

ISLE Metadata Initiative (IMDI)

PART 1 B

**Metadata Elements
for
Catalogue Descriptions**

Version 3.0.13

August 2009

INDEX

1	INTRODUCTION	3
2	CATALOGUE ELEMENTS OVERVIEW	4
3	METADATA ELEMENT DEFINITIONS	6
3.1	CATALOGUE	6
3.1.1	<i>Catalogue . Name</i>	7
3.1.2	<i>Catalogue . Title</i>	7
3.1.3	<i>Catalogue . Id</i>	7
3.1.4	<i>Catalogue . Description</i>	7
3.1.5	<i>Catalogue . Subject Languages</i>	7
3.1.6	<i>Catalogue . Document Languages</i>	7
3.1.7	<i>Catalogue . Location</i>	7
3.1.8	<i>Catalogue . Content Type</i>	8
3.1.9	<i>Catalogue . Format</i>	8
3.1.10	<i>Catalogue . Quality</i>	9
3.1.11	<i>Catalogue . Smallest Annotation Unit</i>	9
3.1.12	<i>Catalogue . Applications</i>	9
3.1.13	<i>Catalogue . Date</i>	10
3.1.14	<i>Catalogue . Project</i>	10
3.1.15	<i>Catalogue . Publisher</i>	10
3.1.16	<i>Catalogue . Author</i>	10
3.1.17	<i>Catalogue . Size</i>	10
3.1.18	<i>Catalogue . Distribution Form</i>	10
3.1.19	<i>Catalogue . Access</i>	10
3.1.20	<i>Catalogue . Pricing</i>	10
3.1.21	<i>Catalogue . Contact Person</i>	11
3.1.22	<i>Catalogue . Reference Link</i>	11
3.1.23	<i>Catalogue . Metadata Link</i>	11
3.1.24	<i>Catalogue . Publications</i>	11
3.1.25	<i>Catalogue . Keys</i>	11
3.2	SUB-SCHEMAS	11
3.2.1	<i>Subject Languages . Language</i>	11
3.2.2	<i>Document Languages . Language</i>	12
4	VOCABULARIES	13
4.1	CATALOGUE . CONTENT TYPE	13
4.2	CATALOGUE . FORMAT	13
4.2.1	<i>Catalogue . Format . Text</i>	13
4.2.2	<i>Catalogue . Format . Audio</i>	13
4.2.3	<i>Catalogue . Format . Video</i>	13
4.2.4	<i>Catalogue . Format . Image</i>	13
4.3	CATALOGUE . SMALLEST ANNOTATION UNIT	13
4.4	CATALOGUE . APPLICATIONS	13
	REFERENCES	14
	APPENDIX A : CATALOGUE METADATA INVENTORY	15
	APPENDIX B : REVISION HISTORY	16

1 Introduction

In discussing the first set of IMDI metadata elements with various people we came to the conclusion that some metadata elements that describe “published” corpora at the top level (for instance the number of CD’s) are not present in the IMDI set or are inappropriate to describe at the Session level. We therefore tried to make an inventory of such elements using existing description formalisms used by institutions that deal with “published corpora” such as [\[ELRA\]](#) and [\[LDC\]](#). We call the set of metadata elements that describe “published corpora” at the top-level “catalogue” metadata elements for language resources.

This proposal has been put forward after studying the current catalogue structure of ELRA, LDC and the UHLCS metadata requirements (compiled by Pirkko Suihkonen). It takes as a starting point all the fields specified by ELRA / LDC and looks for any matching UHLCS fields (see Appendix A).

What we have done is to categorise the fields and try to determine which of them are already covered in the IMDI proposal for Sessions. Also we try to determine which elements of the IMDI proposal for Sessions would better fit in this catalogue metadata set or would make sense to be duplicated in the catalogue set.

It would be feasible to implement this catalogue metadata set for language resources in Dublin-Core [\[DCMES\]](#). This is to be expected since it is metadata about published material without much information on the constituting parts. Therefore we have made a suggestion for DC equivalents with all the proposed catalogue elements. Sometimes it appears that a choice is possible, this has been indicated by a question mark.

Some concluding remarks:

- We are not yet sure that a separate IMDI catalogue metadata set is useful in all cases. It would seem so if you were looking for a corpus you can obtain as a unit and which has probably been produced for your needs (e.g. language technology applications).
- The flat description of a whole corpus by this set seems to make it ideal for easy identification of corpora by less-specific (in comparison to the IMDI Session vocabulary) search such as OAI.

All proposed DC qualifiers are our own except for Subject.language which is [\[OLAC\]](#)’s.

2 Catalogue Elements Overview

Element Name		DC Equivalent	Definition	Encoding	Connection to Session Elements
Name		Title.short	The name of the corpus	string	duplicate
Title		Title	The title of the corpus	string	duplicate
Id +		ID	A Unique identifier for the corpus. For example an ISBN.	string	None
Description +		Description	A description of the corpus	string	duplicate
Document Languages		Language	The languages used for describing/annotating the corpus	Document Languages . Language *	Summation of
Subject Languages		Subject. language ([OLAC] qualifier)	The languages subject of analysis	Subject Languages . Language *	Summation of
Location +		Coverage	Groups the information about the location of where the corpus content was made	continent, country, region	Generalisation
	Continent		The continent of where the corpus content was made	closed controlled vocabulary	
	Country		The country of where the corpus content was made	ISO3166-1	
	Region		The region or sub-region of where the corpus content was made	string	
	Address		The address of where the corpus content was created	string	
Content type +		Type? Subject.type?	The type of the corpus	open vocabulary list Type = { <i>written, speech, terminology</i> } Subtype = { <i>Corpus, Monoling. Lex., Multiling. Lex., Telephone speech, Desktop/microphone, Multimodal/Multi media, Other speech related</i> } See ELRA	Generalisation Might use: - Session . Content . Genre and SubGenre
Format		Format	Groups information about the formats used in the corpus	text, audio, video	Implicit in resource description
	Text		The format of the text used in the corpus	open vocabulary list	
	Audio		The format of the audio used in the corpus	open vocabulary list	
	Video		The format of the video used in the corpus	open vocabulary list	

	Image		The format of images used in the corpus	open vocabulary list	
Quality		Format.quality	Groups information about the quality of the corpus content	audio, video	
	Text		The quality of the text data in the corpus	closed controlled vocabulary {1..5}	
	Audio		The quality of the audio data in the corpus	closed controlled vocabulary {1..5}	
	Video		The quality of the video data in the corpus	closed controlled vocabulary {1..5}	
	Image		The quality of the image data in the corpus	closed controlled vocabulary {1..5}	
Smallest Annotation Unit		Format.unit	The smallest unit of annotation used in the corpus	open vocabulary {paragraph, utterance, word, phoneme, ...}	Summation of Might use: - WrittenResource . SubType (in case of Annotation)
Applications		Type.application ? Subject.application?	The application domain of the corpus	Open Vocabulary List. Based on LDC corpus catalogue.	Summation of
Date		Date.Issued	Publishing date	YYYY-MM-DD	none
Project + (Name, Title, Id, Contact, Description *)		Creator.project		See Session Metadata	duplicate
Publisher +		Publisher	An entity responsible for making the resource available	string	none
Author +		Creator	An entity primarily responsible for making the content of the resource	string	Session.Project.Contact.Name (can be different contact names in different sessions)
Size		Format.Extent	Total size of the corpus	string	total sum
Distribution Form		Type.distribution? Format.distribution?	How are the corpora distributed	open vocabulary list {CD, ftp, ... }	none
Access		Rights	The access conditions of the corpus	sub-schema	none
Pricing		Rights.price	The price of the corpus	string	none
Contact Person			Person to be contacted about the catalogue	string	
Reference Link			URL to the catalogue	string	
Metadata Link			URL to the metadata for the catalogue	string	
Publications			List of any publications related to the catalogue	string	
Keys					

3 Metadata Element Definitions

The elements for session descriptions are defined using the following attributes:

- *Element/Group Name*
A name of the element or grouping.
- *Identifier*
A unique identifier assigned to the element.
- *Definition*
A statement that clearly represents the concept and essential nature of the data element.
- *Encoding*
A statement that describes how the content of the element is encoded.
- *Comment*
Remarks concerning the application of the data element.
Dublin Core equivalent: some elements can be mapped with the Dublin Core Metadata Element Set [[DCMES](#)]. If this is possible, the Dublin Core equivalent of the IMDI element will be named here.
Example: sometimes an example helps to clarify the use of the element. If this is the case, the example will be mentioned here.

3.1 Catalogue

Group: Catalogue
Identifier: Catalogue
Definition: Groups information about a published corpus.
Encoding: Catalogue . Name
Catalogue . Title
Catalogue . Id +
Catalogue . Description +
Catalogue . Subject Languages
Catalogue . Document Languages
Catalogue . Location
Catalogue . Content Type
Catalogue . Format
Catalogue . Quality
Catalogue . Smallest Annotation Unit
Catalogue . Applications
Catalogue . Date
Catalogue . Project
Catalogue . Publisher
Catalogue . Author
Catalogue . Size
Catalogue . Distribution Form
Catalogue . Access
Catalogue . Pricing
Catalogue . Contact Person
Catalogue . Reference Link
Catalogue . Metadata Link
Catalogue . Publications
Catalogue . Keys

Comments:

3.1.1 Catalogue . Name

Element: Catalogue . Name
Identifier: Catalogue . Name
Definition: Name of the corpus.
Encoding: string
Comments:

3.1.2 Catalogue . Title

Element: Catalogue . Title
Identifier: Catalogue . Title
Definition: Title of the corpus.
Encoding: string
Comments:

3.1.3 Catalogue . Id

Element: Catalogue . Id
Identifier: Catalogue . Id
Definition: Unique identifier for the corpus.
Encoding: string
Comments: This can be an ISBN.

3.1.4 Catalogue . Description

Element: Catalogue . Name
Identifier: Catalogue . Name
Definition: Description of the corpus.
Encoding: string
Comments:

3.1.5 Catalogue . Subject Languages

Element: Catalogue . Subject Language
Identifier: Catalogue . SubjectLanguage
Definition: Languages subject of analysis.
Encoding: Subject Languages . Description *
Subject Languages . Language *
Comments:

3.1.6 Catalogue . Document Languages

Element: Catalogue . Document Languages
Identifier: Catalogue . DocumentLanguages
Definition: The languages used for describing and annotating the corpus data.
Encoding: Document Languages . Description *
Document Languages . Language *
Comments:

3.1.7 Catalogue . Location

Group: Catalogue . Location
Identifier: Catalogue . Location
Definition: Groups information about the location of where the corpus content was recorded or originated.
Encoding: Location . Continent
Location . Country
Location . Region
Comments:

Catalogue . Location . Continent

Element: Catalogue . Location . Continent
Identifier: Catalogue . Location . Continent
Definition: The continent of where the corpus content was recorded or originated.

Encoding: Closed controlled vocabulary { Africa, Antarctica, Asia, Australia, Europe, North America, Oceania, South America }.

Comments:

Catalogue . Location . Country

Element: Catalogue . Location . Country

Identifier: Catalogue . Location . Country

Definition: The country where the corpus content was recorded or originated.

Encoding: Closed controlled vocabulary. The country is encoded with a two-letter code as described by [\[ISO3166-1\]](#).

Comments:

Catalogue . Location . Region

Element: Catalogue . Location . Region

Identifier: Catalogue . Location . Region

Definition: The region or sub-region of where the corpus content was recorded or originated.

Encoding: string

Comments: This element can also be used to describe sub-regions.

Examples: europe, the netherlands, gelderland, achterhoek.

Catalogue . Location . Address

Element: Catalogue . Location . Address

Identifier: Catalogue . Location . Address

Definition: The address of where the corpus content was recorded or originated.

Encoding: string

Comments:

3.1.8 Catalogue . Content Type

Element: Catalogue . Content Type

Identifier: Catalogue . ContentType

Definition: The type of the corpus.

Encoding: Open vocabulary list '[Catalogue . Content . Type](#)' (4.1).

Comments:

3.1.9 Catalogue . Format

Group: Catalogue . Format

Identifier: Catalogue . Format

Definition: Groups information about the formats used in the corpus.

Encoding: Format . Text

Format . Audio

Format . Video

Format . Image

Comments:

Catalogue . Format . Text

Element: Catalogue . Format . Text

Identifier: Catalogue . Format . Text

Definition: The formats of the text used in the corpus.

Encoding: Open Vocabulary List '[Catalogue . Format . Text](#)' (4.2.1)

Comments:

Catalogue . Format . Audio

Element: Catalogue . Format . Audio

Identifier: Catalogue . Format . Audio

Definition: The formats of the audio data used in the corpus.

Encoding: Open Vocabulary List '[Catalogue . Format . Audio](#)' (4.2.2)

Comments:

Catalogue . Format . Video

Element: Catalogue . Format . Video
Identifier: Catalogue . Format . Video
Definition: The formats of the video data used in the corpus.
Encoding: Open Vocabulary List '[Catalogue . Format . Video](#)' (4.2.3)
Comments:

Catalogue . Format . Image

Element: Catalogue . Format . Image
Identifier: Catalogue . Format . Image
Definition: The formats of the image data used in the corpus.
Encoding: Open Vocabulary List 'Catalogue . Format . Image' (4.2.3)
Comments:

3.1.10 Catalogue . Quality

Group: Catalogue . Quality
Identifier: Catalogue . Quality
Definition: Groups information about the quality of the corpus content.
Encoding: string
Comments:

Catalogue . Quality . Text

Element: Catalogue . Quality . Text
Identifier: Catalogue . Quality . Text
Definition: The quality of the text data in the corpus.
Encoding: Closed controlled vocabulary {1 .. 5, Unknown, Unspecified}.
Comments:

Catalogue . Quality . Audio

Element: Catalogue . Quality . Audio
Identifier: Catalogue . Quality . Audio
Definition: The quality of the audio data in the corpus.
Encoding: Closed controlled vocabulary {1 .. 5, Unknown, Unspecified}.
Comments:

Catalogue . Quality . Video

Element: Catalogue . Quality . Video
Identifier: Catalogue . Quality . Video
Definition: The quality of the video data in the corpus.
Encoding: Closed controlled vocabulary {1 .. 5, Unknown, Unspecified}.
Comments:

Catalogue . Quality . Image

Element: Catalogue . Quality . Image
Identifier: Catalogue . Quality . Image
Definition: The quality of the image data in the corpus.
Encoding: Closed controlled vocabulary {1 .. 5, Unknown, Unspecified}.
Comments:

3.1.11 Catalogue . Smallest Annotation Unit

Element: Catalogue . Smallest Annotation Unit
Identifier: Catalogue . SmallestAnnotationUnit
Definition: The smallest annotation unit used in the corpus.
Encoding: Open vocabulary '[Catalogue . Smallest Annotation Unit](#)' (4.3).
Comments:

3.1.12 Catalogue . Applications

Element: Catalogue . Applications

Identifier: Catalogue . Applications
Definition: The application domain of the corpus.
Encoding: Open vocabulary list '[Catalogue . Application](#)' (4.4).
Comments:

3.1.13 Catalogue . Date

Element: Catalogue . Date
Identifier: Catalogue . Date
Definition: The publishing date of the corpus.
Encoding: The date is encoded according to a profile of [[ISO8601](#)] as described in [[W3CDTF](#)] and follows the YYYY-MM-DD format.
Comments:

3.1.14 Catalogue . Project

Element: Catalogue . Project
Identifier: Catalogue . Project
Definition: The project for which the corpus was originally created.
Encoding: See Project in the Session Metadata document
Comments:

3.1.15 Catalogue . Publisher

Element: Catalogue . Publisher
Identifier: Catalogue . Publisher
Definition: An entity responsible for making the resource available.
Encoding: string
Comments:

3.1.16 Catalogue . Author

Element: Catalogue . Authors
Identifier: Catalogue . Authors
Definition: An entity primarily responsible for making the content of the resource.
Encoding: string
Comments:

3.1.17 Catalogue . Size

Element: Catalogue . Size
Identifier: Catalogue . Size
Definition: Total size of the corpus.
Encoding: string
Comments:

3.1.18 Catalogue . Distribution Form

Element: Catalogue . Distribution Form
Identifier: Catalogue . DistributionForm
Definition: How are the corpora distributed.
Encoding: string
Comments:

3.1.19 Catalogue . Access

Group: Catalogue . Access
Identifier: Catalogue . Access
Definition: Groups information about access rights.
Encoding: Access (sub-schema)
Comments:

3.1.20 Catalogue . Pricing

Element: Catalogue . Pricing
Identifier: Catalogue . Pricing
Definition: The price of the corpus.

Encoding: string
Comments:

3.1.21 Catalogue . Contact Person

Element: Catalogue . Contact Person
Identifier: Catalogue . ContactPerson
Definition: Person to be contacted about the catalogue.
Encoding: string
Comments:

3.1.22 Catalogue . Reference Link

Element: Catalogue . Reference Link
Identifier: Catalogue . ReferenceLink
Definition: URL to the catalogue.
Encoding: string
Comments:

3.1.23 Catalogue . Metadata Link

Element: Catalogue . Metadata Link
Identifier: Catalogue . MetadataLink
Definition: URL to the metadata for the catalogue
Encoding: string
Comments:

3.1.24 Catalogue . Publications

Element: Catalogue . Publications
Identifier: Catalogue . Publications
Definition: List of any publications related to the catalogue.
Encoding: string
Comments:

3.1.25 Catalogue . Keys

Element: Catalogue . Keys
Identifier: Catalogue . Keys
Definition:
Encoding: See Keys in Session Metadata document
Comments:

3.2 Sub-schemas

3.2.1 Subject Languages . Language

Group: Subject Languages . Language
Identifier: SubjectLanguages . Language
Definition: Groups information about the language subject of analysis.
Encoding: Subject Languages . Language . Id
Subject Languages . Language . Name
Subject Languages . Language . Dominant
Subject Languages . Language . Source Language
Subject Languages . Language . Target Language
Subject Languages . Language . Description *

Comments:

Subject Languages . Language . Id

Element: Subject Languages . Language . Id
Identifier: SubjectLanguages . Language . Id
Definition: Specifies a unique code to identify the language.
Encoding: See 'Language Identifier Encoding' in the document Metadata Elements for Session Descriptions.
Comments: Dublin Core equivalent: DC:Language

Subject Languages . Language . Name

Element: Subject Languages . Language . Name
Identifier: SubjectLanguages . Language . Name
Definition: A human understandable name of the language.
Encoding: string
Comments: In general the names from the [[ETHNOLOGUE](#)] list from SIL International are recommended.

Subject Languages . Language . Dominant

Element: Subject Languages . Language . Dominant
Identifier: SubjectLanguages . Language . Dominant
Definition: Specifies that this is the most frequently used language in the corpus.
Encoding: {True, False, Unknown, Unspecified}.
Comments:

Subject Languages . Language . Source Language

Element: Subject Languages . Language . Source Language
Identifier: SubjectLanguages . Language . SourceLanguage
Definition: Specifies that this is the source language in the corpus.
Encoding: {True, False, Unknown, Unspecified}.
Comments: Used for language acquisition corpora.

Subject Languages . Language . Target Language

Element: Subject Languages . Language . Target Language
Identifier: SubjectLanguages . Language . TargetLanguage
Definition: Specifies that this is the target language in the corpus.
Encoding: {True, False, Unknown, Unspecified}.
Comments: Used for language acquisition corpora.

Subject Languages . Language . Description

Element: Subject Languages . Language . Description
Identifier: SubjectLanguages . Language . Description
Definition: Elaborate description of the language.
Encoding: [Description](#) (sub-schema)
Comments:

3.2.2 Document Languages . Language

Group: Document Languages . Language
Identifier: DocumentLanguage . Language
Definition: Groups information about the a language used for describing and annotating corpus data.
Encoding: Document Languages . Language . Id
DocumentLanguages . Language . Name
Comments:

Document Languages . Language . Id

Element: Document Languages . Language . Id
Identifier: DocumentLanguages . Language . Id
Definition: Specifies a unique code to identify the language.
Encoding: See 'Language Identifier Encoding' in the document Metadata Elements for Session Descriptions.
Comments: Dublin Core equivalent: DC:Language

Document Languages . Language . Name

Element: Document Languages . Language . Name
Identifier: DocumentLanguages . Language . Name
Definition: A human understandable name of the language.
Encoding: string

Comments: In general the names from the [[ETHNOLOGUE](#)] list from SIL International are recommended.

Details about the sub-schemas **Access**, **Project**, **Description** and **Keys** can be found in the document 'Metadata Elements for Session Descriptions'.

4 Vocabularies

4.1 Catalogue . Content Type

Open vocabulary list:

- Written
- Speech
- Terminology
- Corpus
- Monolingual Lexicon
- Multilingual Lexicon
- Telephone speech
- Desktop/microphone
- Multimodal/Multimedia
- Other speech related

4.2 Catalogue . Format

4.2.1 Catalogue . Format . Text

Open vocabulary list:

<http://www.mpi.nl/IMDI/Schema/WrittenResource-Format.xml>

4.2.2 Catalogue . Format . Audio

Open vocabulary list:

<http://www.mpi.nl/IMDI/Schema/CatalogueFormatAudio.xml>

4.2.3 Catalogue . Format . Video

Open vocabulary list:

<http://www.mpi.nl/IMDI/Schema/CatalogueFormatVideo.xml>

4.2.4 Catalogue . Format . Image

Open vocabulary list:

<http://www.mpi.nl/IMDI/Schema/CatalogueFormatImage.xml>

4.3 Catalogue . Smallest Annotation Unit

Open vocabulary:

<http://www.mpi.nl/IMDI/Schema/CatalogueAnnotationUnit.xml>

4.4 Catalogue . Applications

Open vocabulary list (based on the LDC Corpus Catalogue):

<http://www.mpi.nl/IMDI/Schema/CatalogueApplicationType.xml>

References

- [DCMES] Dublin Core Metadata Element Set
<http://dublincore.org/documents/dces/>
- [ETHNOLOGUE] Ethnologue language name index
<http://www.sil.org/ethnologue/names/>
- [ELRA] European Language Resources Association, <http://www.icp.grenet.fr/ELRA/>
- [ISO639-1] Code for the representation of names of languages, International Organization for Standardization (ISO), 1988.
- [ISO639-2]
Codes for the representation of names of languages - part 2: alpha-3 code, International Organization for Standardization (ISO), 1998.
<http://lcweb.loc.gov/standards/iso639-2/langhome.html>
- [ISO3166-1]
Codes for the representation of names of countries, International Organization for Standardization (ISO), 1997.
<http://www.din.de/gremien/nas/nabd/iso3166ma/codlstp1/index.html>
- [ISO8601] Data elements and interchange formats - Information interchange - Representation of dates and times, International Organization for Standardization (ISO), 2000.
- [LANGID] Language Identification and IT: Addressing problems of linguistic diversity on a global scale, Peter Constable and Gary Simons, SIL International, 2000.
<http://www.sil.org/silewp/2000/001/>
- [LDC] Linguistic Data Consortium, <http://morph ldc.upenn.edu/>
- [OAI] Open Archives Initiative, <http://www.openarchives.org/>
- [OLAC] Open Language Archives Community, <http://www.language-archives.org/>
- [RFC1766] Tags for the identification of language
<http://www.ietf.org/rfc/rfc1766.txt>
specifies a two letter code taken from [ISO639-1], followed optionally by a two letter country code taken from [ISO3166-1]
- [W3CDTF] Date and Time Formats, W3C Note
<http://www.w3.org/TR/NOTE-datetime>

Appendix A : Catalogue Metadata Inventory

	UHLCS	ELRA	LDC
Name	Corpus Title	Item Name	Item Name
Author			Authors
Project	Projects Sponsors		Project(s) +
Date		Creation date	
Catalogue Id's		ELRA Id	LDC Catalogue Id. NIST Catalogue Id. ISBN Catalogue Id.
Type/Quality		Type = { <i>written, speech, terminology, Tools & software</i> } Subtype = { <i>Corpus, Monoling. Lex., Multiling. Lex., Telephone speech, Desktop/microphone, Mutimodal/Multimedia, Other speech related</i> }	Data type = { <i>Lexicon, Speech, Text</i> } Data source = { <i>Broadcast, conversation, microphone, mobile-radio, newswire, parallel, pronunciation, telephone, varied</i> }
Format/Quality			Sample Frequency Sample Format
Subject Language	Main-language Languages	Language	Language(s) +
Documentation	Description of the corpus		Online documentation Readme file
Applications			Application types based on the LDC corpus catalogue http://wave ldc.upenn.edu/Catalog/
Media	Physical Storage format Server OS type Exploitation tools		Number of CD's (or) ftp
Distributor	Distributor name, address	Implicit	Implicit
Pricing		Member price Non-member price	Member price Non-member price Membership year (when free for members)
Access/ Licensing	Location Access how-to		Member license Non-member license

Appendix B : Revision History

Version 3.0.13

Date: 6 August 2009; MPI ISLE Team

Removed Catalogue . Begin Year
Removed Catalogue . End Year
Added Catalogue . Format . Image
Added Catalogue . Quality . Text
Added Catalogue . Quality . Image
Revised Catalogue . Authors (changed to Catalogue . Author)
Revised Catalogue . Author (allowed multiple occurrences)
Revised Catalogue . Project (allowed multiple occurrences)

Version 3.0.10

Date: 17 June 2009; MPI ISLE Team

Added Catalogue . Contact Person
Added Catalogue . Begin Year
Added Catalogue . End Year
Added Catalogue . Reference Link
Added Catalogue . Metadata Link
Added Catalogue . Publications
Revised Catalogue . Location (allowed multiple occurrences and added address element that was in schema anyway)
Revised Catalogue . Content Type (allowed multiple occurrences)
Revised Catalogue . Publisher (allowed multiple occurrences)

Version: 3.0.0

Date: 12 November 2004; MPI ISLE Team

Synchronized with IMDI 3.x
Added values for Catalogue . Application from LDC catalogue
Added Catalogue . Keys
Added encoding for Catalogue . Format . Text
Added encoding for Catalogue . Format . Audio
Added encoding for Catalogue . Format . Video
Changed 'Subject Language' into 'Subject Languages'
Changed 'Document Language' into 'Document Languages'
Added sub-schema 'Subject Languages . Language'
Added sub-schema 'Document Languages . Language'
Corrected CV 'Content Type' to CV list

Version: 2.1

Date: 8 June 2001; MPI ISLE Team

First frozen element set.

Version: 2.0

Date: 4 June 2001; MPI ISLE Team

Smallest Annotation Unit was added
Size of Corpus was added

Version: 1.0

Date: 23 February 2001; MPI ISLE Team

First version