This document is part of the Network of Excellence "Multilingual Europe Technology Alliance (META-NET)",
co- funded by the 7th Framework Programme of the European Commission through the T4ME grant agreement no.: 249119.

A Network of Excellence forging the
Multilingual Europe Technology Alliance

Documentation and User Manual of the META-SHARE Metadata Model

Editors: Penny Labropoulou, Elina Desipri

Authors: Elina Desipri, Maria Gavrilidou, Penny Labropoulou, Stelios Piperidis (R.C. Athena/ILSP), Francesca Frontini, Monica Monachini (ILC/CNR), Victoria Arranz, Valérie Mapelli (ELDA), Gil Francopoulo (LIMSI), Thierry Declerck (DFKI)

Date: 2011/11/23

Dissemination Level: Public

Grant agreement no.	249119
Project acronym	T4ME Net (META-NET)
Project full title	Technologies for the Multilingual European Information Society
Funding scheme	Network of Excellence
Coordinator	Prof. Hans Uszkoreit (DFKI)
Start date, duration	1 February 2010, 36 months
Distribution	Public
Contractual date of delivery	November 30, 2011
Actual date of delivery	December 21, 2011
Deliverable number	D7.2.3
Deliverable title	Documentation and User manual of the META-SHARE metadata model
Type	Report
Status and version	Final version (for metadata & platform v1.1)
Number of pages	158
Contributing partners	DFKI, CNR, ILSP, CNRS, ELDA
WP leader	ILSP
Task leader	ILSP
Authors	Elina Desipri, Maria Gavrilidou, Penny Labropoulou, Stelios Piperidis (R.C. "Athena"/ILSP), Francesca Frontini, Monica Monachini (ILC/CNR), Victoria Arranz, Valérie Mapelli (ELDA), Gil Francopoulo (CNRS), Thierry Declerq (DFKI)
EC project officer	Hanna Klimek
The partners in META-NET are:	Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI), Germany Barcelona Media (BM), Spain Consiglio Nazionale Ricerche – Istituto di Linguistica Computazionale "Antonio Zampolli" (CNR), Italy Institute for Language and Speech Processing, R.C. "Athena" (ILSP), Greece Charles University in Prague (CUP), Czech Republic Centre National de la Recherche Scientifique – Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur (CNRS), France Universiteit Utrecht (UU), Netherlands Aalto University (AALTO), Finland Fondazione Bruno Kessler (FBK), Italy Dublin City University (DCU), Ireland Rheinisch-Westfälische Technische Hochschule Aachen (RWTH), Germany Jozef Stefan Institute (JSI), Slovenia Evaluations and Language Resources Distribution Agency (ELDA), France

For copies of reports, updates on project activities and other META-NET-related information, contact:
DFKI GmbH
META-NET
Dr. Georg Rehm
Alt-Moabit 91c
10559 Berlin, Germany

office@meta-net.eu
Phone: +49 (30) 23895-1833
Fax: +49 (30) 23895-1810

Copies of reports and other material can also be accessed via http://www.meta-net.eu

Table of Contents

1 Executive Summary

The current deliverable presents the META-SHARE metadata schema v1.0, as implemented in the META-SHARE XSD's v1.0 released to (META-NET and PSP partners) in July 2011 for text corpora and lexical/conceptual resources and its supplement for audio corpora, tools and language descriptions (simplified/refactored version) as implemented in November.
It is meant to act as a user manual, providing explanations on the model contents for LRs providers and LRs curators that wish to describe their resources in accordance to it.
Work on the schema is ongoing and changes/updates to the model are constantly being made; where appropriate, some changes that are already under way are documented in this deliverable.
N.B. It should be noted that the accompanying set of XSD's has been created separately from the editor and there might be some discrepancies between the two; where these have been identified, they are mentioned as such in the deliverable.

2 Introduction

The current deliverable documents the metadata model proposed for the description of Language Resources (LRs) made available through META-SHARE, the open distributed facility for the sharing and exchange of resources of META-NET. A more detailed account of the theoretical principles and a general introduction to the model can be found in [Gavrilidou et al., 2011].
In the context of META-SHARE, the term **metadata** refers to descriptions of LRs, encompassing both ***data*** (textual, multimodal/multimedia and lexical data, grammars, language models etc.) and ***technologies*** (tools/services) used for their processing. These are also found in the literature as Language Resources and Technologies (LRTs).

3 Model essentials

3.1 Basic concepts

The META-SHARE metadata model has been implemented in the current version as an XML schema.
The mechanism we have adopted is the **component-based mechanism** (Component MetaData Infrastructure, CMDI) grouping together semantically coherent elements to form components and providing relations between them [Broeder et al., 2008].
More specifically, **elements** are used to encode specific descriptive features of the LRs. To cater for semantic consistency with other related schemas and models, a link to existing elements in the Dublin Core (DC,

www.dublincore.org) and the ISO Data Category Registry (ISO DCR, [ISO 12620, 2009]) is provided; where necessary, the new elements will populate the ISO DCR.

In addition, the notion of **relations**[1] has been introduced to give information on linking features between resources. Relations hold between the various forms of a LR (e.g. raw and annotated resource), different LRs (e.g. a language resource and the tool that has been used to create it etc.) - irrespective of whether these are included in the META-SHARE repository or not - as well as peripheral resources (e.g. standards used, related documentation etc.).

The set of all the components and elements describing specific LR types and subtypes represent the ***profile*** of this type. Obviously, certain components include information common to all types of resources (e.g. identification, contact, licensing information etc.) and are, thus, used for all LRs, while others (e.g. components including information on the contents, annotation etc.) differ across types.

In order to accommodate flexibility, the elements belong to two basic levels of description:

- an initial level providing the basic elements for the description of a resource (**minimal schema**), and
- a second level with a higher degree of granularity (**maximal schema**), providing detailed information on a resource and covering all stages of LR production and use.

The minimal schema contains those elements considered indispensable for LR description (from the provider's perspective) and identification (from the consumer's perspective). It takes into account the views expressed in the user survey conducted in the framework of WP7 (see [Federmann et al., 2011]) concerning which features are considered sufficient to give a sound "identity" to a resource.

These two levels contain four classes of elements: the first level contains Mandatory (M) and Condition-dependent Mandatory (MC) elements (i.e. to be filled in when specific conditions are met), while the second level includes Recommended (R) and Optional (O) elements.

3.2 The META-SHARE ontology

META-SHARE takes a global view on resources, aiming to provide users not only with a catalogue of LRs (data and tools) but also with information that can be used to enhance their exploitation. For instance, research papers that document the production of a resource as well as standards used and guidelines followed are informative for LR users and advisory for prospective LR producers.

In the proposed META-SHARE ontology (Figure 1), a distinction is made between LR per se and all other related resources/entities, such as reference documents related to the resource (papers, reports, manuals etc.), persons / organizations involved in their creation and use (creators, distributors etc.), related projects and activities (funding projects, activities of usage etc.) and licenses (for the distribution of the LRs).

1 In the current version of the model, relations have not been formally implemented and are represented as elements.

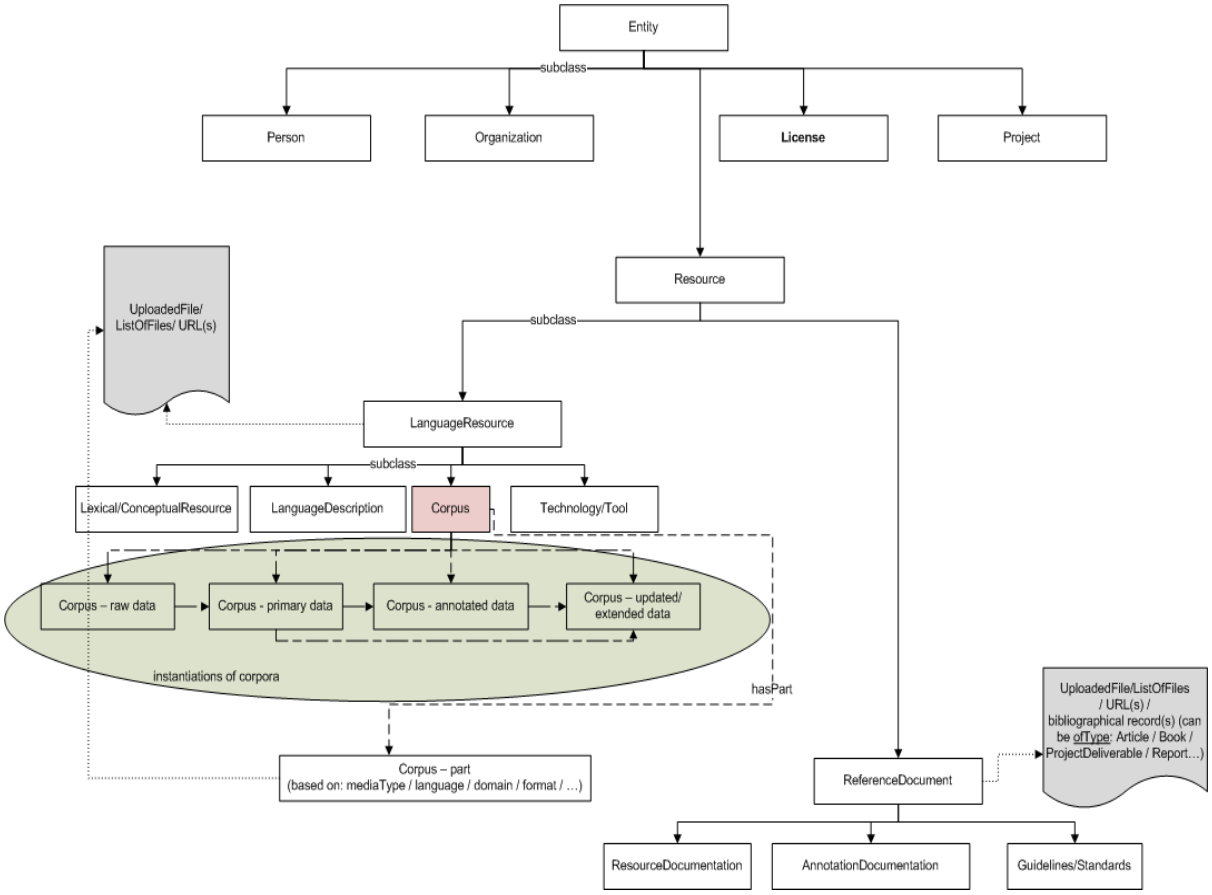

Figure : META-SHARE ontology excerpt

3.3 Proposed LR taxonomy

Central to the model is the LR taxonomy, which allows us to organize the resources in a structured way, taking into consideration the specificities of each type.

The proposed LR taxonomy constitutes an integral part of the metadata model, whereby the types of LRs (attributes and values) belong to the element set. The basic element used to categorize LRs in types that lead to coherent sets of descriptions is the ***resourceType*** with the following values:

- corpus (including written/text, oral/spoken, multimodal/multimedia corpora)
- lexical / conceptual resource (including terminological resources, word lists, semantic lexica, ontologies etc.)
- language description (including grammars, language models etc.)
- tool / service (including basic processing tools, applications, web services etc. required for processing data resources)
- evaluation package[2] (for packages of datasets, tools and metrics used for evaluation purposes).

Central to the description of the LRs in the META-SHARE context is also the ***mediaType*** element, which specifies the form/physical medium of the resource. The notion of medium is preferred over the written/spoken/multimodal distinction, as it has clearer semantics and

2 Metadata for these types are not included in this version.

allows us to view LRs as a set of modules, each of which can be described through a distinctive set of features. Thus, the following *mediaType* values are foreseen:

- text,
- audio
- image[3]
- video[3]
- sensorimotor[3].

A resource may consist of parts belonging to different types of media: for instance, a multimodal corpus includes a video part (moving image), an audio part (dialogues) and a text part (subtitles and/or transcription of the dialogues); a multimedia lexicon includes the text part, but also a video and/or an audio part; a sign language resource is also a good example for a resource with various media types. Similarly, tools can be applied to resources of different media types: e.g. a tool can be used both for video and for audio files. Thus, for each part of the resource, the respective feature set (components and elements) should be used: e.g. for a spoken corpus and its transcriptions, the audio feature set will be used for the audio part and the text feature set for the transcribed part.

3.4 Basic contents & structure of the model

The core of the model is the *Resource* component (Figure 2), which contains all the information relevant for the description of a LR. It subsumes components and elements that combine together to provide this description. A broad distinction can be made between the "administrative" components, which are common to all LRs, and the components that are idiosyncratic to a specific LR type.

The *ContentInfo* component groups together information on the contents of the resource, including the two basic elements that are used for the categorisation and further description of the resource, i.e. *resourceType* and *mediaType*. Each one of the values of these two elements gives rise to a new component, respectively:

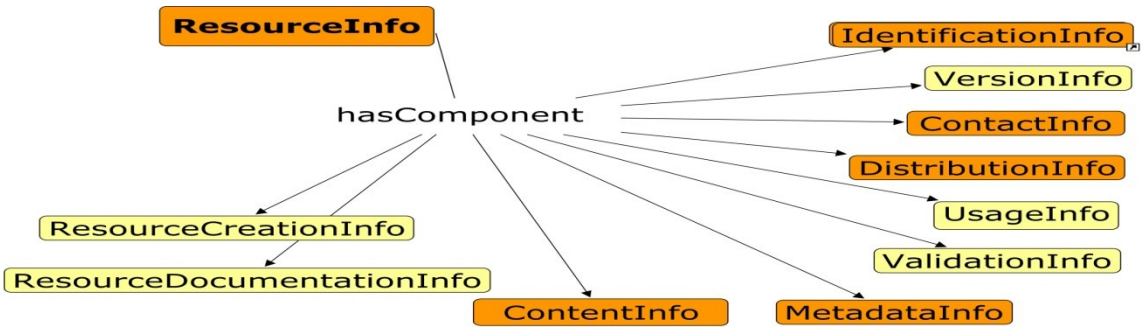

Figure 2: Common components for all LRs
(N.B.: Orange = Mandatory, Yellow = Recommended/Optional)

- *CorpusInfo*, *LexicalConceptualResourceInfo*, *LanguageDescriptionInfo*, *ToolServiceInfo* and *EvaluationPackageInfo* include information specific to each LR type
- *TextInfo*, *AudioInfo*, *VideoInfo*, *ImageInfo* and *SensorimotorInfo* provide information depending on the media type of a resource.

A set of five components enjoy a "special" status, in the sense that they can be attached to various components performing different roles, namely *PersonInfo*, *OrganizationInfo*, *CommunicationInfo*, *ProjectInfo* and *SizeInfo*. For instance, *SizeInfo* can be used either for the size of a whole resource or, in combination with another component, to describe the size of parts of the resource (e.g. per domain, per language etc.); *PersonInfo* is used for contact persons, resource creators, license signatories, annotators of a corpus etc.

4 Structure of the document

The following paragraphs of the deliverable present in detail the model[3]. More specifically, the first section includes the five "special" components of the model, followed by the components that are common to all LR types, and then the resource-type components in the following order: corpora, lexical/conceptual resources, tools/services and, finally, language descriptions.

For each **component** the following information is provided:

- definition: a short statement explaining the semantics of the component inside META-SHARE;
- type: it typically takes the value "component"; the value "special status component" is used for elements that are typed as one of the five special status components;
- elements: the set of elements/components included in the component, with a hyperlink to the explanation of the element/component itself; for each element, further information is provided as to its status (*mandatory*, *condition-dependent mandatory*, *recommended*[4], *optional*) and repeatability (*1* for non-repeatable vs. *unbounded* for

3 In this version of the deliverable, only text and audio corpora, lexical/conceptual resources (only text part), tools/services (only for audio and text) and language descriptions (grammars) are presented.

4 Recommended components and elements are implemented as optional in the current version of the editor.

repeatable ones); if the element/component is condition-dependent, the specific condition is mentioned

- component: used instead of "elements" for the special status components.

For **elements**, the accompanying information includes:

- definition: a short statement explaining its semantics in the META-SHARE context
- type: with values
    - string: free text
    - integer
    - boolean: yes/no
    - myString: free text in any language (the "lang" attribute must be used to specify the language of the text)[5]
    - emailAddress: pattern of email addresses
    - date: date, to be written according to the ISO-8601 standard
    - httpURI: pattern of url's
    - myStringURI: either free text in any language or pattern of url
    - closed controlled vocabulary: the value must be selected from a list of values contained in a controlled vocabulary
    - open controlled vocabulary: the value can be selected from a list of values contained in a controlled vocabulary, but users are also allowed to enter their own values[6]
- value space: reference to the controlled vocabulary; where possible, widely used (best practice) or standardized controlled vocabularies are preferred
- values: if the controlled vocabulary is specific to META-SHARE, the set of values are listed together with definitions where necessary
- examples
- DCLINK: the name of the corresponding element of the Dublin Core schema, provided for mapping purposes

5 Not yet implemented in the editor.

6 In the current version of the editor, users can select the value "other" but not add their own values.

- ISOcatLINK: the name of the corresponding element of the ISOcat DCR[7].

In certain cases, specific components may be re-used at different places of the schema, with a restricted set of elements; e.g. *TextInfo* for Lexical/Conceptual Resources does not include the *AnnotationInfo*, *TextCreationInfo* and *TextClassification* components which are normally used for the description of text corpora. These cases are marked as such.

5 Special status components

5.1 PersonInfo

definition	Used to group together information relevant to persons; to be used mainly for contact persons, resource creators, validators etc. for whom personal data (at least an email) can be provided
type	component
elements	surname *Status: optional* *Repeatability: 1* givenName *Status: optional* *Repeatability: 1* position *Status: optional* *Repeatability: 1* CommunicationInfo *Status: mandatory* *Repeatability: 1* affiliation *Status: optional* *Repeatability: 1*

5.1.1 surname

definition	Surname (family name) of a person
type	myString
value space	
values	

7 In the current version of the deliverable, it's left empty.

examples	Smith;von Kamp;de Gruyter
DCLINK	
ISOcatLINK	

5.1.2 givenName

definition	Given name (first name) of a person; initials can also be used
type	myString
value space	
values	
examples	John;John Jr.;J.;John K.;J.K.
DCLINK	
ISOcatLINK	

5.1.3 position

definition	Position of a person if affiliated to an organization (e.g. director, president, head of unit, etc.)
type	myString
value space	
values	
examples	director; president of the organization
DCLINK	
ISOcatLINK	

5.1.4 affiliation

definition	Groups information on organization to whome the person is affiliated
type	special status component
component	OrganizationInfo

5.2 OrganizationInfo

definition	Used to group together information on organizations
type	component
elements	organizationName *Status: mandatory*

	Repeatability: 1 organizationShortName *Status: optional* *Repeatability: 1* departmentName *Status: optional* *Repeatability: 1* CommunicationInfo *Status: mandatory* *Repeatability: 1*

5.2.1 organizationName

definition	Full name of an organization
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

5.2.2 organizationShortName

definition	Short name (abbreviation, acronym etc.) used for an organization
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

5.2.3 departmentName

definition	Name of the department/unit (e.g. specific university faculty/department, department/unit of a research organization or private company etc.)
type	myString
value space	
values	

examples	
DCLINK	
ISOcatLINK	

5.3 CommunicationInfo

definition	Groups information on communication details (address etc.)[8]
type	component
elements	address *Status: optional* *Repeatability: 1* zipCode *Status: optional* *Repeatability: 1* city *Status: optional* *Repeatability: 1* country *Status: optional* *Repeatability: 1* telephoneNumber *Status: optional* *Repeatability: unbounded* faxNumber *Status: optional* *Repeatability: unbounded* email *Status: mandatory* *Repeatability: unbounded* url *Status: optional* *Repeatability: unbounded*

8 The element *region* is added in the next version.

5.3.1 address

definition	Postal address of a person or organization - street and number
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

5.3.2 zipCode

definition	Postal address of a person or organization - zip code
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

5.3.3 city

definition	Postal address of a person or organization - city/town/village
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

5.3.4 country

definition	Postal address of a person or organization - country; list of values to be taken from ISO 3166
type	controlled vocabulary
value space	ISO 3166, Codes for the representation of names of countries and their subdivisions
values	
examples	
DCLINK	

ISOcatLINK	

5.3.5 telephoneNumber

definition	The telephone number of a person or an organization; recommended format: +_international code_city code_number
type	tel
value space	number
values	
examples	
DCLINK	
ISOcatLINK	

5.3.6 faxNumber

definition	The Fax number of a person or an organization; recommended format: +_international code_city code_number
type	tel
value space	number
values	
examples	
DCLINK	
ISOcatLINK	

5.3.7 email

definition	The email address of a person or an organization; if it is unknown, please use the "unknown@example.com" invalid address
type	emailAddress
value space	
values	
examples	
DCLINK	
ISOcatLINK	

5.3.8 url

definition	URL of a person or organization
type	httpURI
value space	
values	
examples	
DCLINK	
ISOcatLINK	

5.4 ProjectInfo

definition	Groups together information related to a project (either funded by external funds or by own funds)
type	component
elements	projectID *Status: optional* *Repeatability: 1* projectName *Status: mandatory* *Repeatability: 1* projectShortName *Status: optional* *Repeatability: 1* fundingType *Status: mandatory* *Repeatability: unbounded* fundingCountry *Status: recommended* *Repeatability: unbounded* funder *Status: recommended* *Repeatability: unbounded* projectStartDate *Status: optional* *Repeatability: 1* projectEndDate *Status: optional* *Repeatability: 1*

	url *Status: optional* *Repeatability: unbounded*

5.4.1 projectID

definition	A unique identifier identifying the project
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

5.4.2 projectName

definition	The full name of the project that led to the creation of the resource
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

5.4.3 projectShortName

definition	A short name or abbreviation of the project that led to the creation of the resource
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

5.4.4 fundingType

definition	Type of funding of the project

type	open controlled vocabulary	
value space	MS-fundingType	
values		value definition other ownFunds nationalFunds euFunds
examples		
DCLINK		
ISOcatLINK		

5.4.5 fundingCountry

definition	Funding country, in case of national funding; use ISO3166
type	controlled vocabulary
value space	ISO 3166, Codes for the representation of names of countries and their subdivisions
values	
examples	
DCLINK	
ISOcatLINK	

5.4.6 funder

definition	Name of the funder of the project i.e. a private organization, company etc.
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

5.4.7 projectStartDate

definition	project starting date
type	date
value space	
values	

examples	
DCLINK	
ISOcatLINK	

5.4.8 projectEndDate

definition	project end date
type	date
value space	
values	
examples	
DCLINK	
ISOcatLINK	

5.4.9 url

definition	url of the project
type	httpURI
value space	
values	
examples	
DCLINK	
ISOcatLINK	

5.5 SizeInfo

definition	SizeInfo Element
type	component
elements	size *Status: mandatory* *Repeatability: 1* sizeUnitMultiplier[9] *Status: mandatory* *Repeatability: 1* sizeUnit *Status: mandatory*

9 *sizeUnitMultiplier* is dropped in the next version.

	Repeatability: 1

5.5.1 size

definition	The size of the resource with regard to the SizeUnit measurement in form of a number
type	number
value space	
values	
examples	
DCLINK	
ISOcatLINK	

5.5.2 sizeUnit

definition	Specification of the unit of size that is used when providing information on the size of a resource	
type	open controlled vocabulary	
value space	MS-sizeUnit[10]	
values		Value definition terms entries turns utterances articles files items seconds elements units minutes hours texts sentences bytes tokens words keywords idiomaticExpressions neologisms multiWordUnits expressions synsets

10 Values have been added in the next version; note also that for language descriptions, it contains an extra value, namely "rules".

	classes concepts lexicalTypes phoneticUnits syntacticUnits semanticUnits predicates phonemes diphones T-HPairs syllables other
examples	
DCLINK	
ISOcatLINK	

5.5.3 sizeUnitMultiplier

definition		
type	closed controlled vocabulary	
value space	MS-sizeUnit	
values	value	definition
	unit	
	tera	
	mega	
	kilo	
	hundred	
examples		
DCLINK		
ISOcatLINK		

6 Resource – Common components

definition	Used to group together all information required for the description of language resources[11]
elements	IdentificationInfo

11 A new component, *RelationInfo*, is added in the next version.

	Status: mandatory *Repeatability: 1* contactPerson *Status: mandatory* *Repeatability: unbounded* VersionInfo *Status: recommended* *Repeatability: 1* DistributionInfo *Status: mandatory* *Repeatability: 1* ValidationInfo *Status: recommended* *Condition: resourceType=corpus or lexicalConceptualResource* *Repeatability: unbounded* ResourceCreationInfo *Status: recommended* *Repeatability: 1* UsageInfo *Status: recommended* *Repeatability: 1* MetadataInfo *Status: mandatory* *Repeatability: 1* ResourceDocumentationInfo *Status: recommended* *Repeatability: unbounded* ContentInfo *Status: mandatory* *Repeatability: 1*

6.1 IdentificationInfo

definition	Groups together information needed to identify the resource
type	component
elements	resourceName *Status: mandatory*

	Repeatability: 1 resourceShortName *Status: optional* *Repeatability: 1* pid[12] *Status: mandatory* *Repeatability: 1* url *Status: recommended* *Repeatability: unbounded* identifier *Status: optional* *Repeatability: unbounded*

6.1.1 resourceName

definition	The name by which the resource is known; if there are two forms (a full and a short name), please use this for the full name and "resourceShortName" for the short alternative
type	myString
value space	
values	
examples	British National Corpus; Penn Treebank; EXMARaLDA
DCLINK	dc:title
ISOcatLINK	

6.1.2 resourceShortName

definition	Short form (abbreviation, acronym etc.) used to identify the language resource
type	myString
value space	
values	
examples	BNC; PTB
DCLINK	
ISOcatLINK	

12 A new element, *metaShareId*, is introduced in the next version while *pid* is implemented as a type of *identifier*.

6.1.3 pid

definition	persistent identifier
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

6.1.4 url

definition	The URL of a web site providing information on the language resource (e.g. description, on creation, samples, contact info, info on modes of access etc.)
type	httpURI
value space	
values	
examples	
DCLINK	
ISOcatLINK	

6.1.5 identifier

definition	Identifier used for the resource, such as the one from the ELRA or LDC catalogues, the pid or an internal identifier used by the resource provider; the attribute "type" is obligatorily used for further specification[13]
type	free text
value space	
examples	
DCLINK	dc:identifier
ISOcatLINK	

6.2 contactPerson[14]

definition	Used for giving information on contact person for the resource
type	special status component
component	PersonInfo

13 The attribute "type" is not included in the current XSD's and the editor.

14 It corresponds to the Person tab of the editor

DCLINK	
ISOcatLINK	

6.3 VersionInfo

definition	Groups information on version/release of the resource
type	component
elements	version *Status: mandatory* *Repeatability: 1* lastDateUpdated *Status: optional* *Repeatability: 1* revision *Status: optional* *Repeatability: 1* updateFrequency *Status: optional* *Repeatability: 1*

6.3.1 version

definition	Any string (usu. number) that identifies the version of a metadata description, a resource or a tool/web service
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

6.3.2 lastDateUpdated

definition	Date of last updating of the version/release
type	date
value space	
values	
examples	
DCLINK	dc:date
ISOcatLINK	

6.3.3 revision

definition	Account of the revisions made from previous versions of the resource; this could also be a link to a document
value space	date
values	
examples	
DCLINK	
ISOcatLINK	

6.3.4 updateFrequency

definition	The frequency with which the resource or the tool/service is updated
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

6.4 DistributionInfo

definition	Groups information on the distribution of the resource[15]
type	component
elements	availability *Status: mandatory* *Repeatability: 1*

15 A new element, *distributionRightsHolder*, is added in the next version.

	iprHolder *Status: optional* *Repeatability: unbounded* LicenseInfo *Status: mandatory*[16] *Condition: availability="available-restrictedUse" or "available-unrestrictedUse"* *Repeatability: unbounded*

6.4.1 availability

definition	Availability status of the resource; restrictionsOfUse can be further used to indicate the specific terms of availability	
type	closed controlled vocabulary	
value space	MS-availability	
values		value definition available-unrestrictedUse available-restrictedUse notAvailable underNegotiation
examples		
DCLINK	dc:rights	
ISOcatLINK		

6.4.2 iprHolder

definition	Groups information on person/organization who holds the IPR (could be different from creator and distributor)
type	special status component
elements	choice between PersonInfo OrganizationInfo *Status: optional* *Repeatability: unbounded*

16 Condition-dependent mandatory in the next version

6.4.3 LicenseInfo

definition	Groups information on licenses for the resource; can be repeated to allow for different modes of access and restrictions of use (e.g. free for academic use, on-a-fee basis for commercial use, download of a sample for free use etc.)
type	component
elements	license *Status: mandatory* *Repeatability: 1* restrictionsOfUse *Status: condition-dependent mandatory* *Condition: availability="available-restrictedUse"* *Repeatability: 1*[17] price *Status: optional* *Repeatability: 1* distributionAccessMedium *Status: mandatory* *Repeatability: 1*[18] executionLocation *Status: condition-dependent mandatory* *Condition: distributionMedium="accessibleThroughInterface" or "webExecutable"* *Repeatability: unbounded* downloadLocation *Status: condition-dependent mandatory* *Condition: availability="available-restrictedUse"* *Repeatability: unbounded* availabilityStartDate *Status: optional* *Repeatability: 1* availabilityEndDate *Status: optional* *Repeatability: 1* distributor *Status: recommended*

17 unbounded in next version

18 unbounded in next version

	Repeatability: unbounded licenseSignatory *Status: recommended* *Repeatability: unbounded*

1.1.1.1 license

definition	The license of use for the resource[19]	
type	open controlled vocabulary	
value space	MS-license	
values		Value definition AGPL LGPL CC_BY-NC-ND CC_BY-NC-SA CC_BY-NC CC_BY-ND CC_BY-SA CC_BY MSCommons ELRA_EVALUATION ELRA_VAR ELRA_END_USER proprietary CC CLARIN_PUB CLARIN_ACA-NC CC_BY-SA_3.0 LGPLv3 CLARIN_ACA CLARIN_RES Princeton_Wordnet GPL GeneralLicenseGrant GFDL CC_BY-NC-SA_3.0 ApacheLicense_V2.0 BSD-style other underNegotiation
examples		
DCLINK	dc:rights	
ISOcatLINK		

19 Changes and addition of values in next version

1.1.1.2 restrictionsOfUse

definition	The restrictions imposed by the type of the license	
type	open controlled voacabulary	
value space	MS-restrictionsOfUse	
values		Value definition other noModifications informResourceOwner redeposit onlyMSmembers academic-nonCommercialUse evaluationUse commercialUse attribution shareAlike noDerivatives
examples		
DCLINK	dc:rights	
ISOcatLINK		

1.1.1.3 price

definition	The costs that are required to access the resource, a fragment of the resource or to use the tool/service
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

1.1.1.4 distributionAccessMedium

definition	The medium (channel) used for delivery or providing access to the resource	
type	open controlled vocabulary	
value space	MS-distributionAccessMedium	
values		value definition webExecutable other paperCopy hardDisk bluRay DVD-R CD-ROM

	downloadable accessibleThroughInterface
examples	
DCLINK	
ISOcatLINK	

1.1.1.5 executionLocation

definition	Where the service providing access to a resource is being executed
type	httpURI
value space	
values	
examples	
DCLINK	
ISOcatLINK	

1.1.1.6 downloadLocation

definition	where the resource can be downloaded from
type	httpURI
value space	
values	
examples	
DCLINK	
ISOcatLINK	

1.1.1.7 availabilityStartDate

definition	Start date of availability of a resource
type	
value space	date
values	
examples	
DCLINK	dc:date
ISOcatLINK	

1.1.1.8 availabilityEndDate

definition	End date of availability of a resource
type	
value space	date

values	
examples	
DCLINK	dc:date
ISOcatLINK	

1.1.1.9 distributor

definition	Groups information on person/organization distributing the resource (could be different from creator)
type	special status component
component	choice between PersonInfo OrganizationInfo *Status: optional* *Repeatability: unbounded*
DCLINK	~dc:publisher
ISOcatLINK	

1.1.1.10 licenseSignatory

definition	Groups information on person who is legally responsible to sign the license (could be different from creator, distributor or rightsholder)
type	special status component
component	PersonInfo

6.5 ValidationInfo

definition	Groups information on validation of a resource; it can be repeated to allow for different validations (e.g. formal validation of the whole resource; content validation of one part of the resource etc.). For tools please use the EvaluationInfo component included in the ToolServiceInfo instead.
type	component
elements	validated *Status: mandatory* *Repeatability: 1* validationType *Status: optional* *Repeatability: 1*

	validationMode *Status: optional* *Repeatability: 1* validationModeDetails *Status: optional* *Repeatability: 1* validationReport *Status: optional* *Repeatability: 1* validationTool *Status: optional* *Repeatability: 1* validationExtent *Status: optional* *Repeatability: 1* validationExtentDetails *Status: optional* *Repeatability: 1* validator *Status: optional* *Repeatability: unbounded* sizePerValidationType *Status: optional* *Repeatability: 1*

6.5.1 validated

definition	The validation status of the resource; please, use "yes" even for partially validated resources
type	boolean
value space	
values	
examples	
DCLINK	
ISOcatLINK	

6.5.2 validationType

definition	The validation type applied	
type	closed controlled vocabulary	
value space	MS-validationType	
values	Value formal content	definition
examples		
DCLINK		
ISOcatLINK		

6.5.3 validationMode

definition	The validation methodology applied	
type	closed controlled vocabulary	
value space	MS-validationMode	
values	value manual automatic mixed interactive	definition
examples		
DCLINK		
ISOcatLINK		

6.5.4 validationModeDetails

definition	Textual field for additional information on validation
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

6.5.5 validationReport

definition	Short account of the validation details or link to the validation report
type	myStringURI
value space	
values	
examples	
DCLINK	
ISOcatLINK	

6.5.6 validationTool

definition	Name of the tool used for the validation of the resource
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

6.5.7 validationExtent

definition	The resource coverage in terms of validated data	
type	closed controlled vocabulary	
value space	MS-validationExtent	
values	value	definition
	full	
	partial	
examples		
DCLINK		
ISOcatLINK		

6.5.8 validationExtentDetails

definition	information on size or other details of partially validated data; to be used if only part of the resource has been validated and as an alternative to SizeInfo if the validated part cannot be counted otherwise
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

6.5.9 validator

definition	Groups information on person/organization who validated the resource
type	special status component
component	choice between PersonInfo OrganizationInfo

6.5.10 sizePerValidationType

definition	For information on size of the validated part of a resource
type	linked omponent
component	SizeInfo

6.6 ResourceCreationInfo

definition	Groups information on creation procedure, tools etc. of a resource
type	component
elements	creationStartDate *Status: recommended* *Repeatability: 1* creationEndDate *Status: recommended* *Repeatability: 1* resourceCreator *Status: recommended*

	Repeatability: unbounded FundingInfo *Status: optional* *Repeatability: unbounded*

6.6.1 creationStartDate

definition	The date in which the creation process was started
type	date
value space	
values	
examples	
DCLINK	dc:date
ISOcatLINK	

6.6.2 creationEndDate

definition	The date in which the creation process was completed
type	date
value space	
values	
examples	
DCLINK	dc:date
ISOcatLINK	

6.6.3 resourceCreator

definition	Person or organization that has created the resource
type	special status component
component	choice between PersonInfo OrganizationInfo
DCLINK	dc:creator
ISOcatLINK	

6.6.4 FundingInfo[20]

definition	Groups information on all projects that have funded the resource; repeat for each project; internal funding of a resource is also thought of as a kind of project
type	special status component
elements	ProjectInfo *Status: Mandatory* *Repeatability: unbounded*

6.7 UsageInfo

definition	Groups information on usage (both intended and actual use, i.e. how it has already been used)
type	component
elements	accessTool *Status: optional* *Repeatability: unbounded* toolAssociatedWith *Status: optional* *Repeatability: unbounded* ForeseenUseInfo *Status: recommended* *Repeatability: 1*[21] ActualUseInfo *Status: recommended* *Repeatability: unbounded*

6.7.1 accessTool

definition	Tool used to access/view a resource (e.g. a corpus workbench); alternative to the relation hasAsAccessTool
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

20 To be replaced by the element *fundingProject* in the next version.

21 Repeatable in the next version

6.7.2 toolAssociatedWith

definition	Indicates another resource that the resource described uses for its operation (e.g. a tagger using a lexicon as one of its components, a corpus used for training a tool)
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

6.7.3 ForeseenUseInfo

definition	Groups information on the use for which the resource was created
type	component
elements	foreseenUse *Status: recommended*[22] *Repeatability: 1* useNLPSpecific *Status: recommended* *Repeatability: unbounded*

1.1.1.11foreseenUse

definition	Classification of the foreseen use of the resource (why it was made); if foreseenUse=nlpApplications, specify all nlp applications in the same component; if foreseenUse is both humanUse and nlpApplications, then the component must be repeated, one for humanUse and one for nlpApplications	
type	closed controlled vocabulary	
value space	MS-foreseenUse	
values		value definition humanUse nlpApplications

22 Obligatory in next version.

examples	
DCLINK	
ISOcatLINK	

1.1.1.12 useNLPSpecific

definition	Specific NLP application for which the resource was created or where it has been used, e.g. speech synthesis, testbed, word disambiguation etc.)[23]	
type	open controlled vocabulary	
value space	MS-useNLPSpecific	
values	value acquisition avatarSynthesis automaticPersonRecognition automaticSpeechRecognition automaticTextGeneration automaticTextSummarization bilingualLexiconInduction contradictionDetection coreferenceResolution derivationalMorphologicalAnalysis discourseAnalysis documentClassification emotionGeneration emotionRecognition entityMentionRecognition eventExtraction expressionRecognition faceRecognition faceVerification humanoidAgentSynthesis informationExtraction informationRetrieval intra-documentCoreferenceResolution knowledgeDiscovery knowledgeRepresentation languageIdentification languageModelling languageModelsTraining lemmatization lexiconAccess lexiconAcquisitionFromCorproa lexiconEnhancement lexiconExtractionFromLexica lexiconFormatConversion lexiconMerging lexiconVisualization	definition

23 Changes and deletions of values in next version

	linguisticResearch lipTrackingAnalysis machineTranslation morphologicalAnalysis morphosyntacticTagging multimediaDevelopment multimediaDocumentProcessing namedEntityRecognition naturalLanguageGeneration naturalLanguageUnderstanding opinionMining other parsing personIdentification persuasiveExpressionMining posTagging qualitativeAnalysis questionAnswering readingAndWritingAidApplications semanticRoleLabelling semanticWeb sentimentAnalysis signLanguageGeneration signLanguageRecognition speakerIdentification speakerVerification speechAnalysis speechAssistedVideoControl speechLipsCorrelationAnalysis speechRecognition speechSynthesis speechToSpeechTranslation speechUnderstanding speechVerification spellChecking spokenDialogueSystems summarisation talkingHeadSynthesis temporalExpressionRecognition terminologyExtraction textCategorisation textMining texToSpeechSynthesis textualEntailment topicDetection_Tracking userAuthentication visualSceneUnderstanding voiceControl webServices wordSenseDisambiguation
examples	
DCLINK	
ISOcatLINK	

6.7.4 ActualUseInfo

definition	Groups information on how the resource has already been used
type	component
elements	actualUse *Status: mandatory* *Repeatability: 1* useNLPSpecific *Status: recommended* *Repeatability: unbounded* publication *Status: optional* *Repeatability: unbounded* outcome *Status: optional* *Repeatability: unbounded* actualUseDetails *Status: optional* *Repeatability: 1* usageProject *Status: optional* *Repeatability: unbounded*

1.1.1.13 actualUse

definition	Classification of the use of the resource	
type	closed controlled vocabulary	
value space	MS-actualUse	
values	value humanUse nlpApplications	definition
examples		
DCLINK		
ISOcatLINK		

1.1.1.14 publication

definition	Titles of research papers documenting the usage of a resource
type	myStringURI
value space	
values	
examples	
DCLINK	
ISOcatLINK	

1.1.1.15 outcome

definition	Outcome/product of the resource (e.g. terminological list as the result of term extraction corpus); alternative to the relation hasOutcome
type	myStringURI
value space	
values	
examples	
DCLINK	
ISOcatLINK	

1.1.1.16 actualUseDetails

definition	Description of the usage of the resource
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

1.1.1.17 usageProject

definition	Information on the project in which the resource has been used
type	special status component
component	ProjectInfo

6.8 MetadataInfo

definition	Groups information on the metadata record itself
type	component
elements	source

	Status: optional *Repeatability: unbounded* originalMetadataSchema *Status: optional* *Repeatability: unbounded* harvestingDate [24] *Status: optional* *Repeatability: 1* metadataCreationDate *Status: mandatory* *Repeatability: 1* originalMetadataLink *Status: optional* *Repeatability: 1* metadataCreator *Status: optional* *Repeatability: unbounded* metadataLastDateUpdated *Status: optional* *Repeatability: 1* metadataLanguage *Status: optional* *Repeatability: unbounded* revision *Status: optional* *Repeatability: 1*

6.8.1 source

definition	Catalogue/Repository from which the harvesting was made (CLARIN, OLAC, META,...); open issue: value to be assigned automatically depending on where the metadata has been harvested from
type	open controlled vocabulary[25]
value space	MS-source

24 To be removed in next version.

25 Replaced by "string" in next version.

values	Value definition CLARIN OLAC METASHARE E LREmap CESAR META-NORD METANET4U PANACEA other
examples	
DCLINK	
ISOcatLINK	

6.8.2 originalMetadataSchema

definition	metadata schema originally used for the resource
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

6.8.3 harvestingDate

definition	The date of harvesting of this metadata description for records that have been harvested
type	date
value space	
values	
examples	
DCLINK	dc:date
ISOcatLINK	

6.8.4 metadataCreationDate

definition	The date of creation of this metadata description; if the metadata recorded has

	resulted from conversion after harvesting, this is also concidered the metadataCreationDate
type	date
value space	
values	
examples	
DCLINK	dc:date
ISOcatLINK	

6.8.5 originalMetadataLink

definition	Link to the metadata of the original source; to be automatically assigned in harvesting process
type	httpURI
value space	
values	
examples	
DCLINK	
ISOcatLINK	

6.8.6 metadataCreator

definition	person that created the metadata if inserted by META-SHARE editor; to be automatically assigned
type	special status component
component	PersonInfo
DCLINK	~dc:contributor
ISOcatLINK	

6.8.7 metadataLastDateUpdated

definition	Date of last updating of the metadata record; to be automatically assigned each time the record is updated
type	date
value space	
values	
examples	
DCLINK	dc:date
ISOcatLINK	

6.8.8 metadataLanguage

definition	An identifier of the language in which the metadata description was written; for current version, default value should be English
type	controlled vocabulary
value space	ISO 639-3:2007, Codes for the representation of names of languages – *Part 3:* Alpha-3 code for comprehensive coverage of languages[26]
values	
examples	
DCLINK	
ISOcatLINK	

6.8.9 revision

definition	account of the revisions of the metadata or link to a document with revisions
type	myStringURI
value space	
values	
examples	
DCLINK	
ISOcatLINK	

6.9 ResourceDocumentationInfo

definition	Groups together information on (papers, etc.) describing the resource
type	component
elements	publication *Status: optional* *Repeatability: unbounded* toolDocumentationType *Status: optional* *Condition: resourceType=toolService* *Repeatability: unbounded* samplesLocation *Status: recommended*

26 To be changed to IETF BCP47 in v2

	Repeatability: unbounded

6.9.1 publication

definition	Paper, manual etc. for the resource; alternative to the relation isDocumentedIn
type	myStringURI
value space	
values	
examples	
DCLINK	
ISOcatLINK	

6.9.2 toolDocumentationType

definition	for tools only, type of documentation	
type	open controlled vocabulary	
value space	MS-toolDocumentationType	
values		value definition online manual helpFunctions none other
examples		
DCLINK		
ISOcatLINK		

6.9.3 samplesLocation

definition	URL with samples of the resource or, in the case of tools, of samples of the output
type	httpURI
value space	
values	
examples	
DCLINK	
ISOcatLINK	

6.10 ContentInfo

definition	Groups together information on contents of the resource
type	component
elements	description *Status: mandatory* *Repeatability: 1* resourceType *Status: mandatory* *Repeatability: 1* mediaType *Status: condition-dependent mandatory* *Condition: resourceType="corpus" or "lexicalConceptualResource" or "languageDescription"*[27] *Repeatability: unbounded*

6.10.1 description

definition	Description of the resource in prose
type	myString
value space	
values	
examples	
DCLINK	dc:description
ISOcatLINK	

6.10.2 resourceType

definition	type of the resource	
type	closed controlled vocabulary	
value space	MS-resourceType	
values	value	definition
	corpus	for text, speech and m
	lexicalConceptualResource	includes lexica, or

27 The condition will be implemented in the next version

		word lists etc.
	languageDescription	covers language mod
	technologyToolService [28]	used for tools, syster etc.
	evaluationPackage	used for datasets and evaluation
examples		
DCLINK	dc:type	
ISOcatLINK		

6.10.3 mediaType

definition	Specification of the media type of the resource; can be multiple if the resource is a multimodal set	
type	closed controlled vocabulary	
value space	MS-mediaType	
values	value text audio video image sensorimoto r	definition
examples		
DCLINK	dc:type	
ISOcatLINK		

7 Corpora

7.1 Resource

definition	extension of Resource for Corpora
type	component
elements	**Resource – common components** IdentificationInfo

28 *toolService* in the next version

	contactPerson VersionInfo DistributionInfo ValidationInfo ResourceCreationInfo UsageInfo MetadataInfo ResourceDocumentationInfo ContentInfo [N.B.: resourceType=corpus] **Additional components**[29]

29 CorpusInfo is not included in the current version of the editor and corresponding XSD's.

	TextInfo *Status: Condition-dependent mandatory* *Condition: mediaType=text* *Repeatability: unbounded* AudioInfo *Status: Condition-dependent mandatory* *Condition: mediaType=audio* *Repeatability: unbounded*

7.2 TextInfo

definition	Groups together information on the text component of a text resource
type	component
elements	LingualityInfo *Status: mandatory* *Repeatability: 1* LanguageInfo *Status: mandatory* *Repeatability: unbounded* TextCreationInfo *Status: recommended* *Repeatability: 1* SizeInfo *Status: mandatory*[30] *Repeatability: 1* TextFormatInfo *Status: recommended* *Repeatability: unbounded* CharacterEncodingInfo *Status: recommended* *Repeatability: unbounded* DomainInfo *Status: recommended* *Repeatability: unbounded* TimeCoverageInfo *Status: recommended*

30 Recommended in next version

	Repeatability: unbounded GeographicCoverageInfo *Status: recommended* *Repeatability: unbounded* TextClassificationInfo *Status: recommended* *Repeatability: unbounded* AnnotationInfo [31] *Status: mandatory* *Repeatability: unbounded*

7.2.1 LingualityInfo

definition	Groups information on linguality and modalities
type	component
element	lingualityType *Status: mandatory* *Repeatability: 1* multilingualityType *Status: optional* *Repeatability: 1* multilingualityTypeDetails *Status: optional* *Repeatability: 1* modalityType *Status: optional* *Repeatability: 1*

1.1.1.18 lingualityType

definition	Indicates whether the resource includes one, two ore more languages	
type	closed controlled vocabulary	
value space	MS-lingualityType	
values		value definition monolingual

31 Recommended in next version.

	bilingual multilingual
examples	
DCLINK	
ISOcatLINK	

1.1.1.19 multilingualityType

definition	Indicates whether the corpus is parallel or comparable or mixed
type	open controlled vocabulary[32]
values	MS-multilingualityType
value space	Value definition comparable parallel other
examples	
DCLINK	
ISOcatLINK	

1.1.1.20 multilingualityTypeDetails

definition	Free text statement giving further information on multilinguality of a resource (e.g. translation of a text, direction of translation, …)
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

1.1.1.21 modalityType

definition	information on modality
type	string[33]
value space	
values	
examples	

32 added value in next version

33 open controlled vocabulary in next version

DCLINK	
ISOcatLINK	

7.2.2 LanguageInfo

definition	extension of LanguageInfo for texts
type	component
elements	languageCoding [34] *Status: mandatory* *Repeatability: 1* languageId *Status: mandatory* *Repeatability: 1* languageName *Status: mandatory* *Repeatability: 1* languageScript *Status: optional* *Repeatability: 1* sizePerLanguage *Status: optional* *Repeatability: unbounded* [for SizeUnit] LanguageVarietyInfo *Status: optional* *Repeatability: unbounded*

1.1.1.22 languageCoding

definition	Designation of the standard used to code the name of the languages
type	controlled vocabulary
value space	ISO 639-3:2007, Codes for the representation of names of languages – *Part 3:* Alpha-3 code for comprehensive coverage of languages[35]
values	
examples	

34 languageCoding, languageId and languageScript do not appear in the editor.

35 To be replaced by IETF BCP47; related elements (*languageName, languageScript*) will be updated subject to this change.

DCLINK	
ISOcatLINK	

1.1.1.23 languageId

definition	Identifier of the language that is included in the resource or supported by the tool/service
type	controlled vocabulary
value space	ISO 639-3:2007, Codes for the representation of names of languages – *Part 3:* Alpha-3 code for comprehensive coverage of languages
values	
examples	
DCLINK	dc:language
ISOcatLINK	

1.1.1.24 languageName

definition	A human understandable name of the language that is used in the resource or supported by the tool/service
type	controlled vocabulary
value space	ISO 639-3:2007, Codes for the representation of names of languages – *Part 3:* Alpha-3 code for comprehensive coverage of languages
values	
examples	
DCLINK	dc:language
ISOcatLINK	

1.1.1.25 languageScript

definition	Indication of the writing system used to represent the language in form of a four letter code as it is defined in ISO-15924
type	controlled vocabulary
value space	ISO-15924, Codes for the representation of names of scripts
values	
examples	
DCLINK	
ISOcatLINK	

1.1.1.26 sizePerLanguage

definition	used to provide info on size per language component
type	special status component
component	SizeInfo

1.1.1.27 LanguageVarietyInfo

definition	Groups information on language varieties of a resource (e.g. dialects); repeated for different language varieties
type	component
elements	languageVarietyType *Status: mandatory* *Repeatability: 1* languageVarietyName *Status: mandatory* *Repeatability: 1* sizePerLanguageVariety *Status: optional* *Repeatability: unbounded* [for SizeUnit]

1.1.1.27.1 languageVarietyType

definition	Type of the language variety	
type	open controlled vocabulary	
value space	MS-languageVarietyType	
values	value	definition
	dialect	
	jargon	
	other	
examples		
DCLINK		
ISOcatLINK		

1.1.1.27.2 languageVarietyName

definition	Name of the language variety that occurs in the resource
type	myString
value space	
values	
examples	

DCLINK	
ISOcatLINK	

1.1.1.27.3 sizePerLanguageVariety

definition	size per language variety of a resource
type	special status component
component	SizeInfo

7.2.3 TextCreationInfo

definition	Groups together information on the raw corpus creation (selection of texts, structural encoding thereof); for annotation, use the Annotation component
type	component
elements	originalSource *Status: recommended* *Repeatability: unbounded* creationMode *Status: recommended* *Repeatability: 1* creationModeDetails *Status: optional* *Repeatability: 1* creationTool *Status: optional* *Repeatability: unbounded*

1.1.1.28 originalSource

definition	Indicates the original resources that were at the base of the creation/derivation process; alternative to the relation hasOriginalSource
type	myString
value space	
values	
examples	
DCLINK	dc:source
ISOcatLINK	

1.1.1.29 creationMode

definition	A first indication as to the mode of creation of the resource	
type	closed controlled vocabulary	
value space	MS-creationMode	
values	value automatic manual mixed interactiv e	definition
examples		
DCLINK		
ISOcatLINK		

1.1.1.30 creationModeDetails

definition	Used to supply more details as to the creation methods and processes of the resource
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

1.1.1.31 creationTool

definition	Indicates the tool with help of which the resource was created; alternative to the preferred relation hasAsCreationTool
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

7.2.4 TextFormatInfo

definition	Groups information on the format(s) of a resource; repeated if parts of the resource are in different formats
type	component

elements	mime-type *Status: mandatory* *Repeatability: unbounded* sizePerTextFormat *Status: optional* *Repeatability: unbounded*

1.1.1.32 mime-type

definition	Specification of the mime-type of the resource which is a formalized specifier for the format included or a mime-type that the tool/service accepts; value to be taken from a subset of the official mime types	
type	controlled vocabulary[36]	
value space		value definition other text/plain application/pdf application/msword text/xml text/sgml text/html text/rtf application/zip
values		
examples		
DCLINK	dc:format	
ISOcatLINK		

1.1.1.33 sizePerTextFormat

definition	used to give info on size of parts with different format
type	special status component
component	SizeInfo

7.2.5 CharacterEncodingInfo

definition	Groups together information on character encoding of the resource; repeated if parts of the resource have different character encodings

36 In the XSD there is no enumeration; in the next version the value space will be a subset of the IANA mime-types (http://www.iana.org/assignments/media-types/index.html).

type	component
elements	characterEncoding *Status: mandatory* *Repeatability: unbounded* characterSet *Status: optional* *Repeatability: unbounded* sizePerCharacterEncoding *Status: optional* *Repeatability: unbounded*

1.1.1.34 characterEncoding

definition	Name of the character encoding used in the resource or accepted by the tool/service	
type	closed controlled vocabulary	
value space	MS-characterEncoding	
values		value definition ISO-8859-1 UTF-8 ISO-2022 other
examples		
DCLINK		
ISOcatLINK		

1.1.1.35 characterSet

definition	The repertoire of characters used in the resource; a range of characters (non-coded character set) or a coded character set as defined in RFC 2050	
type	closed controlled vocabulary	
value space	MS-characterSet	
values		value definition UCS ISO-8859-6 other
examples		
DCLINK		
ISOcatLINK		

1.1.1.36 sizePerCharacterEncoding

annotation	used to give info on size of parts with different character encoding
type	special status component
componen t	SizeInfo

7.2.6 DomainInfo

definitio n	Groups together information on domains of a resource; can be repeated for parts of the resource with distinct domain[37]
type	component
elements	domain

37 The element *conformanceToClassificationScheme* is to be added in the next version.

	Status: mandatory *Repeatability: unbounded* sizePerDomain *Status: optional* *Repeatability: unbounded*

1.1.1.37 domain

definition	Indicates the application domain of the resource or the tool/service	
type	open controlled vocabulary	
value space	MS-domain	
values		value definition sports environment law_politics medicine tourism science banking entertainment literature education business general biomedicine economy other
examples		
DCLINK	dc:subject	
ISOcatLINK		

1.1.1.38 sizePerDomain

definition	size of subpart of a resource per domain
type	special status component
component	SizeInfo

7.2.7 TimeCoverageInfo

definition	Groups together information on time classification of a resource

type	component
elements	timeCoverage *Status: mandatory* *Repeatability: 1* sizePerTimeCoverage *Status: optional* *Repeatability: unbounded*

1.1.1.39 timeCoverage

definition	The time period that the content of a resource is about; DC-Definition: The temporal topic of the resource
type	myString
value space	
values	
examples	
DCLINK	dc:coverage
ISOcatLINK	

1.1.1.40 sizePerTimeCoverage

definition	used to provide info on size per time period of a resource
type	special status component
component	SizeInfo

7.2.8 GeographicCoverageInfo

definition	Groups information on geographic classification of a resource
type	component
elements	geographicCoverage *Status: mandatory* *Repeatability: 1* sizePerGeographicCoverage *Status: optional* *Repeatability: unbounded*

1.1.1.41 geographicCoverage

definition	Geographic region related to the resource; for countries, recommended use of ISO-3166
type	myString
value space	
values	
examples	
DCLINK	dc:coverage
ISOcatLINK	

1.1.1.42 sizePerGeographicCoverage

definition	used to provide info on size per geographically distinct section of a resource
type	special status component
component	SizeInfo

7.2.9 TextClassificationInfo

definition	Groups together information on text type/genre of the resource
type	component
elements	textGenre *Status: optional* *Repeatability: 1* textType *Status: optional* *Repeatability: 1* register *Status: optional* *Repeatability: 1* subject_topic *Status: optional* *Repeatability: 1* conformanceToClassificationScheme *Status: optional* *Repeatability: 1* sizePerTextClassification *Status: optional*

	Repeatability: unbounded

1.1.1.43 textGenre

definition	Genre: The conventionalized discourse or text types of the content of the resource, based on extra-linguistic and internal linguistic criteria	
type	closed controlled vocabulary[38]	
value space	MS-textGenre	
values		value definitio n IPTC OLAC PAROLE LDC ELRA
examples		
DCLINK	dc:subject	
ISOcatLINK		

1.1.1.44 textType

definition	For text corpora that have already been using text type classification
type	myString
value space	
values	
examples	
DCLINK	dc:subject
ISOcatLINK	

1.1.1.45 register

definition	For corpora that have already been using register classification
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

38 To be replaced in the next version

1.1.1.46 subject_topic

definition	For text corpora that have already been using subject classification
type	myString
value space	
values	
examples	
DCLINK	dc:subject
ISOcatLINK	

1.1.1.47 conformanceToClassificationScheme

definition	For reference to external classification schemes (e.g. Library of Congress Subject headings etc.)
type	myStringURI
value space	
values	
examples	
DCLINK	
ISOcatLINK	

1.1.1.48 sizePerTextClassification

definition	used to give info on size of parts with different text classification
type	special status component
component	SizeInfo

7.2.10 AnnotationInfo

definition	Groups information on the annotated part(s) of a resource; must be repeated for each part that contains different annotations (different annotation type)
type	component
elements	annotationType *Status: mandatory* *Repeatability: 1* annotationStandoff *Status: optional* *Repeatability: 1* segmentationLevel *Status: condition-dependent mandatory*[39] *Condition: annotationType=segmentation* *Repeatability: unbounded* annotationFormat *Status: recommended* *Repeatability: 1* tagset *Status: recommended*

39 Mistakenly marked as mandatory in the XSD for text corpora. The condition will be implemented in the next version of the editor

	Repeatability: 1 tagsetLanguageId *Status: optional* *Repeatability: 1* theoreticModel *Status: optional* *Repeatability: unbounded* annotationManual *Status: optional* *Repeatability: 1* conformanceToStandardsBestPractice *Status: optional* *Repeatability: 1*[40] annotationTool *Status: recommended* *Repeatability: unbounded* annotationMode *Status: recommended* *Repeatability: 1* annotationModeDetails *Status: optional* *Repeatability: 1* annotationStartDate *Status: optional* *Repeatability: 1* annotationEndDate *Status: optional* *Repeatability: 1* interannotatorAgreement *Status: optional* *Repeatability: 1* intraannotatorAgreement *Status: optional* *Repeatability: 1* sizePerAnnotation *Status: optional*

40 unbounded in next version.

	Repeatability: 1 annotator *Status: optional* *Repeatability: unbounded*

1.1.1.49 annotationType

definition	The annotation level of the text corpus	
type	open controlled vocabulary	
value space	MS-annotationType	
values		value definition alignment audienceReactions discourseAnnotation discourseAnnotation-coreference discourseAnnotation-discourseRelations lemmatization morphosyntacticAnnotation morphosyntacticAnnotation-bPosTagging morphosyntacticAnnotation-PosTagging other questionTopicalTarget segmentation semanticAnnotation semanticAnnotation-certaintyLevel semanticAnnotation-entityMentions semanticAnnotation-Events semanticAnnotation-namedEntities semanticAnnotation-polarity semanticAnnotation-semanticClasses semanticAnnotation-semanticRelations semanticAnnotation-semanticRoles semanticAnnotation-temporalExpressions semanticAnnotation-wordSenses speechActs stemming structuralAnnotation syntacticAnnotation-shallowParsing syntacticAnnotation-subcategorizationFrames syntacticAnnotation-treebanks syntacticosemanticAnnotation-links textualEntailment
examples		
DCLINK		
ISOcatLINK		

1.1.1.50 annotationStandoff

definition	Indicates whether the annotation was created inline or in a stand-off fashion	
type	boolean	
value space		
values		
examples		
DCLINK		
ISOcatLINK		

1.1.1.51 segmentationLevel

definition	Indicates whether the annotation was created inline or in a stand-off fashion	
type	closed controlled vocabulary[41]	
value space	MS-segmentationLevel	
values		value definition paragraph sentence clause word wordGroup utterance
examples		
DCLINK		
ISOcatLINK		

1.1.1.52 annotationFormat

definition	Specifies the annotation format that is used since often the mime type will not be sufficient for machine processing	
type	closed controlled vocabulary	
value space	MS-annotationFormat	
values		value definition application/pdf text/html text/plain text/xml text/x-chat text/x-cut text/x-eaf+xml text/x-esf text/x-lexus-config+xml

41 Mistakenly marked as closed; this is already fixed for audio corpora and will be fixed for all in next version.

	text/x-lexus-resource+xml l text/x-lmf+xml text/x-shoebox-language text/ x-shoebox-text text/ x-shoebox-type text/ x-shoebox-lexicon text/ x-toolbox-text text/ x-toolbox-lexicon text/ x-cgn-bpt+xml text/ x-cgn-pri+xml text/ x-cgn-prx+xml text/ x-cgn-tag+xml text/ x-cgn-tig+xml AIF BAS MT TRS

	Unknown
examples	
DCLINK	dc:format
ISOcatLINK	

1.1.1.53 tagset

definition	Specifies the name/reference/url of the tagset used in the annotation of the resource or used by the tool/service
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

1.1.1.54 tagsetLanguageId

definition	Indicates the language of the tagset itself, expressed in the values of IETF and iso639-3
type	controlled vocabulary
value space	ISO 639-3:2007, Codes for the representation of names of languages – Part 3: Alpha-3 code for comprehensive coverage of languages[42]
values	
examples	
DCLINK	
ISOcatLINK	

1.1.1.55 theoreticModel

definition	Name of the theoretic model underlying the annotation task and/or reference (URL or bibliographic reference) to informative material about the theoretic model used
type	myString
value space	
values	
examples	

42 To be replaced by IETF BP47

DCLINK	
ISOcatLINK	

1.1.1.56 annotationManual

definition	Bibliographic reference or httpURI link to the annotation manual; alternative to the relation AnnotationDocumentedIn	
type	myStringURI	
value space		
values		
examples		
DCLINK		
ISOcatLINK		

1.1.1.57 conformanceToStandardsBestPractice

definition	Name of the standard/best practice to which the tagset used for the annotation conforms (e.g. MULTEXT, PDT, Time-ML etc.)	
type	open controlled vocabulary	
value space	MS-standards	
values		value definition EAGLES XCES MULTEXT T PDT Time-ML GrAF SYNAF TEI other
examples		
DCLINK		
ISOcatLINK		

1.1.1.58 annotationTool

definition	Gives the name or the url of a tool used for annotating a resource	
type	myStringURI	
value space		
values		

examples	
DCLINK	
ISOcatLINK	

1.1.1.59 annotationMode

definition	Indicates whether the resource was created manually or by automatic processes	
type	closed controlled vocabulary	
value space	MS-annotationMode	
values		Value definition automatic manual mixed interactive
examples		
DCLINK		
ISOcatLINK		

1.1.1.60 annotationModeDetails

definition	Short description of the annotation process	
type	myString	
value space		
values		
examples		
DCLINK		
ISOcatLINK		

1.1.1.61 annotationStartDate

definition	Start date of annotation	
type	date	
value space		
values		
examples		
DCLINK	dc:date	
ISOcatLINK		

1.1.1.62 annotationEndDate

definition	End date of annotation
type	date
value space	
values	
examples	
DCLINK	dc:date
ISOcatLINK	

1.1.1.63 interannotatorAgreement

definition	An indication of the inter-annotator agreement if appropriate methods where applied
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

1.1.1.64 intraannotatorAgreement

definition	An indication of the intra-annotator agreement if appropriate methods where applied
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

1.1.1.65 sizePerAnnotation

definition	used to give info on size of partially annotated resources
type	component
component	SizeInfo

1.1.1.66 annotator

definition	used to give info on annotators of a resource
type	special status component
componen t	PersonInfo

7.3 AudioInfo

definition	Groups together information on the audio module of a resource
type	component
elements	LingualityInfo [as defined for TextInfo] *Status: mandatory* *Repeatability: 1* LanguageInfo [as defined for TextInfo] *Status: mandatory* *Repeatability: unbounded* AudioContentInfo *Status: recommended* *Repeatability: 1* AudioSizeInfo[43] *Status: mandatory* *Repeatability: unbounded* AudioFormatInfo *Status: recommended* *Repeatability: unbounded* AudioSettingInfo *Status: recommended* *Repeatability: 1* DomainInfo [as defined for TextInfo] *Status: recommended* *Repeatability: unbounded* TimeCoverageInfo [as defined for TextInfo] *Status: recommended* *Repeatability: unbounded* GeographicCoverageInfo [as defined for TextInfo]

43 In XSD 1.0 (before simplification/refactoring) it was named *SizeInfo* [extension of the basic SizeInfo].

	Status: recommended *Repeatability: unbounded* AudioClassificationInfo *Status: recommended* *Repeatability: unbounded* AudioAnnotationInfo [44] *Status: recommended* *Repeatability: unbounded* AudioRecordingInfo *Status: recommended* *Repeatability: 1*

44 In the XSD's v1 before refactoring/simplification, this was implemented as *AnnotationInfo* but it was exactly the same. Please also note that there are differences between this and the *AnnotationInfo* of text corpora (and this is why it has been renamed): the elements *annotationType* and *segmentationLevel* have different sets of values, the element *annotatedElements* is used only for the audio corpora, the element *conformanceToStandardsBestPractice* is renamed for audio corpora as *conformanceToStandardsBestPractices*, it is of type "myStringURI" and it is repeatable and, finally, the order of the components slightly differs.

7.3.1 AudioContentInfo

definition	Groups together information on the contents of the audio part of a resource
type	component
elements	speechItems *Status: optional* *Repeatability: unbounded* nonSpeechItems *Status: optional* *Repeatability: unbounded* textualDescription *Status: optional* *Repeatability: 1*

1.1.1.67 speechItems

definition	distinct elements that are pronounced and annotated as such	
type	open controlled vocabulary	
value space	MS-speechItems	
values		value definition isolatedWords isolatedDigits naturalNumbers bankAccount other
examples		
DCLINK		
ISOcatLINK		

1.1.1.68 nonSpeechItems

definition	distinct elements that maybe included in the audio	
type	open controlled vocabulary	
value space	MS-nonSpeechItems	
values		Value definition notes tempo sounds other
examples		

DCLINK	
ISOcatLINK	

1.1.1.69 textualDescription

definition	legend of the soundtrack
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

7.3.2 AudioSizeInfo

definition	SizeInfo Element for Audio parts of a resource
type	component
elements	size *Status: mandatory* *Repeatability: 1* sizeUnitMultiplier[45] *Status: mandatory* *Repeatability: 1* sizeUnit *Status: mandatory* *Repeatability: 1* DurationOfEffectiveSpeechInfo *Status: optional* *Repeatability: unbounded* DurationOfAudioInfo *Status: optional* *Repeatability: unbounded*

1.1.1.70 DurationOfEffectiveSpeechInfo

definition	Groups together information on the duration of effective speech
type	component

45 *sizeUnitMultiplier* is dropped in the next version.

elements	size *Status: mandatory* *Repeatability: 1* sizeUnit *Status: mandatory* *Repeatability: 1*

1.1.1.70.1size

definition	The size of the resource with regard to the SizeUnit measure number
type	integer
value space	
values	
examples	
DCLINK	
ISOcatLINK	

1.1.1.70.2 sizeUnit

definition		Specification of the unit of size that is used when providing the size of a resource
type	closed controlled vocabulary	
value space	MS-DurationOfEffectiveSpeechInfo/sizeUnit	
values		value definitio n hours minutes seconds
examples		
DCLINK		
ISOcatLINK		

1.1.1.71 DurationOfAudioInfo

definition	Groups together information on the size of audio parts; for silences, music etc.
type	component
elements	size *Status: mandatory* *Repeatability: 1*

	sizeUnit *Status: mandatory* *Repeatability: 1*

1.1.1.71.1 size

definition		The size of the resource with regard to the SizeUnit measure number
type	number	
value space		
values		
examples		
DCLINK		
ISOcatLINK		

1.1.1.71.2 sizeUnit

definition		Specification of the unit of size that is used when providing the size of a resource
type	closed controlled vocabulary	
value space	MS-DurationOfEffectiveSpeechInfo/sizeUnit	
values		value definition hours minutes seconds
examples		
DCLINK		
ISOcatLINK		

7.3.3 AudioFormatInfo

definition	Groups together information on the format of the audio part of a resource
type	component
elements	mime-type *Status: mandatory* *Repeatability: unbounded* signalEncoding *Status: recommended*

	Repeatability: unbounded samplingRate *Status: recommended* *Repeatability: 1* quantization *Status: optional* *Repeatability: 1* byteOrder *Status: optional* *Repeatability: 1* signConvention *Status: optional* *Repeatability: 1* compression *Status: optional* *Repeatability: 1* compressionName *Status: optional* *Repeatability: unbounded*

	compressionLoss *Status: optional* *Repeatability: 1* audioQualityMeasuresIncluded *Status: optional* *Repeatability: 1* numberOfTracks *Status: optional* *Repeatability: 1* recordingQuality *Status: optional* *Repeatability: 1* sizePerAudioFormat *Status: optional* *Repeatability: unbounded*

1.1.1.72 signalEncoding

definition	what encoding the audio type uses	
type	open controlled vocabulary	
value space	MS-signalEncoding	
values		value definition aLaw microLaw linearPC M µ-law other
examples		
DCLINK		
ISOcatLINK		

1.1.1.73 samplingRate

definition	format of files contained in the resource in Hertz
type	string
value space	
values	
examples	
DCLINK	
ISOcatLIN	

K	

1.1.1.74 quantization

definition	the number of bits for each audio sample	
type	closed controlled vocabulary	
value space	MS-quantization	
values		Value definition 8 16 32 64
examples		
DCLINK		
ISOcatLINK		

1.1.1.75 byteOrder

definition	byte order of 2 or more bytes sample	
type	closed controlled vocabulary	
value space	MS-byteOrder	
vale		value definition lowHi hiLow littleEndian bigEndian
examples		
DCLINK		
ISOcatLINK		

1.1.1.76 signConvention

definition	binary representation of numbers	
type	closed controlled vocabulary	
value space	MS-signConvention	
value		value definition signed unsigned
examples		

DCLINK	
ISOcatLINK	

1.1.1.77 compression

definition	whether the audio is compressed or not
type	boolean
value space	
values	
examples	
DCLINK	
ISOcatLINK	

1.1.1.78 compressionName

definition	compression name	
type	open controlled vocabulary	
value space	MS-compressionName	
values	value	definition
	flac	
	shorten	
	mp3	
	vorbis	
	other	
examples		
DCLINK		
ISOcatLINK		

1.1.1.79 compressionLoss

definition	whether there is loss due to compression
type	boolean
value space	
values	
examples	
DCLINK	
ISOcatLINK	

1.1.1.80 audioQualityMeasuresIncluded

definition	audio quality measures	
type	open controlled vocabulary	
value space	MS-audioQualityMeasuresIncluded	
values	value	definition
	SNR	
	crossTalk	
	clippingRate	
	backgroundNoise	
	other	
examples		
DCLINK		
ISOcatLINK		

1.1.1.81 numberOfTracks

definition	specification of the number of audio channels	
type	open controlled vocabulary	
value space	MS-numberOfTracks	
values	value	definition
	mono	
	stereo	
examples		
DCLINK		
ISOcatLINK		

1.1.1.82 recordingQuality

definition	Indication of the audio recording quality	
type	open controlled vocabulary	
value space	MS-recordingQuality	
values	value	definition
	veryLow	
	low	
	medium	
	high	
	veryHigh	

examples	
DCLINK	
ISOcatLINK	

1.1.1.83 sizePerAudioFormat

definition	Used to give info on size of parts of a resource that differ as to the format
type	special status component
component	SizeInfo

7.3.4 AudioSettingInfo

definition	Groups together information on the setting of the audio part of a resource
type	component
elements	typeOfSituationOfCommunication *Status: optional* *Repeatability: 1* speechSetting *Status: optional* *Repeatability: unbounded* speechTask *Status: optional* *Repeatability: unbounded* audience *Status:* 1 *Repeatability: unbounded* noiseLevel *Status: optional* *Repeatability: 1* interactivity *Status: optional* *Repeatability: 1*

1.1.1.84 typeOfSituationOfCommunication

definition	type of speech state
type	open controlled vocabulary

value space	MS-typeOfSituationOfCommunication	
value		value definition readSpeech plannedSpeech semiPlannedSpeech spontaneousSpeech emotionalSpeech other
examples		
DCLINK		
ISOcatLINK		

1.1.1.85 speechSetting

definition	conversational type	
type	closed controlled vocabulary	
value space	MS-speechSetting	
values		value definition monologue dialogue multilogue
examples		
DCLINK		
ISOcatLINK		

1.1.1.86 speechTask

definition	task defined for the conversation	
type	open controlled vocabulary	
value space	MS-speechTask	
values		value definition meeting lecture frogStory pearStory mapTask onlineEducationalGame rolePlay wordGame wizardOfOz other
examples		

DCLINK	
ISOcatLINK	

1.1.1.87 audience

definition	Indication of the intended audience size	
type	closed controlled vocabulary	
value space	MS-audience	
values	value no few some largePublic	definition
examples		
DCLINK		
ISOcatLINK		

1.1.1.88 noiseLevel

definition	level of background noise	
type	closed controlled vocabulary	
value space	MS-noiseLevel	
values	value low medium high	definition
examples		
DCLINK		
ISOcatLINK		

1.1.1.89 interactivity

definition	conversational interaction between speakers	
type	closed controlled vocabulary	
value space	MS-interactivity	
values	value interactive nonInteractive	definition

		semiInteractiv e overlapping
examples		
DCLINK		
ISOcatLIN K		

7.3.5 AudioClassificationInfo

definition	Groups together information on audio type/genre of the resource
type	component
elements	audioGenre *Status: mandatory* *Repeatability: 1* speechGenre *Status: recommended* *Repeatability: 1* subject_topic *Status: optional* *Repeatability: 1* register *Status: optional* *Repeatability: 1* conformanceToClassificationScheme *Status: optional* *Repeatability: 1* sizePerAudioClassification *Status: optional* *Repeatability: unbounded*

1.1.1.90 audioGenre

definition	a first indication of type of sounds recorded	
type	closed controlled vocabulary	
value space	MS-audioGenre	
values		value definition speech song

	instrumenta l
examples	
DCLINK	
ISOcatLIN K	

1.1.1.91 speechGenre

definition	Genre: The conventionalized discourse or text types of the content of the resource, based on extra-linguistic and internal linguistic criteria; the values here are intended only for speech	
type	open controlled vocabulary	
value space	MS-speechGenre	
values		value definition broadcast news meeting lecture spontaneous emotional_expressiv e wideband airTrafficControl animalSpeech other
examples		
DCLINK		
ISOcatLIN K		

1.1.1.92 subject_topic

definition	Topics of the specific pieces recorded
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLIN K	

1.1.1.93 register

definition	For corpora that have already been using register classification
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

1.1.1.94 conformanceToClassificationScheme

definition	For reference to external classification schemes (e.g. Library of Congress Subject headings etc.); name/reference/url can be used
type	myStringURI
values	
examples	
DCLINK	
ISOcatLINK	

1.1.1.95 sizePerAudioClassification

definition	Used to give info on subparts of a resource which differ as to the classification parameter
type	special status component
component	SizeInfo

7.3.6 AudioAnnotationInfo

definition	Groups information on the annotated part(s) of an audio corpus; must be repeated for each part that contains different annotations (different annotation type) N.B. Please, see footnote 44 for differences with AnnotationInfo of text corpora
type	component
elements	annotationType[46] *Status: mandatory* *Repeatability: 1* annotatedElements [47] *Status: optional* *Repeatability: unbounded* annotationStandoff *Status: optional* *Repeatability: 1* segmentationLevel [48] *Status: condition-dependent mandatory*[49] *Condition: annotationType=segmentation* *Repeatability: unbounded* annotationFormat *Status: recommended* *Repeatability: 1* tagset *Status: recommended* *Repeatability: 1*

46 The element *annotationType* has a different set of values for text and audio corpora.

47 In the current version, this is only added to the *AudioInfo*.

48 The element *segmentationLevel* has a different set of values for text and audio corpora

49 The condition will be implemented in the next version of the editor

	tagsetLanguageId *Status: optional* *Repeatability: 1* conformanceToStandardsBestPractices *Status: optional* *Repeatability: unbounded* theoreticModel *Status: optional* *Repeatability: unbounded* annotationManual *Status: optional* *Repeatability: 1* annotationMode *Status: recommended* *Repeatability: 1* annotationModeDetails *Status: optional* *Repeatability: 1* annotationTool *Status: recommended* *Repeatability: unbounded* annotationStartDate *Status: optional* *Repeatability: 1* annotationEndDate *Status: optional* *Repeatability: 1* sizePerAnnotation *Status: optional* *Repeatability: 1* interannotatorAgreement *Status: optional* *Repeatability: 1* intraannotatorAgreement *Status: optional* *Repeatability: 1* annotator

	Status: optional *Repeatability: unbounded*

1.1.1.96 annotationType

definition	The annotation level of the text corpus	
type	open controlled vocabulary	
value space	MS-annotationType	
values	value	definition
	alignment	
	discourseAnnotation	
	discourseAnnotation-audienceReactions	
	discourseAnnotation-coreference	
	discourseAnnotation-dialogueActs	
	discourseAnnotation-discourseRelations	
	lemmatization	
	morphosyntacticAnnotation-bPosTagging	
	morphosyntacticAnnotation-posTagging	
	other	
	segmentation	
	semanticAnnotation	
	semanticAnnotation-certaintyLevel	
	semanticAnnotation-emotions	
	semanticAnnotation-entityMentions	
	semanticAnnotation-events	
	semanticAnnotation-namedEntities	
	semanticAnnotation-polarity	
	semanticAnnotation-questionTopicalTarget	
	semanticAnnotation-semanticClasses	
	semanticAnnotation-semanticRelations	
	semanticAnnotation-semanticRoles	
	semanticAnnotation-speechActs	
	semanticAnnotation-temporalExpressions	
	semanticAnnotation-textualEntailment	
	semanticAnnotation-wordSenses	
	speechAnnotation	
	speechAnnotation-orthographicTranscription	
	speechAnnotation-paralanguageAnnotation	
	speechAnnotation-phoneticTranscription	
	speechAnnotation-prosodicAnnotation	
	speechAnnotation-soundEvents	
	speechAnnotation-soundToTextAlignment	
	speechAnnotation-speakerIdentification	
	speechAnnotation-speakerTurns	
	stemming	
	structuralAnnotation	
	syntacticAnnotation-shallowParsing	
	syntacticAnnotation-subcategorizationFrame	
	s	
	syntacticAnnotation-treebanks	
	syntacticosemanticAnnotation-links	
	translation	

	transliteration
examples	
DCLINK	
ISOcatLINK	

1.1.1.97 annotatedElements

definition	annotated elements; N.B. used only for audio
type	open controlled vocabulary
value space	MS-annotatedElements
values	value definition speakerNoise backgroundNoise mispronunciations truncation discourseMarkers other
examples	
DCLINK	
ISOcatLINK	

1.1.1.98 segmentationLevel

definition	Indicates whether the annotation was created inline or in a stand-off fashion
type	open controlled vocabulary
value space	MS-segmentationLevel
values	value definition paragraph sentence clause word wordGroup utterance topic signal phoneme syllable phrase diphone prosodicBoundaries other frame scene

	shot
examples	
DCLINK	
ISOcatLINK	

1.1.1.99 conformanceToStandardsBestPractices

definition	Name of the standard/best practice to which the tagset used for the annotation conforms (e.g. MULTEXT, PDT, Time-ML etc.)
type	myStringURI
value space	
values	
examples	
DCLINK	
ISOcatLINK	

7.3.7 AudioRecordingInfo

definition	Groups together information on the recording of the audio part of a corpus
type	component
elements	originalSource *Status: optional* *Repeatability: unbounded* recordingMode *Status: optional* *Repeatability: 1* recordingModeDetails *Status: optional* *Repeatability: 1* recordingDeviceType *Status: optional* *Repeatability: unbounded* recordingDeviceTypeDetails *Status: optional* *Repeatability: 1* recordingPlatformSoftware *Status: optional*

	Repeatability: unbounded recordingEnvironment *Status: optional* *Repeatability: unbounded* sourceChannel *Status: optional* *Repeatability: 1* sourceChannelType *Status: optional* *Repeatability: unbounded* sourceChannelName *Status: optional* *Repeatability: unbounded* sourceChannelDetails *Status: optional* *Repeatability: 1* recorder *Status: optional* *Repeatability: unbounded* AudioCaptureInfo *Status: optional* *Repeatability: unbounded*

1.1.1.100 originalSource

definition	The main sources used for the creation of the resource (dictionaries, grammars, lexica, corpora, ...)
type	myString
value space	
values	
examples	
DCLINK	dc:source
ISOcatLINK	

1.1.1.101 recordingMode

definition	Whether the audio was recorded manually or automatically
type	closed controlled vocabulary
value space	MS-recordingMode

values	value definition automatic manual mixed interactiv e
examples	
DCLINK	
ISOcatLIN K	

1.1.1.102 recordingModeDetails

definition	Desription of the recording mode
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLIN K	

1.1.1.103 recordingDeviceType

definition	The nature of the recording platform hardware and the storage medium
type	open controlled vocabulary
value space	MS-recordingDeviceType
values	value definition analogCassetteRecorder digitalAudioTapeRecorde r minidiskRecorder pcCard cdRecorder hardDiskRecorder other
examples	
DCLINK	
ISOcatLIN K	

1.1.1.104 recordingDeviceTypeDetails

definition	Free text description of the recoding device
type	myString

value space	
values	
examples	
DCLINK	
ISOcatLINK	

1.1.1.105 recordingPlatformSoftware

definition	The software used for the recording platform	
type	open controlled vocabulary	
value space	MS-recordingPlarformSoftware	
values		value definition cubase audition_adobe soundforge_sony other
examples		
DCLINK		
ISOcatLINK		

1.1.1.106 recordingEnvironment

definition	Where the recording took place	
type	open controlled vocabulary	
value space	MS-recordingEnvironment	
values		value definition office car studio publicPlace industrial other
examples		
DCLINK		
ISOcatLINK		

1.1.1.107 sourceChannel

definition	Information on the audio source channel
type	open controlled vocabulary

value space	MS-sourceChannel	
values		value definition internet radio tv telephone webCam other
examples		
DCLINK		
ISOcatLINK		

1.1.1.108 sourceChannelType

definition	Type of the source channel	
type	open controlled vocabulary	
value space	MS-sourceChannelType	
values		value definition ISDN GSM 3G CDMA other
examples		
DCLINK		
ISOcatLINK		

1.1.1.109 sourceChannelName

definition	The name of the specific source recorded
type	string
value space	
values	
examples	
DCLINK	
ISOcatLINK	

1.1.1.110 sourceChannelDetails

definition	The details of the channel equipment used (brand, type,...)
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

1.1.1.111 recorder

definition	Information on the recorder(s) of the audio resource
type	special status component
component	PersonInfo

1.1.1.112 AudioCaptureInfo

definition	Groups together information on the capture of the audio part of a corpus
type	component
elements	capturingDeviceType *Status: optional* *Repeatability: unbounded* capturingDeviceTypeDetails *Status: optional* *Repeatability: 1* PersonSourceSetInfo *Status: recommended* *Repeatability: 1*

1.1.1.112.1 capturingDeviceType

definition	the transducers through which the audio is captured	
type	open controlled vocabulary	
value space	MS-capturingDeviceType	
values		Value definition studioEquipment microphone microphoneArray embeddedMicrophone largeMembraneMicrophone

	laryngograph other
examples	
DCLINK	
ISOcatLINK	

1.1.1.112.2 capturingDeviceTypeDetails

definition	free text description for further information on the capturing device
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

1.1.1.112.3 PersonSourceSetInfo

definition	Information on the persons (speakers, video participants, etc.) in the audio/video/sensorimotor parts of the resource
type	component
elements	numberOfPersons *Status: optional* *Repeatability: unbounded* ageOfPersons *Status: optional* *Repeatability: unbounded* ageRangeStart *Status: optional* *Repeatability: unbounded* ageRangeEnd *Status: optional* *Repeatability: unbounded* sexOfPersons *Status: optional* *Repeatability: unbounded* originOfPersons *Status: optional* *Repeatability: unbounded*

	dialectAccentOfPersons *Status: optional* *Repeatability: unbounded* geographicDistributionOfPersons *Status: optional* *Repeatability: unbounded* hearingImpairmentOfPersons *Status: optional* *Repeatability: unbounded* speakingImpairmentOfPersons *Status: optional* *Repeatability: unbounded* numberOfTrainedSpeakers *Status: optional* *Repeatability: unbounded* speechInfluences *Status: optional* *Repeatability: unbounded* participant (ParticipantInfo)[50] *Status: optional* *Repeatability: unbounded*

50 In the XSD named as "participant"

1.1.1.112.3.1 numberOfPersons

definition	the number of the persons participating in the audio
type	integer
value space	
values	
examples	
DCLINK	
ISOcatLINK	

1.1.1.112.3.2 ageOfPersons

definition	the age range of the group of participants; repeat the element if needed		
type	closed controlled vocabulary		
value space	MS-ageOfPersons		
values		value	definition
		child	
		teenager	
		adult	
		elderly	
examples			
DCLINK			
ISOcatLINK			

1.1.1.112.3.3 ageRangeStart

definition	Start of age range
type	string
value space	
values	
examples	
DCLINK	
ISOcatLINK	

1.1.1.112.3.4 ageRangeEnd

definition	End of age range
type	string
value space	
values	
examples	
DCLINK	

ISOcatLINK	

1.1.1.112.3.5 sexOfPersons

definition	the gender of the group of persons participating in the audio	
type	closed controlled vocabulary	
value space	MS-sexOfPersons	
values	value male female mixed unknown	definition
examples		
DCLINK		
ISOcatLINK		

1.1.1.112.3.6 originOfPersons

definition	speaker language origin	
type	closed controlled vocabulary	
value space	MS-originOfPersons	
values	value native nonNative mixed unknown	definition
examples		
DCLINK		
ISOcatLINK		

1.1.1.112.3.7 dialectAccentOfPersons

definition	speaker dialect information
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLIN	

K	

1.1.1.112.3.8 geographicDistributionOfPersons

definition	speaker geographic coverage
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

1.1.1.112.3.9 hearingImpairmentOfPersons

definition	whether the group contains persons with hearing impairments	
type	closed controlled vocabulary	
value space	MS-hearingImpairmentOfPersons	
values	value	definition
	yes	
	no	
	mixed	
examples		
DCLINK		
ISOcatLINK		

1.1.1.112.3.10 speakingImpairmentOfPersons

definition	whether the group contains persons with speakingimpairments	
type	closed controlled vocabulary	
value space	MS-hearingImpairmentOfPersons	
values	value	definition
	yes	
	no	
	mixed	
examples		
DCLINK		
ISOcatLINK		

1.1.1.112.3.11 numberOfTrainedSpeakers

definition	number of speakers that have been trained for the specific task
type	integer
value space	
values	
examples	
DCLINK	
ISOcatLINK	

1.1.1.112.3.12 speechInfluences

definition	factors influencing speech	
type	open controlled vocabulary	
value space	MS-speechInfluences	
values		value definition alcohol sleepDeprivation hyperbaric other
examples		
DCLINK		
ISOcatLINK		

1.1.1.112.3.13 ParticipantInfo

definition	Information on the individual person participating in the audio/video/sensorimotor part of the resource
type	component
elements	alias *Status: optional* *Repeatability: unbounded* ageGroup *Status: optional* *Repeatability: unbounded* age *Status: optional* *Repeatability: unbounded* sex *Status: optional* *Repeatability: unbounded* origin *Status: optional* *Repeatability: unbounded* placeOfLiving *Status: optional* *Repeatability: unbounded* placeOfBirth *Status: optional* *Repeatability: unbounded* placeOfChildhood *Status: optional* *Repeatability: unbounded* dialectAccent *Status: optional* *Repeatability: unbounded* speakingImpairment *Status: optional* *Repeatability: unbounded* hearingImpairment *Status: optional* *Repeatability: unbounded* smokingHabits

	Status: optional *Repeatability: unbounded* vocalTractConditions *Status: optional* *Repeatability: unbounded* profession *Status: optional* *Repeatability: unbounded* height *Status: optional* *Repeatability: unbounded* weight *Status: optional* *Repeatability: unbounded* trainedSpeaker *Status: optional* *Repeatability: unbounded* placeOfSecondEducation *Status: optional* *Repeatability: unbounded* educationLevel *Status: optional* *Repeatability: unbounded*

7.3.7.1.1.1.1 alias

definition	name used instead of the real one
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

7.3.7.1.1.1.2 ageGroup

definition	
type	closed controlled vocabulary
value space	MS-ageGroup

values	value: child, teenager, adult, elderly definition:
examples	
DCLINK	
ISOcatLINK	

7.3.7.1.1.1.3 age

definition	
type	free text
value space	
values	
examples	
DCLINK	
ISOcatLINK	

7.3.7.1.1.1.4 sex

definition	
type	closed controlled vocabulary
value space	MS-sex
values	value: male, female, unknown definition:
examples	
DCLINK	
ISOcatLINK	

7.3.7.1.1.1.5 origin

definition	
type	closed controlled vocabulary
value space	MS-origin

values	value definition native nonNativ e unknown
examples	
DCLINK	
ISOcatLIN K	

7.3.7.1.1.1.6 placeOfLiving

definition	
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLIN K	

7.3.7.1.1.1.7 placeOfBirth

definition	
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLIN K	

7.3.7.1.1.1.8 placeOfChildhood

definition	
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLIN K	

7.3.7.1.1.1.9 dialectAccent

definition	
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

7.3.7.1.1.1.10 speakingImpairment

definition	
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

7.3.7.1.1.1.11 hearingImpairment

definition	
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

7.3.7.1.1.1.12 smokingHabits

definition	
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

7.3.7.1.1.1.13 vocalTractConditions

definition		
type	closed controlled vocabulary	
value space	MS-vocalTractConditions	
values		value definition dentalProsthesis
examples		
DCLINK		
ISOcatLINK		

7.3.7.1.1.1.14 profession

definition	
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

7.3.7.1.1.1.15 height

definition	
type	integer
value space	
values	
examples	
DCLINK	
ISOcatLINK	

7.3.7.1.1.1.16 weight

definition	
type	integer
value space	
values	
examples	
DCLINK	
ISOcatLINK	

7.3.7.1.1.1.17 trainedSpeaker

definition	
type	boolean
value space	
values	
examples	
DCLINK	
ISOcatLINK	

7.3.7.1.1.1.18 placeOfSecondEducation

definition	
type	integer
value space	
values	
examples	
DCLINK	
ISOcatLINK	

7.3.7.1.1.1.19 educationLevel

definition	
type	integer
value space	
values	
examples	
DCLINK	
ISOcatLINK	

8 Lexical/Conceptual Resources

8.1 Resource

definition	extension of Resource for LexicalConceptualResources
elements	**Resource – common components** IdentificationInfo ContentInfo [N.B.: resourceType=lexicalConceptualResource; mediaType=text] DistributionInfo

	contactPerson MetadataInfo VersionInfo ValidationInfo UsageInfo ResourceDocumentationInfo ResourceCreationInfo **Additional components** LexicalConceptualResourceInfo *Status: Mandatory* *Repeatability: 1* TextInfo [restriction for lexical/conceptual resources] *Status: Mandatory* *Repeatability: unbounded*

8.2 LexicalConceptualResourceInfo

definition	Groups together information specific to lexical/conceptual resources
type	component
elements	lexicalConceptualResourceType *Status: mandatory* *Repeatability: 1* LexicalConceptualResourceCreationInfo *Status: optional*

	Repeatability: 1 LexicalConceptualResourceEncodingInfo *Status: recommended* *Repeatability: 1*

8.2.1 lexicalConceptualResourceType

definition	subtype of lexicalConceptualResources	
type	open controlled vocabulary	
value space	MS-lexicalConceptualResourceType	
values		value definition wordList computationalLexicon ontology wordnet thesaurus framenet terminologicalResource machineReadableDictionary lexicon other
examples		
DCLINK		
ISOcatLINK		

8.2.2 LexicalConceptualResourceCreationInfo

definition	Groups all information regarding the creation process of the lexicalConceptualResource
type	component
elements	originalSource *Status: Optional* *Repeatability: unbounded* creationMode *Status: Optional* *Repeatability: 1* creationModeDetails *Status: Optional* *Repeatability: 1*

	creationTool *Status: Optional* *Repeatability: unbounded*

1.1.1.113 originalSource

definition	The main sources used for the creation of the resource (dictionaries, grammars, lexica, corpora, ...)
type	myString
value space	
values	
examples	
DCLINK	dc:source
ISOcatLINK	

1.1.1.114 creationMode

definition	A first indication as to the mode of creation of the resource	
type	open controlled vocabulary	
value space	MS-creationMode	
values		value definition automatic manual mixed interactive
values		
examples		
DCLINK		

ISOcatLINK	

1.1.1.115 creationModeDetails

definition	free text used to supply more details as to the creation methods and processes of the resource
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

1.1.1.116 creationTool

definition	Indicates the tool with help of which the resource was created; alternative to the preferred relation hasAsCreationTool
type	myStringURI
value space	
values	
examples	
DCLINK	
ISOcatLINK	

8.2.3 LexicalConceptualResourceEncodingInfo

definition	Groups all information regarding the contents of lexical/conceptual resources
type	component
elements	encodingLevel *Status: Mandatory* *Repeatability: unbounded* conformanceToStandardsBestPractice *Status: recommended* *Repeatability: unbounded* theoreticModel *Status: optional* *Repeatability: unbounded* linguisticInformation *Status: recommended* *Repeatability: unbounded* externalRef *Status: Optional* *Repeatability: unbounded* extratextualInformation *Status: Optional* *Repeatability: unbounded*

1.1.1.117 encodingLevel

definition	Information on the contents of the lexicalConceptualResource as regards the

	linguistic level of analysis	
type	open controlled vocabulary	
value space	MS-encodingLevel	
values		value definition phonetics phonology semantics morpholog y syntax pragmatics other
examples		
DCLINK		
ISOcatLIN K		

1.1.1.118 conformanceToStandardsBestPractice

definition	Name of the standard/best practice to which the tagset used for the annotation conforms (e.g. MULTEXT, PDT, Time-ML etc.)	
type	open controlled vocabulary	
value space	MS-lexiconStandards	
values		value definition LMF TMF PAROLE WordNet FrameNe t COMLEX

	SIMPLE LC-STAR EAGLES OLIF LADL other
examples	
DCLINK	
ISOcatLINK	

1.1.1.119 theoreticModel

definition	Name of the theoretic model underlying the lexicalConceptualResource and/or reference (URL or bibliographic reference) to informative material about the theoretic model used
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

1.1.1.120 linguisticInformation

definition	a more detailed account of the linguistic information contained in the lexicalConceptualResource
type	open controlled vocabulary
value space	MS-linguisticInformation

values	value definition accentuation lemma lemma-MultiWordUnits lemma-Variants lemma-Abbreviations lemma-Compounds lemma-CliticForms partOfSpeech morpho-Case morpho-Gender morpho-Number morpho-Degree morpho-IrregularForms morpho-Mood morpho-Tense morpho-Person morpho-Aspect morpho-Voice morpho-Auxiliary morpho-Inflection morpho-Reflexivity syntax-SubcatFrame semantics-Traits semantics-SemanticClass semantics-CrossReferences semantics-Relations semantics-Relations-Hyponyms semantics-Relations-Hyperonyms semantics-Relations-Synonyms semantics-Relations-Antonyms semantics-Relations-Troponyms semantics-Relations-Meronyms usage-Frequency usage-Register

	usage-Collocations usage-Examples usage-Notes definition/gloss translationEquivalent phonetics-Transcription semantics-Domain semantics-EventType semantics-SemanticRoles statisticalType morpho-Derivation other semantics-QualiaStructure syntacticoSemanticLinks
examples	
DCLINK	
ISOcatLINK	

1.1.1.121 externalRef

definition	Another resource to which the lexicalConceptualResource is linked (e.g. link to a wordnet or ontology)
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

1.1.1.122 extratextualInformation

definition	An indication of the extratextual information contained in the lexicalConceptualResouce; can be used as an alternative to audio, image, videos etc. for cases where these are not considered an important part of the lcr
type	open controlled vocabulary
value space	MS-extratextualInformation
values	value definition images videos soundRecordings other
examples	
DCLINK	

ISOcatLINK	

8.3 TextInfo [restriction for lexical/conceptual resources]

definition	restriction of TextInfo for LexicalConceptualResources & Language descriptions [no TextCreation, Annotation & TextClassification] N.B. The TextInfo of Lexical/Conceptual Resources includes in the current version mistakenly two optional components (TextCreationInfo and AnnotationInfo); these are not included in the restricted TextInfo of Language descriptions.
type	component
elements	LingualityInfo *Status: Mandatory* *Repeatability: 1* LanguageInfo *Status: Mandatory* *Repeatability: unbounded* TextCreationInfo[51] *Status: Optional* *Repeatability: 1* SizeInfo *Status: Mandatory* *Repeatability: unbounded* TextFormatInfo *Status: Optional* *Repeatability: unbounded* CharacterEncodingInfo *Status: Optional* *Repeatability: unbounded* DomainInfo *Status: Optional* *Repeatability: unbounded* TimeCoverageInfo *Status: Optional* *Repeatability: unbounded*

51 This is wrongly included for Lexical/Conceptual Resources in the current version of the editor and the XSD's; to be fixed in next version; note, however, that the restricted *TextInfo* of language descriptions is correct.

	GeographicCoverageInfo *Status: Optional* *Repeatability: unbounded* AnnotationInfo[52] *Status: Optional* *Repeatability: unbounded*

9 Tools/Services

9.1 Resource

definition	extension of Resource for Tools/Services
elements	**Resource – common components** IdentificationInfo contactPerson VersionInfo DistributionInfo ValidationInfo ResourceCreationInfo UsageInfo MetadataInfo ResourceDocumentationInfo ContentInfo [N.B.: resourceType=technologyToolService] **Additional components** ToolServiceInfo *Status: Mandatory* *Repeatability: 1*

9.2 ToolServiceInfo

definition	Groups together elements required for the description of tools/services
type	component
elements	toolServiceType *Status: mandatory*

52 This is wrongly included for Lexical/Conceptual Resources in the current version of the editor and the XSD's; to be fixed in next version; note, however, that the restricted *TextInfo* of language descriptions is correct.

	Repeatability: 1 toolServiceSubtype *Status: optional* *Repeatability: unbounded* languageDependent *Status: mandatory* *Repeatability: 1* InputInfo[53] *Status: mandatory* *Repeatability: 1* OutputInfo[54] *Status: recommended* *Repeatability: 1* ToolServiceOperationInfo *Status: mandatory* *Repeatability: 1* ToolServiceEvaluationInfo *Status: recommended* *Repeatability: 1* ToolServiceCreationInfo *Status: recommended* *Repeatability: 1*

9.2.1 toolServiceType

definition	type of the toolService; select one of the recommended values		
type	open controlled vocabulary		
value space	MS-toolServiceType		
values		value	definition
		tool	a device that perform tasks listed in th component

53 Simplified in the current version of the editor and the accompanying XSD's; components *CorpusInfo*, *LexicalConceptualResourceInfo*, *TextInfo* and *AudioInfo*, which were present in v1 before the simplification/refactoring, will be added in the next version.

54 Simplified in the current version of the editor and the accompanying XSD's; components *CorpusInfo*, *LexicalConceptualResourceInfo*, *TextInfo* and *AudioInfo*, which were present in v1 before the simplification/refactoring, will be added in the next version.

	service	a form in which NI and delivered to a us of acquiring an corresponding tools
	platform	a technology that eas of new tools and serv
	suiteOfTools	a more of less organisation of tools combinaton of tasks
	infrastructure architecture	a technology that s development of together with all the
	nlpDevelopmentEnvironment	a technology th development of da lexicons, grammars, included in an Ar Platform
	other	
examples		
DCLINK		
ISOcatLINK		

9.2.2 toolServiceSubtype

definition	subtype of tool Service; free text; examples can be parser, tagger, annotator, corpus workbench etc.
type	myString
value space	
values	

examples	
DCLINK	
ISOcatLINK	

9.2.3 languageDependent

definition	Indicates whether the operation of the toolService is language dependent or not
type	boolean
value space	
values	
examples	
DCLINK	
ISOcatLINK	

9.2.4 InputInfo

definition	Groups together information on the requirements set on the input resource of a tool or service[55]
type	component
elements	resourceType *[restriction]* *Status: recommended* *Repeatability: unbounded* mediaType *Status: mandatory* *Repeatability: unbounded* modalityType *Status: optional* *Repeatability: unbounded*

1.1.1.123 resourceType

definition	resourceType restricted for use with tools and services
type	closed controlled vocabulary
value space	MS-resourceType

55 The components *CorpusInfo*, *LexicalConceptualResourceInfo*, *TextInfo* and *AudioInfo* will be added in the next version to cater for the requirements on the input resources

values	value definition corpus lexicalConceptualResource languageDescription
examples	
DCLINK	
ISOcatLINK	

1.1.1.124 modalityType

definition	modalityType restricted for use with tools and services
type	open controlled vocabulary
value space	MS-modalityType
values	Value definition bodyGesture facialExpression voice combinationOfModalities signLanguage spokenLanguage writtenLanguage other
examples	
DCLINK	
ISOcatLINK	

9.2.5 OutputInfo

definition	Groups together information on the requirements set on the output of a tool or service[56]
type	component
elements	resourceType *[restriction]* *Status: recommended* *Repeatability: unbounded* mediaType

56 The components *CorpusInfo*, *LexicalConceptualResourceInfo*, *TextInfo* and *AudioInfo* will be added in the next version to cater for the requirements on the output of a tool/service

	Status: mandatory *Repeatability: unbounded* modalityType *Status: optional* *Repeatability: unbounded*

9.2.6 ToolServiceOperationInfo

definition	Groups together information on the operation of a tool or service
type	component
elements	operatingSystem *Status: mandatory* *Repeatability: unbounded* RunningEnvironmentInfo *Status: recommended* *Repeatability: 1* runningTime *Status: recommended* *Repeatability: 1*

1.1.1.125 operatingSystem

definition	The operating system on which the tools will be runing	
type	open controlled vocabulary	
value space	MS-operatingSystem	
values		value definition os-independen t windows linux unix mac-OS other
examples		
DCLINK		
ISOcatLINK		

1.1.1.126 runningTime

definition	running time for the tool/service; recommended format: expressed in seconds
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

1.1.1.127 RunningEnvironmentInfo

definition	Groups together information on the running environment of a tool or service
type	component
elements	requiredSoftware *Status: recommended* *Repeatability: unbounded* requiredHardware *Status: optional* *Repeatability: unbounded* requiredLRs *Status: optional* *Repeatability: unbounded* runningEnvironmentDetails *Status: optional* *Repeatability: 1*

1.1.1.127.1 requiredSoftware

definition	If an addtional software should be installed before running the tool; this software is not included in the delivery of the tool
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

1.1.1.127.2 requiredHardware

definition	If for running a tool, specific hardware is required	
type	open controlled vocabulary	
value space	MS-requiredHardware	
values	Value graphicCard microphone ocrSystem specialHardwareEquipment t none other	annotation
examples		
DCLINK		
ISOcatLINK		

1.1.1.127.3 requiredLRs

definition	If for running a tool, specific LRs (e.g. a grammar, a list of words etc.) are required
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

1.1.1.127.4 runningEnvironmentDetails

definition	free text description for the running environment
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

9.2.7 ToolServiceEvaluationInfo

definitio	Groups together information on the evaluation status of a tool or service

n	
type	component
elements	evaluated *Status: mandatory* *Repeatability: 1* evaluationLevel *Status: optional* *Repeatability: unbounded* evaluationType *Status: optional* *Repeatability: unbounded* evaluationCriteria *Status: optional* *Repeatability: unbounded* evaluationMeasure *Status: optional* *Repeatability: unbounded* evaluationReport *Status: recommended* *Repeatability: unbounded* evaluationTool *Status: recommended* *Repeatability: unbounded* evaluationDetails *Status: recommended* *Repeatability: 1* evaluator *Status: optional* *Repeatability: unbounded*

1.1.1.128 evaluated

definition	Indicates whether the tool or service has been evaluated
type	boolean
value space	
values	
examples	
DCLINK	

ISOcatLINK	

1.1.1.129 evaluationLevel

definition	evaluation level	
type	closed controlled vocabulary	
value space	MS-evaluationLevel	
values	value technological usage impact diagnostic	definition
examples		
DCLINK		
ISOcatLINK		

1.1.1.130 evaluationType

definition	evaluation type	
type	closed controlled vocabulry	
value space	MS-evaluationType	
values	value glassBox blackBox	defintion
examples		
DCLINK		
ISOcatLINK		

1.1.1.131 evaluationCriteria

definition	evaluation criteria	
type	closed controlled vocabulry	
value space	MS-evaluationCriteria	
values	value extrinsic intrinsic	definition
examples		

DCLINK	
ISOcatLINK	

1.1.1.132 evaluationMeasure

definition	evaluation measure	
type	closed controlled vocabulry	
value space	MS-evaluationMeasure	
values	value human automatic	definition
examples		
DCLINK		
ISOcatLINK		

1.1.1.133 evaluationReport

definition	bibliographical record of or link to a report describing the evaluation process, tool, method etc. of the tool or service
type	myStringURI
value space	
values	
examples	
DCLINK	
ISOcatLINK	

1.1.1.134 evaluationTool

definition	name or id or url of the tool used for the evaluation of the tool or service
type	myStringURI
value space	
values	
examples	
DCLINK	
ISOcatLINK	

1.1.1.135 evaluationDetails

definition	free text element to add any information on the evaluation of a tool or service
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

1.1.1.136 evaluator

definition	Groups information on person/organization who evaluated the tool
type	special status component
component	choice between PersonInfo OrganizationInfo *Status: optional* *Repeatability: unbounded*

9.2.8 ToolServiceCreationInfo

definition	Groups together information on the creation of a tool or service
type	component
elements	implementationLanguage *Status: recommended* *Repeatability: unbounded* formalism *Status: optional* *Repeatability: unbounded* creationDetails *Status: optional* *Repeatability: 1*

1.1.1.137 implementationLanguage

definition	The programming languages needed for allowing user contributions, or for runing the tools, in case no executables are available

type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

1.1.1.138 formalism

definition	formalism (e.g. GATE, etc.) used to implement the tool/service
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

1.1.1.139 creationDetails

definition	free text element for additional information on the creation of a tool or service
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

10Language descriptions

10.1Resource

definition	extension of Resource for Language Descriptions
elements	**Resource – common components** IdentificationInfo contactPerson VersionInfo DistributionInfo ValidationInfo

	ResourceCreationInfo UsageInfo MetadataInfo ResourceDocumentationInfo ContentInfo [N.B.: resourceType=languageDescription; mediaType=text] **Additional components** LanguageDescriptionInfo *Status: Mandatory* *Repeatability: 1* TextInfo [restriction for language descriptions] *Status: Mandatory* *Repeatability: unbounded*

10.2 LanguageDescriptionInfo

definition	Groups together information on language descriptions (grammars, language models etc.)
type	component
elements	languageDescriptionType *Status: Mandatory* *Repeatability: 1* LanguageDescriptionCreationInfo *Status: recommended* *Repeatability: 1* LanguageDescriptionEncodingInfo *Status: Mandatory* *Repeatability: 1* LanguageDescriptionOperationInfo *Status: Optional* *Repeatability: 1* LanguageDescriptionPerformanceInfo *Status: recommended* *Repeatability: 1*

10.2.1 languageDescriptionType

definition	type of the language description	
type	open controlled vocabulary	
value space	MS-languageDescriptionType	
values	value	definition
	grammar	
	n-gramMode	
	l	
	other	
examples		
DCLINK		
ISOcatLINK		

10.2.2 LanguageDescriptionCreationInfo

definition	Groups together information on the creation of the ld
type	component
elements	originalSource *Status: Optional* *Repeatability: 1* creationMode *Status: Optional* *Repeatability: 1* creationModeDetails *Status: Optional* *Repeatability: unbounded* creationTool *Status: Optional* *Repeatability: unbounded* formalism *Status: Optional* *Repeatability: 1*

1.1.1.140 originalSource

definition	The main sources used for the creation of the resource (other grammars, corpora etc.)
type	

value space	
values	
examples	
DCLINK	dc:source
ISOcatLINK	

1.1.1.141 creationMode

definition	A first indication as to the mode of creation of the resource	
type	closed controlled vocabulary	
value space	MS-creationMode	
values		Value definition automatic manual mixed interactive
examples		
DCLINK		
ISOcatLINK		

1.1.1.142 creationModeDetails

definition	Used to supply more details as to the creation methods and processes of the resource
type	
value space	
values	
examples	
DCLINK	
ISOcatLINK	

1.1.1.143 creationTool

definition	Indicates the tool with help of which the resource was created; alternative to the preferred relation hasAsCreationTool
type	
value space	
values	
examples	

DCLINK	
ISOcatLINK	

1.1.1.144 formalism

definition	Reference (name, bibliographic reference or link to url) for the formalism used for the LanguageDescription
type	myStringURI
value space	
values	
examples	
DCLINK	
ISOcatLINK	

10.2.3 LanguageDescriptionEncodingInfo

definition	Groups together information on the contents of the LanguageDescriptions
type	component
elements	encodingLevel *Status: Mandatory* *Repeatability: unbounded* conformanceToStandardsBestPractice *Status: Optional* *Repeatability: unbounded* theoreticModel *Status: Optional* *Repeatability: unbounded* grammaticalPhenomenaCoverage *Status: Optional* *Repeatability: unbounded* task *Status: Optional* *Repeatability: unbounded* weightedGrammar *Status: Optional* *Repeatability: 1*

1.1.1.145 encodingLevel

definition	Information on the linguistic levels covered by the grammar	
type	open controlled vocabulary	
value space	MS-encodingLevel	
values		value definition phonetics phonology semantics morpholog y syntax pragmatics other
examples		
DCLINK		
ISOcatLIN K		

1.1.1.146 conformanceToStandardsBestPractice

definition	Name (or link to url or biblio reference) of the standard/best practice to which the LanguageDescriptions conforms
type	myStringURI
value space	
values	
examples	
DCLINK	
ISOcatLIN K	

1.1.1.147 theoreticModel

type	string
definition	Name of the theoretic model underlying the ld and/or reference (URL or bibliographic reference) to informative material about the theoretic model used
value space	
values	
examples	
DCLINK	
ISOcatLIN K	

1.1.1.148 grammaticalPhenomenaCoverage

definition	An indication of the grammatical phenomena covered by the grammar	
type	open controlled vocabulary	
value space	MS-grammaticalPhenomenaCoverage	
values		Value definition clauseStructure ppAttachment npStructure coordination anaphora other
examples		
DCLINK		
ISOcatLINK		

1.1.1.149 task

definition	an indication of the task performed by the grammar	
type	open controlled vocabulary	
value space	MS-task	
values		value definition anaphoraResolution chunking parsing npRecognition titlesParsing definitionsParsing analysis generation other
examples		
DCLINK		
ISOcatLINK		

1.1.1.150 weightedGrammar

definition	whether the grammar contains numeric weights
type	boolean
value space	
values	
examples	

DCLINK	
ISOcatLINK	

10.2.4 LanguageDescriptionOperationInfo

definition	Groups together information on the operation requirements of the LanguageDescriptions
type	component
elements	LexiconRequirementsInfo *Status: mandatory* *Repeatability: 1*

1.1.1.151 LexiconRequirementsInfo

definition	Groups together information on requirements for lexica set by the LanguageDescriptions
type	component
elements	relatedLexicon *Status: Mandatory* *Repeatability: 1* attachedLexiconPosition *Status: Optional* *Repeatability: 1* compatibleLexiconType *Status: Optional* *Repeatability: unbounded*

1.1.1.151.1 relatedLexicon

definition	position of the lexica that must or can be used with the grammar	
type	closed controlled vocabulary	
value space	MS-relatedLexicon	
values	Value	definition
	included	
	attached	
	none	

examples	
DCLINK	
ISOcatLINK	

1.1.1.151.2 attachedLexiconPosition

definition	position of attached lexicon
type	string
value space	
values	
examples	
DCLINK	
ISOcatLINK	

1.1.1.151.3 compatibleLexiconType

definition	type of lexicon that can be used with the grammar; for external lexica	
type	open controlled vocabulary	
value space	MS-compatibleLexiconType	
value		value definition wordnet wordlist morphologicalLexicon other
examples		
DCLINK		
ISOcatLINK		

10.2.5 LanguageDescriptionPerformanceInfo

definition	Groups together information on the performance of the LanguageDescriptions
type	component
elements	robustness *Status: recommended* *Repeatability: 1* shallowness *Status: recommended*

	Repeatability: 1 output *Status: recommended* *Repeatability: 1*

1.1.1.152 robustness

definition	free text statement on the robustness of the grammar
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

1.1.1.153 shallowness

type	myString
definition	free text statement on the shallowness of the grammar
value space	
values	
examples	
DCLINK	
ISOcatLINK	

1.1.1.154 output

definition	whether the output is structures or a statement of grammaticality (grammatical/ungrammatical)
type	myString
value space	
values	
examples	
DCLINK	
ISOcatLINK	

10.3 TextInfo [restriction for language descriptions]

definition	restriction of TextInfo for LexicalConceptualResources & Language descriptions [no TextCreation, Annotation & TextClassification]

	N.B. The TextInfo of Lexical/Conceptual Resources includes in the current version mistakenly two optional components (TextCreationInfo and AnnotationInfo); these are not included in the restricted TextInfo of Language descriptions.
type	component
elements	LingualityInfo *Status: Mandatory* *Repeatability: 1* LanguageInfo *Status: Mandatory* *Repeatability: unbounded* SizeInfo[57] *Status: Mandatory* *Repeatability: unbounded* TextFormatInfo *Status: Optional* *Repeatability: unbounded* CharacterEncodingInfo *Status: Optional* *Repeatability: unbounded* DomainInfo *Status: Optional* *Repeatability: unbounded* TimeCoverageInfo *Status: Optional* *Repeatability: unbounded* GeographicCoverageInfo *Status: Optional* *Repeatability: unbounded*

11 Appendix - Bibliography

Broeder, D., T. Declerck, E. Hinrichs, S. Piperidis, L. Romary, N. Calzolari and P. Wittenburg, “Foundation of a Component-based Flexible Registry for Language

57 For language descriptions, the *sizeUnit* of *SizeInfo* contains an extra value, namely "rules".

Resources and Technology", *Proceedings of the 6th International Conference of Language Resources and Evaluation*, 2008.

Federmann, C., B. Georgantopoulos, R. del Gratta, O. Hamon, B. Magnini, D. Mavroeidis, S. Piperidis, M. Schroeder, M. Speranza, *META-NET Deliverable D7.1.1 – META-SHARE Functional and Technical Specification*, http://t4me.dfki.de/intranet/document_repository/deliverables/wp07-infrastructure-functional-and-technical-specification/meta-net-d7.1.1-final.pdf/view, 2011

Gavrilidou, M., P. Labropoulou, S. Piperidis, M. Speranza, M. Monachini, V. Arranz, G. Francopoulo, *META-NET Deliverable D7.2.1 – Specification of Metadata-Based Descriptions for Language Resources and Technologies,* http://t4me.dfki.de/intranet/document_repository/deliverables/wp07-infrastructure-functional-and-technical-specification/meta-net-d7.2.1-final.pdf/view, 2011.

ISO 12620, *Terminology and other language and content resources – Specification of data categories and management of a Data Category Registry for language resources*. http://www.isocat.org, 2009.