

NDL Metadata Schema

Manual: Version 2.0

**Development of National Digital Library of India: Towards
Building a National Asset**
(A NMEICT Project, MHRD, Govt. of India)



Prepared by NDL Team
2015



Central Library
(ISO 9001:2008 Certified)
Indian Institute of Technology Kharagpur
Kharagpur 721302

Preface

The concept of developing the National Digital Library (NDL) was conceived during the discussion with Prof. P.P. Chakrabarti, Director, IIT Kharagpur and Hon'ble Minister of MHRD, Govt. of India in the month of June, 2014 and subsequent discussion with Dr. B. Sutradhar, Librarian, Central Library, Prof. P.P. Das, Department of Computer Science and Engineering and Prof. P.P. Chakrabarti, Director, IIT Kharagpur, a proposal was made and submitted to Ministry of Human Resource Development (MHRD). The objective of the pilot project of National Digital Library (NDL) is to integrate all the existing digitized contents across educational institutions of the nation to provide single window access to the different group of users of our country. With the objective, several discussions and meeting were held for designing a standards Metadata schema for different types of learning e-content to be used for NDL project. The Metadata summarizes basic information about data, which can make finding and working with particular instances of data easier. For example, author, date created and date modified and file size are examples of very basic document metadata. Having the ability to filter through that metadata makes it much easier for someone to locate a specific document. Based on the discussion and meeting, a preliminary draft for NDL Metadata Schema (ver 1.0) was prepared by Dr. P.K. Bhowmick based on Dublin Core (DC) and IEEE LOM by incorporating the inputs from Prof. P. P. Das, Prof. Sudeshna Sarkar, Dr. P. S. Mukhopadhyay, Dr. P.K. Bhowmick and Dr. B. Sutradhar.

Dr. P. K. Bhowmick had presented NDL metadata (ver 1.0) schema in the meeting held at IIT Delhi on 23 April, 2015, where Prof. P.P. Das, Joint. PI of the NDL Project, Dr. Jagdish Arora, Director, INFLIBNET, Prof. B.D. Gupta, National Coordinator of INDEST-AICTE Consortium, Prof. Uma Kanjila, IGNOU, New Delhi, Dr. B. Sutradhar Co-PI of NDL Project, Dr. S.K. Jalal and Mr. S.G. Roy from Central Library, IIT Kharagpur and Mr. S. Banerjee, CSE, IIT Kharagpur were present.

In the brainstorming session held at IIT Kharagpur during May 21-22, 2015, Prof. P.P. Chakrabarti, Director, IIT Kharagpur, in his welcome address emphasized the mission and vision of NDL to all the members present in the meeting and also requested to bring out **Manual for NDL Metadata Schema (ver-2.0)** to accommodate different types of digital contents. Based on the discussions and feedbacks from the expert members like Prof. Uma

Kanjilal and Dr. P.S. Mukhopadhyay, Dr. Jagdish Arora, Prof. P.P. Das, Prof. A. Basu, Prof. Sudeshna Sarkar, Dr. P.K. Bhowmick, Dr. B. Sutradhar, Dr. Sandip Chakraborty, Mr. Yatrik Patel, Mr. Abhisekh Kumar and Dr. S. K. Jalal, Metadata Manual (ver. 2.0) was drafted incorporating all important elements from metadata standards like Dublin Core, IEEE LOM, LRMI Standards and MPEG 7 into NDL. The drafted version 2.0 of NDL Metadata Schema has been verified thoroughly by Prof. Uma Kanjilal and Dr. P.S. Mukhopadhyay. Prof. Uma Kanjilal has specially contributed the part of Educational Metadata based on LRMI and MPEG 7 for multimedia metadata.

The NDL Metadata Manual (ver 2.0) consists of seven chapters. **Chapter 1** deals with the general concepts which are often used in the manual. **Chapter 2** discusses about the Generic Metadata elements which are essential to describe a digital document. **Chapter 3** deals with Educational Metadata elements based on LRMI schema. **Chapter 4** talks about multimedia metadata elements through MPEG 7 standard to describe objects like audio and video documents. **Chapter 5** deals with metadata based on Shodhganga for Electronic Theses and Dissertation (ETD). **Chapter 6** of the manual deals with few examples of various types of digital documents. **Chapter 7** deals with Vocabulary use especially for the case of educational metadata and some generic metadata.

Acknowledgements

Prof. Partha Pratim Chakrabarti, Director, IIT Kharagpur
Prof. Partha Pratim Das, Professor, CSE, IIT Kharagpur
Prof. Subrata Chattopadhyay, Chairman, C.L & Professor, A & RP, IIT Kharagpur
Prof. Anupam Basu, Professor, CSE, IIT Kharagpur
Prof. Sudeshna Sarkar, Professor, CSE, IIT Kharagpur
Dr. Plaban Kumar Bhowmick, Asst. Professor, CET, IIT Kharagpur
Dr. Jagdish Arora, Director, INFLIBNET, Gujarat
Prof. Uma Kanjilal, Professor, DLIS, IGNOU, New Delhi
Dr. Parthasarathi Mukhopadhyay, Asso. Professor, DLIS, University of Kalyani
Dr. Sandip Chakraborty, Asst. Professor, CSE, IIT Kharagpur
Mr. Mainak Ghosh, Asst. Professor, A & RP, IIT Kharagpur
Dr. B. Sutradhar, Librarian, Central Library, IIT Kharagpur
Dr. Samir Kumar Jalal, Deputy Librarian, Central Library, IIT Kharagpur
Mr. Nanda Gopal Chattopadhyay, CTO, NDL Project, IIT Kharagpur
Mr. Yatrik R. Patel, Scientist D (CS), INFLIBNET Centre
Mr. Abhisekh Kumar, Scientist C (CS), INFLIBNET Centre
Mr. Samrat Guha Roy, Asst. Librarian, Central Library, IIT Kharagpur
Mr. Shibabroto Banerjee, CSE, IIT Kharagpur
Mr. M. Manivannan, Information Analyst, Central Library, IIT Kharagpur

Contents	Page No
Preliminaries	ii-xxii
<i>Preface</i>	<i>ii-iii</i>
<i>Acknowledgement</i>	<i>iv</i>
<i>List of Annexure for Vocabulary Use</i>	<i>ix</i>
<i>List of Abbreviations</i>	<i>x</i>
<i>Executive Summary</i>	<i>xi-xxiii</i>
Details of Manual	1-104
Chapter 1: General Introduction and Concepts	1-7
1.0 Introduction	1
1.1 Generic Metadata	1
1.2 Educational Metadata	2
1.3 Multimedia Metadata (MPEG- 7)	2
1.4 Theses & Dissertation Metadata	2
1.5 Vocabulary for Educational Metadata	2
1.6 Metadata Schema	2
1.6.1 Qualified Dublin Core	2
1.6.2 IEEE LOM	3
1.6.3 LRMI	3
1.6.4 MPEG-7 Standard	3
1.6.5 Shodhganga	4
1.7 Requirement Specification	4
1.8 Required Extensions	4
1.9 Application Profile for Metadata Extension	5
Chapter 2: Generic Metadata	8-49
2.0 Application Profile for Generic Metadata	8
2.1 Contributor	8
2.1.1 Author	9
2.1.2 Illustrator	9
2.1.3 Editor	10

2.1.4	Other	11
2.2	Coverage	11
2.2.1	Temporal	12
2.2.2	Spatial	13
2.3	Creator	15
2.4	Date	16
2.4.1	Accessioned	18
2.4.2	Available	19
2.4.3	Created	19
2.4.4	Issued	20
2.4.5	Submitted	21
2.4.6	Updated	21
2.4.7	Copyright	21
2.5	Description	22
2.5.1	Abstract	22
2.5.2	Sponsorship	23
2.5.3	Table of Contents	23
2.5.4	URI	24
2.6	Format	24
2.6.1	Extent	25
2.6.2	Mime Type	25
2.7	Identifier	26
2.7.1	ISBN	27
2.7.2	ISSN	28
2.7.3	URI	28
2.7.4	Citation	29
2.7.5	Other	29
2.8	Language	29
2.9	Publisher	30
2.10	Relation	31
2.10.1	Is Referenced By	32
2.10.2	Is Part Of	33
2.10.3	Requires	34
2.10.4	Has Part	35

2.10.5	Is Part of Series	37
2.10.6	References	37
2.11	Rights	38
2.11.1	Holder	39
2.11.2	License	40
2.12	Source	41
2.12.1	URI	42
2.13	Subject	42
2.13.1	DDC	43
2.13.2	LCC	44
2.13.3	LCSH	45
2.13.4	MESH	45
2.13.5	Other	46
2.14	Title	47
2.14.1	Alternative title	47
2.15	Type	48
Chapter 3: Educational Metadata		50-64
3.0	Application Profile for Educational Metadata	50
3.1.	Educational Alignment	52
3.1.1	Alignment Type	52
3.1.2	Educational Framework	52
3.1.3	Educational Level	53
3.1.4	Pedagogic Objective	54
3.2.	Educational Role	54
3.3.	Educational Use	55
3.4.	Interactivity type	55
3.5.	Learning Resource Type	56
3.6.	Time Required	58
3.7.	TypicalAgeRange	59
3.8.	UseRightsURL	60
3.9.	isBasedOnURL	60
3.10.	DifficultyLevel	61

3.11.	Accessibility Related Fields	62
Chapter 4: Multimedia Metadata (MPEG-7)		65-72
4.0	Introduction	65
4.1	MPEG-7: the basic concepts	66
4.2	Application of MPEG- 7	68
4.3	Dublin Core to MPEG-7 Mapping	70
Chapter 5: Theses and Dissertation Metadata		73-77
5.0	Application Profile for Thesis & Dissertation	73
5.1	Advisor	74
5.2	Researcher	74
5.3	Awarded	74
5.4	Date	75
5.5	Department	75
5.6	Institution	76
5.7	Place	76
5.8	Degree	76
Chapter 6: Worked Out Examples		78-89
6.0	Introduction	78
6.1	Example 1: Generic Metadata and Educational Metadata	78
6.2	Example 2: Generic Metadata and Educational Metadata	82
6.3	Example 3: Generic Metadata and Educational Metadata	86
Chapter 7: Annexure for Vocabulary Use		90-103
References		104

List of Annexure for Vocabulary Use

S.N	Contents	Page No.
Annexure 2.5.3	Example for Table of Contents	23
Annexure 2.6.2	Vocabulary for Mime Type	25
Annexure 2.8	Vocabulary for Language	29
Annexure 2.15	Vocabulary for Type	48
Annexure 3.1.1	Vocabulary for Alignment Type	52
Annexure 3.1.2	Vocabulary for Educational Alignment – Educational Framework	52
Annexure 3.1.3	Vocabulary for Educational Level	53
Annexure 3.2	Vocabulary for Educational Role	54
Annexure 3.3	Vocabulary for Educational Use	55
Annexure 3.4	Vocabulary for Interactivity Type	55
Annexure 3.5	Vocabulary for Learning Resource Type	56
Annexure 3.6	Vocabulary for Duration (Time Required)	58
Annexure 3.7	Vocabulary for Typical Age Range	59
Annexure 3.10	Vocabulary for Difficulty Level	61
Annexure 3.11	Vocabulary for Accessibility Features	62
Annexure 5.8	Vocabulary for Type of Degree	76

List of Abbreviation

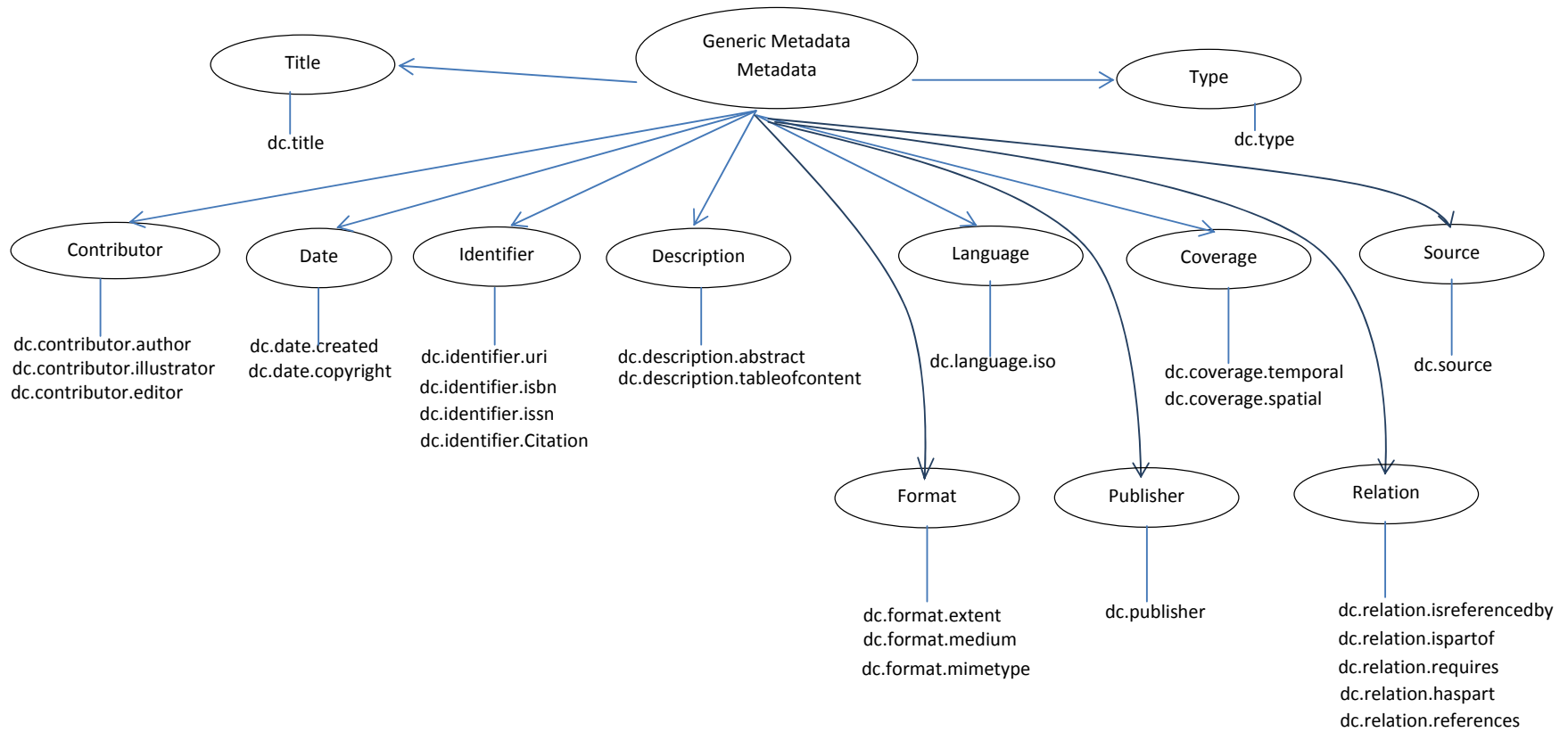
Term	Expansion
DC	Dublin Core
DCAP	Dublin Core Application Profile
DCMES	Dublin Core Metadata Element Set
DCMI	Dublin Core Metadata Elements
DDC	Dewey Decimal Classification
DOI	Digital Object Identifier
ETD	Electronic Theses and Dissertation
ISBN	International Standard Book Number
ISSN	International Standard Serial Number
LCC	Library of Congress Classification
LCSH	Library of Congress Subject Headings
LOM	Learning Object Metadata
LRMI	Learning Resource Metadata Initiative
MARC	MACHINE Readable Catalogue
MESH	Medical Subject Headings
MPEG	Moving Picture Experts Group
NDL	National Digital Library
NMEICT	National Mission in Education through ICT
QDC	Qualified Dublin Core
TOC	Table of Contents
UDC	Universal Decimal Classification
URI	Uniform Resource Identifier

Executive Summary

NDL has been envisaged to be a huge repository of digital content from varying domains and content categories. Variation in content category motivated the exploration of different metadata standards to define metadata schema for National Digital Library. Based on the variation in contents, metadata schema in NDL has been categorized into three classes:

- **Generic Metadata:** This set of metadata describes general attributes of the contents. Generic metadata includes contributor, identifier, date, language, subject etc. This set of metadata fields have been adopted from Dublin Core metadata standard.
- **Educational Metadata:** This set of metadata describes the educational attributes of the resources and helps in enumerating properties of the contents relevant to teaching-learning process. This metadata set includes educational level, type of learning material, educational use etc. This set of metadata has been adopted from Learning Resource Metadata Initiative (LRMI).
- **Thesis Metadata:** Dissertation or thesis related metadata fields are described with this set of metadata. It includes metadata fields like researcher, advisor, degree etc. Shodhganga thesis metadata standard has been used to represent this metadata set.

Generic Metadata



Metadata	Multi-value?	Standard	ShortDescription
dc.contributor.author	multi	lastName, firstName	A set of entities who have authored the content of the resource.
dc.contributor.editor	multi	lastName, firstName	A person or group responsible for editing of an item.
dc.contributor.illustrator	multi	lastName, firstName	A person or group responsible for illustrations contained in an item.
dc.contributor.other	multi	lastName, firstName (may be organization Name)	except the above list
dc.language.iso	*single	ISO 639	Language
dc.coverage.temporal	multi	free text	Coverage will typically include temporal period that has been talked about in the resource
dc.coverage.spatial	multi	free text	Coverage will typically include spatial location (a place name or geographic coordinates)
dc.description.abstract	single	free text	A summary of the resource.
dc.description.sponsorship	multi	free text	Sponsor
dc.description.tableofcontents	single	Refere to the format in Table_of_content worksheet	Table of Contents
dc.description.uri	single	URI	An unambiguous reference to the resource within a given context. (Typical value is URL for accessing the resource)
dc.description	multi	free text	Description

Metadata	Multi-value?	Standard	ShortDescription
dc.identifier.isbn	single	Free text	International Standard Book Number (ISBN)
dc.identifier.issn	single	Free text	International Standard Serial Number (ISSN)
dc.identifier.uri	*single	must be matched with URL regEx	URI
dc.identifier.citation	single	.bib format	Bibliographic format
dc.identifier.other	multi	May be a filename	except ISBN, ISSN, URI
dc.subject.ddc	multi	<DDC code>: <Subject String>	Dewey Decimal Classification (DDC) code
dc.subject.lcc	multi	<LCC code>: <Subject String>	Library of Congress Classification (LCC) code
dc.subject.lcsh	multi	<LCSH code>: <Subject heading String>	Library of Congress Subject Heading (LCSH) code
dc.subject.other	multi	free text	
dc.subject	multi	free text	Subject keywords
dc.relation.haspart	multi	Refere to the format in Table_of_content	If a resource is having multiple parts, this relation list down the titles of individual parts
dc.relation.ispartof	multi	exact "dc.title" value (case sensiive)	Reverse of has part

Metadata	Multi-value?	Standard	ShortDescription
dc.relation.isreferencedby	multi	exact "dc.title" value (case sensiive)	The described resource is referenced, cited, or otherwise pointed to by the referenced resource.
dc.relation.requires	multi	free text	Required software/hardware
dc.relation.references	multi	exact "dc.title" value is taken from the internal repository (case sensiive)	The described resource references, cites, or otherwise points to the referenced resource.
dc.relation	multi		except the above list
dc.source	*single	Free text	Source from which the resource has been acquired
dc.source.uri	*single	URI	URI that locates the source organization
dc.format.extent	multi		May include number of pages, size in Bytes, ISO 8601 for time duration in case of audio or video resource
dc.format.mimetype	single	Controlled Vocabulary	The mimetype of the resource
dc.type	*single	Controlled vocabulary	The nature or genre of the content of the resource.
dc.date.copyright	single	yyyy-mm-dd	
dc.date.created	single	yyyy-mm-dd	
dc.date.issued	single	yyyy-mm-dd	

Metadata	Multi-value?	Standard	ShortDescription
dc.date.submitted	single	yyyy-mm-dd	
dc.title	*single	avoid strings like, Lecture X, Chapter X	Typically, a title will be a name by which the resource is formally known.
dc.title.alternative	multi	avoid strings like, Lecture X, Chapter X	May be used to express title in language other than that used in dc.title
dc.rights.holder	multi		The person or organization that holds the rights in and over the resource.
dc.rights.license	multi		The license in which the resource is covered in. For example, GNU License, Creative Commons etc.
dc.publisher	multi (for book *)	free text	An entity responsible for making the resource available.

Format for Table of Contents

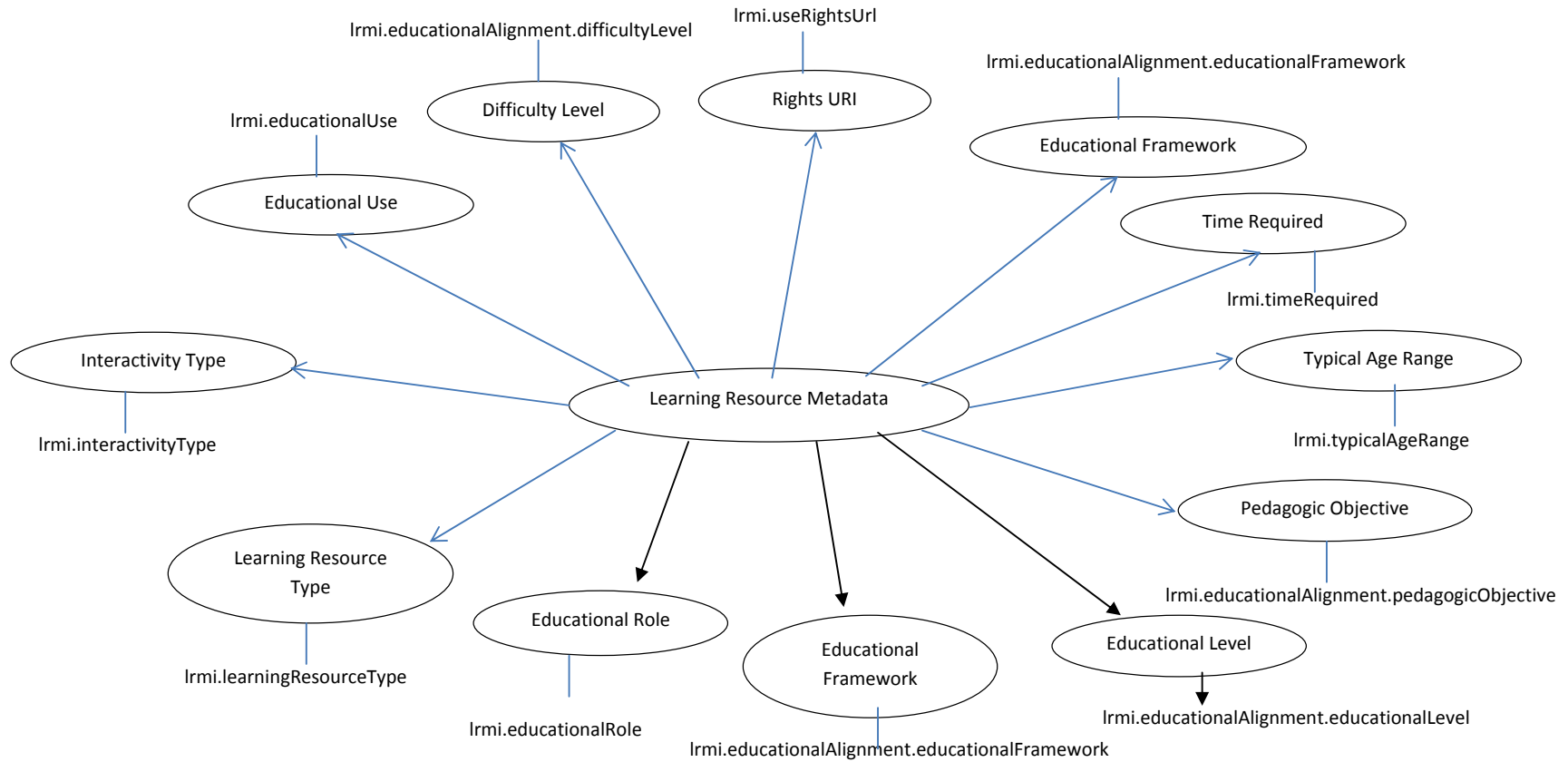
Actual Table of	
Business Vignette: The Relational Revolution	
Chapter 1 : Database Systems	
Why DataBases?	5
Introducing the Database	
Role and Advantages of the DBMS	7
Types of Databases	9
Why Database Design is important	10
Evolution of File System Data Processing	
Manual File Systems	11
File System Redux	14
Summary	20
Key Terms	25
Review Questions	26
Problems	26
Chapter 2 : Data Models	
Data Modeling and Data Models	30
The Importance of Data Models	30
Data Model Basic Building Blocks	31
Business Rules	
Discovering Business Rules	33
Naming Conventions	
Actual Naming	34
Formal Naming	35

Value for Table of Content
{
"Business Vignette": "The Relational Revolution",
"Chapter 1 : Database Systems": {
"Why DataBases?": 5,
"Introducing the Database": {
"Role and Advantages of the DBMS": 7,
"Types of Databases": 9
},
"Why Database Design is important": 10,
"Evolution of File System Data Processing": {
"Manual File Systems": 11,
"File System Redux ": 14
},
"Summary": 20,
"Key Terms": 25,
"Review Questions": 26,
"Problems": 26
},
"Chapter 2 : Data Models": {
"Data Modeling and Data Models": 30,
"The Importance of Data Models": 30,
"Data Model Basic Building Blocks": 31,
"Business Rules": {
"Discovering Business Rules": 33,
"Naming Conventions": {
"Actual Naming": 34,
"Formal Naming": 35
}
}
}
}
}
}

Controlled Vocabulary for Generic Metadata

# dc.format.mimetype	# dc.language.iso	dc.type
WAV	English	Text
AAC	Hindi	Video
MP3	Bengali	Audio
MP4	Assamese	Image
OGG	Bhojpur	Presentation
Flac	Gujarati	Application
MIDI / MID	Kannada	Animation
WMA	Kashmiri	Simulation
GIF	Malayalam	
JPG / JPEG	Marathi	
PNG	Nepali	
BMP	Oriya	
SVG / SVGZ	Punjabi	
TIF / TIFF	Sanskrit	
DjVu / DJV	Sindhi	
CSV	Tamil	
HTM / HTML	Telugu	
TXT	Urdu	
RTF	French	
XML	Italian	
3G2	Spanish	
3GP		
AVI		
MPEG		

Learning Resource Metadata

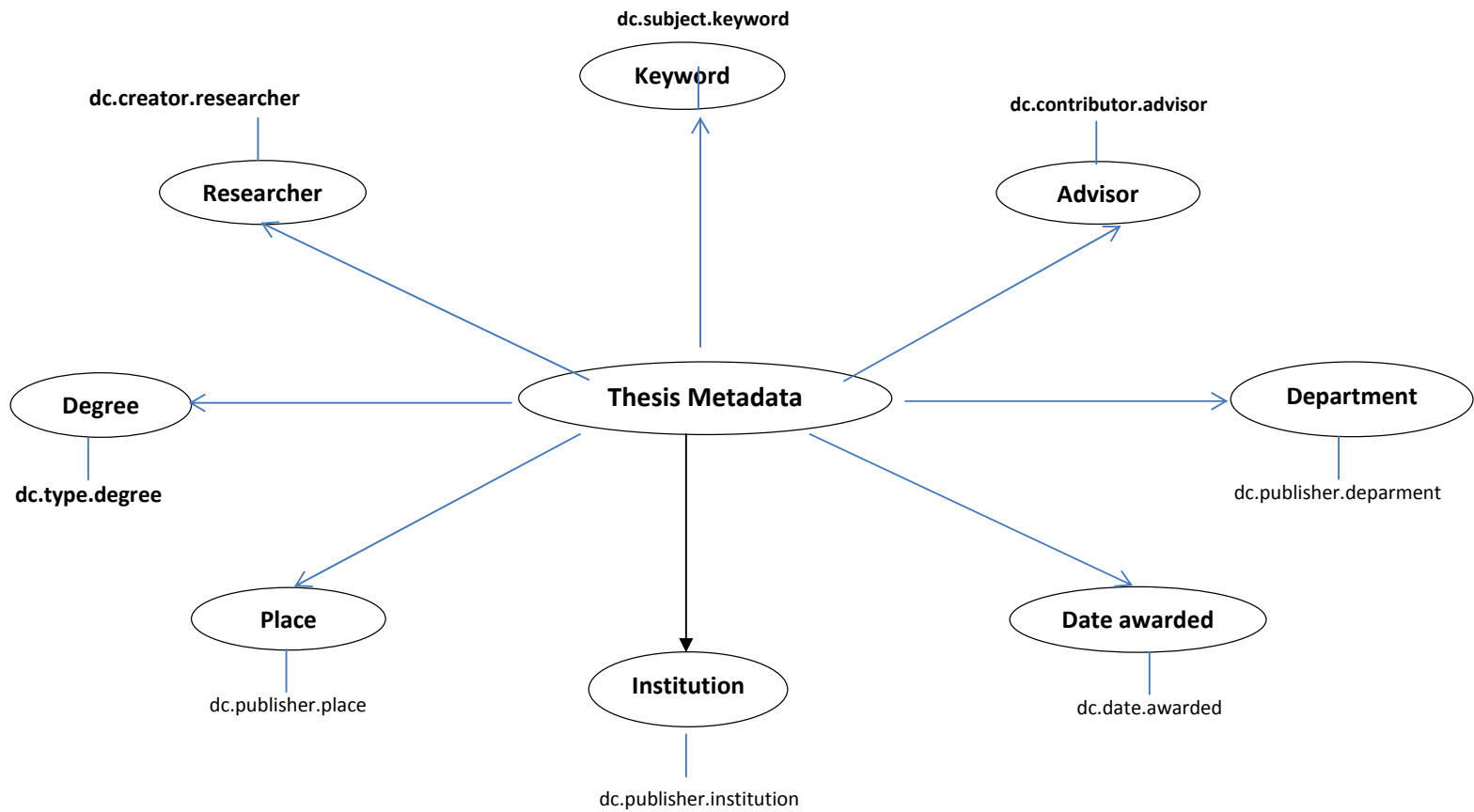


Metadata	Multi-value?	Standard	ShortDescription
lrmi.educationalUse	multi	Controlled Vocabulary	The purpose of the work in the context of education. Ex: “assignment” or “group work”
lrmi.timeRequired	single	ISO 8601	Approximate or typical time it takes to work with or through this learning resource for the typical intended audience. Ex: “P30M” or “P1H25M”
lrmi.typicalAgeRange	multi	Controlled Vocabulary	The typical range of ages of the content’s intended end user. Ex: “7-9” or “18-”
lrmi.interactivityType	single	Controlled Vocabulary	The predominant mode of learning supported by the learning resource. Ex: “active” , “expositive” or “mixed”
lrmi.learningResourceType	multi	Controlled Vocabulary	The predominant type or kind characterizing the learning resource. Ex: “presentation” or “handout”
lrmi.useRightsUrl	multi		The URL where the owner specifies permissions for using the resource.
lrmi.isBasedOnUrl	multi		A resource that was used in the creation of this resource. This term can be repeated for multiple sources.
lrmi.educationalRole	multi	Controlled Vocabulary	The role that describes the target audience of the content. Ex: “student” or “teacher”
lrmi.educationalAlignment.educationalFramework	multi	Controlled Vocabulary	Name of educational bodies to which the resource is aligned to
lrmi.educationalAlignment.educationalLevel	multi	Controlled Vocabulary	Grade level to which the resource is aligned to
lrmi.educationalAlignment.pedagogicObjective	multi	free text	educational objective of the resource
lrmi.educationalAlignment.difficultyLevel	single	Controlled Vocabulary	Level of difficulty of the resource with respect to the target educational level

Controlled Vocabulary

Learning Resource Type	Educational Use	Educational Level	Typical Age Range	Interactivity Type	Educational Role	Difficulty Level
Activity, Audio Lecture, Presentation, Book, Lab Material, Hands On, Handout, Web Course Quiz, Simulation, Lesson Plan, Educational Game, Educational App, Graph, Map, Video Lecture, Question Paper, Technical Manual/ Report, Animation, Exercise, Question Set, Self Assessment, Model Answers, Solution, Article	Analogies , Classifying, Demonstration, Experimental, Classroom Research, Assessment, Brainstroming, Debate, Drill / Practice, Field Trip, Homework, Reading, Inquiry, Lecture, Problem Solving, Project, Reflection, Assignment, Group Work, Self Learning, Solution	upto I, I – IV, V – VIII, IX – X, XI – XII, UG – PG, Career Development / Technical Study, Adult Education	3-7, 8-12, 13-17, 18-22, 22+	Active, Expositive, Mixed	Student, Teacher, Parent, School Admin, State Admin, Country Admin	Easy, Medium, Difficult

Thesis Metadata



Metadata	Multi-value?	Standard	ShortDescription
dc.date.awarded	single	yyyy-mm-dd	Date on which the degree was awarded
dc.type.degree	single	Controlled Vocabulary	Degree awarded to the recipient
dc.creator.researcher	single	lastName, firstName	Researcher
dc.publisher.institution	multi	country, state, city (Capitalize each word)	Name of instituion
dc.publisher.department	single	country, state, city (Capitalize each word)	Name of the department
dc.publisher.place	multi	country, state, city (Capitalize each word)	Place, typically the institute location
dc.publisher.date	single	yyyy-mm-dd	Publishing data
dc.contributor.advisor	multi	lastName, firstName	The supervisors on the thesis