metadata Guidelines for NZ Repositories

These guidelines are a standalone document to supplement NZ Repository metadata requirements. These metadata guidelines are optional and are not comprehensive. They may be used to help inform metadata practices.

2.1. Identifier

Name	QDC Encoding	Encoding Scheme	Format	Notes
Catalogue Record	Identifier	-	Free text.	May be institutional or national.
ISBN	Identifier	URI	URN format (eg. urn:isbn: 9786543212345).	
ISSN	Identifier	URI	URN format (eg. urn:issn:1234- 5678).	
DOI	Identifier		10.1234/5969695	No prefix

2.2. Title

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

Name	QDC Encoding	Encoding Scheme	Format	Notes
Alternative Title	Title.Alternative	-	Free text.	Use is optional

2.3. 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	QDC Encoding	Encoding Scheme	Format	Notes
Institution	Publisher	-	Institution.	Only use the publisher field if the institution is the actual publisher (ie for unpublished work that the Institution can claim to be distributing the work)
Institution and Department	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). See above comment regarding unpublished work
External Publisher	Publisher	-	Free text.	Identifies an external publisher if one exists, such as a journal publisher or conference proceedings publisher.

2.4. Other Description fields

Name	QDC Encoding	Encoding Scheme	Format	Notes
Table of Contents	Description.TableOf Contents	-	Free text	A list of subunits of the content of the resource. A textual separator should be used between items in the table in case the field is displayed as one continuous block of text.
Version	Description.version		Use the NISO Standard Version of Record Accepted Manuscript	http://www.niso.org/publications/rp/RP-8-2008.pdf

2.5. Date

Name	QDC Encoding	Encoding Scheme	Format	Notes
Date accepted	Date.DateAccepted	W3CDTF		Date of acceptance of the resource (e.g. of thesis by university department, of article by journal, etc.).
Date copyrighted	Date.DateCopyrighted	W3CDTF		Useful if there is variance between publication date and release date
Date created	Date.Created	W3CDTF		Date (often a range) of validity of a resource.
Date submitted	Date.DateSubmitted	W3CDTF		Date of submission of the resource (e.g. thesis, articles, etc.).
Date modified	Date.Modified	W3CDTF		
Date valid	Date.Valid	W3CDTF	W3CDTF	W3CDTF

2.6. Subject

DC definition: The topic of the resource.

The Subject field contains controlled or uncontrolled descriptors. 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. NZ Repositories are encouraged to use ANZSRC “Fields of Research” Classification Codes” for subject metadata¹, these are now recommended as best practice. Where it is possible to separately identify temporal and spatial subjects they should appear in Coverage.

Name	QDC Encoding	Encoding Scheme	Format	Notes
ANZSRC	Subject	ANZSRC Classification	Code followed by Descriptor text	Recommended. Repositories may use the full set of descriptors, or only the 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. Example dc.subject.anzsrc:01-Mathematical Sciences dc.subject.anzsrc:0911-Maritime Engineering dc.subject.anzsrc:100510-Wireless Communications
Uncontrolled keywords	Subject	-	Free text	The author usually supplies uncontrolled keywords.
MSH/NUH	Subject	TBD		New Zealand specific.
NZGPN	Subject	TBD		New Zealand specific.

¹ <http://www.arc.gov.au/era/ANZSRC.htm>

Name	QDC Encoding	Encoding Scheme	Format	Notes
LCC	Subject	LCC	LCC.	<p>Most of these schemes are useful for searching (assuming descriptor text is provided, rather than a LC classification code like QA76.9).</p> <p>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 ANZSRC instead for consistency. If authors are going to perform assignment, ANZSRC will be simpler.</p>
LCSH	Subject	LCSH	LCSH.	
DDC	Subject	DDC	DDC.	
UDC	Subject	UDC	UDC.	
MESH	Subject	MESH	MESH.	
ASRC	Subject			ANZSRC Codes supersede ASRC.
Digital Commons subject codes	Subject			No KRIS contributors use the Digital Commons platform anymore.
OECD Research Codes	Subject			ANZSRC Codes should be used in preference to OECD.
PBRF Subject Areas	Subject			Discouraged unless agreed otherwise in late 2010 when PBRF Guidelines for 2012 are issued.
Marsden	Subject			Marsden Codes have been replaced by ANZSRC Codes and should no longer be used.

2.6.1. Controlled subject metadata information.

This lists some options for assigning subject metadata and can be used as a guide.

Scheme	Maintainer	Size	Examples	Notes
Australian Standard Research Classification (ASRC)	Australian Research Council	~1,000	300202 Plant Nutrition	Based on OECD Research Codes Controlled vocabulary XML schemas for ANZSRC subject code allocation is available for most repository software.
Australia New Zealand Standard Research Codes (ANZSRC)	Australian Research Council	~1200	020603 Quantum Information, Computation and Communication	Based on OECD Research Codes. Repositories may use the full set of descriptors, or only the top-level descriptors, or only the top two levels of descriptors. http://www.arc.gov.au/era/ANZSRC.htm
Marsden Fund Fields of Research Classification Codes (Marsden)	The Marsden Fund / Royal Society of New Zealand	~1,000	27000 Biological Sciences 300202 Plant Nutrition	Based on OECD Research Codes. Superseded by ANZSRC. http://marsden.rsnz.org/downloads/2007Classification.pdf
Performance-Based Research Fund (PBRF) Subject Areas	TEC	42	Accounting and finance Visual arts and crafts	Consider splitting “compound” descriptors (e.g. “Accounting and Finance” becomes “Accounting” and also “Finance”). http://www.tec.govt.nz/Documents/Publications/PBRF-Quality-Evaluation-Guidelines-2012.pdf (page 73).
Maori Subject Headings Nga Upoko Tukutuku (MSH/NUT)	LIANZA, Te Te Rōpū Whakahau, National Library	~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		

Scheme	Maintainer	Size	Examples	Notes
Library of Congress Subject Headings (LCSH)	Library of Congress	200,000+	Zoology Forestry and community	Used by University of Auckland in the Library Catalogue but not in ResearchSpace.
Dewey Decimal Classification (DDC)	Online Computer Library Center		330.94 European economy	
Universal Dewey Classification (UDC)	Universal Decimal Classification Consortium		59+636 Zoology and animal breeding	
Medical Subject Headings (MESH)	US National Library of Medicine		C06.301 Digestive System Neoplasms	
OECD Research Codes / CRS Purpose Codes	Organisation for Economic Co-operation and Development	~1,000	32172 transport equipment industry	ASRC and Marsden are apparently both based on these codes. http://www.oecd.org/document/21/0,2340,en_2825_495602_1914325_1_1_1_1.00.html

2.7. Type

Name	QDC Encoding	Encoding Scheme	Format	Notes
DCMI Type Vocabulary	Type	DCMIType	Controlled vocabulary as text tokens (eg. StillImage) Controlled vocabulary as URI (eg. http://purl.org/dc/dcmitype/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.

Name	QDC Encoding	Encoding Scheme	Format	Notes
PBRF List of Research Output Types	Type		Controlled vocabulary.	25 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	Type		Controlled vocabulary.	22 terms. Lacks terms for conference contributions. Can be mapped to Eprints Types.

2.8. Contributor

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

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

Name	QDC Encoding	Encoding Scheme	Format	Notes
Contributor	Contributor	-	As for Author.	
Editor	Contributor	-		
Advisor	Contributor	-		Use for Thesis Supervisor

2.9. Language

DC definition: A language of the resource.

Name	QDC Encoding	Encoding Scheme	Format	Notes
------	--------------	-----------------	--------	-------

Name	QDC Encoding	Encoding Scheme	Format	Notes
Language	Language	RFC3066	RFC 3066 ² 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 ³ codes (two or three letters).	
	Language	RFC1766	RFC 1766 ⁴ codes.	
	Language	-	Text descriptions (e.g. "French").	

2.10. Format

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

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

² <http://www.ietf.org/rfc/rfc3066.txt> (Accessed 2007-02-13.)

³ <http://lcweb.loc.gov/standards/iso639-2/langhome.html> (Accessed 2007-02-143.)

⁴ <http://www.ietf.org/rfc/rfc1766.txt> (Accessed 2007-02-13.)