



**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: 06/03/2012



Table of Contents

1.	Executive Summary.....	14
2.	Introduction	14
3.	Model essentials	15
3.1	Basic concepts.....	15
3.2	The META-SHARE ontology	16
3.3	Proposed LR taxonomy.....	17
3.4	Basic contents & structure of the model.....	18
4.	Presentation structure and conventions	21
5	Special status components	22
5.1	personInfo.....	22
5.1.1	surname	23
5.1.2	givenName	23
5.1.3	sex	24
5.1.4	position	24
5.1.5	affiliation.....	24
5.2	organizationInfo	25
5.2.1	organizationName	25
5.2.2	organizationShortName.....	25
5.2.3	departmentName.....	26
5.3	communicationInfo	26
5.3.1	email.....	27
5.3.2	url.....	27
5.3.3	address.....	28
5.3.4	zipCode.....	28
5.3.5	city.....	28
5.3.6	region	29
5.3.7	country.....	29
5.3.8	telephoneNumber	29
5.3.9	faxNumber	30
5.4	actorInfo	30
5.5	projectInfo	30
5.5.1	projectName.....	31

5.5.2	projectShortName.....	31
5.5.3	projectID	32
5.5.4	fundingType.....	32
5.5.5	fundingCountry.....	32
5.5.6	funder.....	33
5.5.7	projectStartDate.....	33
5.5.8	projectEndDate.....	33
5.6	sizeInfo.....	34
5.6.1	size	34
5.6.2	sizeUnit	34
5.7	documentInfo	36
5.7.1	documentType	38
5.7.2	title	38
5.7.3	author.....	39
5.7.4	editor.....	39
5.7.5	year.....	39
5.7.6	publisher	40
5.7.7	bookTitle	40
5.7.8	journal.....	40
5.7.9	volume.....	41
5.7.10	series	41
5.7.11	pages	41
5.7.12	edition	41
5.7.13	conference.....	42
5.7.14	doi	42
5.7.15	ISSN	42
5.7.16	ISBN.....	43
5.7.17	keywords	43
5.7.18	documentLanguageName	43
5.7.19	documentLanguageId	43
5.8	documentationInfo	44
5.8.1	documentUnstructured	44
5.9	targetResourceInfo	44
5.9.1	targetResourceNameURI.....	45
6	resourceInfo – Common components.....	45
6.1	identificationInfo	46
6.1.1	resourceName	46
6.1.2	resourceShortName	47

6.1.3	description	47
6.1.4	metaShareId.....	48
6.1.5	identifier.....	48
6.2	distributionInfo.....	48
6.2.1	availability.....	49
6.2.1.1	availabilityStartDate	49
6.2.1.2	availabilityEndDate	50
6.2.2	licenceInfo.....	50
6.2.2.1	licence	51
6.2.2.2	restrictionsOfUse	52
6.2.2.3	userNature	53
6.2.2.4	membershipInfo	53
6.2.2.4.1	member	53
6.2.2.4.2	membershipInstitution.....	54
6.2.2.5	distributionAccessMedium.....	54
6.2.2.6	downloadLocation	55
6.2.2.7	executionLocation.....	55
6.2.2.8	price	55
6.2.2.9	attributionText.....	55
6.2.2.10	licensor.....	56
6.2.2.11	distributionRightsHolder	56
6.2.3	iprHolder	56
6.3	contactPerson	57
6.4	metadataInfo.....	57
6.4.1	metadataCreationDate.....	58
6.4.2	metadataCreator	58
6.4.3	source.....	58
6.4.4	originalMetadataSchema	59
6.4.5	originalMetadataLink	59
6.4.6	metadataLanguageName	59
6.4.7	metadataLanguageId	60
6.4.8	metadataLastDateUpdated.....	60
6.4.9	revision.....	61
6.5	versionInfo.....	61
6.5.1	version.....	61
6.5.2	lastDateUpdated	62
6.5.3	updateFrequency	62
6.6	validationInfo.....	62

6.6.1	validated.....	63
6.6.2	validationType	63
6.6.3	validationMode	64
6.6.4	validationModeDetails.....	64
6.6.5	validationExtent.....	65
6.6.6	validationExtentDetails	65
6.6.7	sizePerValidation	65
6.6.8	validationReport	65
6.6.9	validationTool	66
6.6.10	validator	66
6.7	usageInfo.....	66
6.7.1	accessTool	67
6.7.2	resourceAssociatedWith	67
6.7.3	foreseenUseInfo.....	67
6.7.3.1	foreseenUse.....	68
6.7.3.2	useNLPspecific	68
6.7.4	actualUseInfo.....	71
6.7.4.1	actualUse.....	71
6.7.4.2	usageReport	71
6.7.4.3	derivedResource	72
6.7.4.4	usageProject.....	72
6.7.4.5	actualUseDetails	72
6.8	resourceDocumentationInfo.....	72
6.8.1	documentation.....	73
6.8.2	samplesLocation	73
6.8.3	toolDocumentationType	73
6.9	resourceCreationInfo.....	74
6.9.1	resourceCreator	74
6.9.2	fundingProject	74
6.9.3	creationStartDate.....	75
6.9.4	creationEndDate.....	75
6.10	relationInfo	75
6.10.1	relationType.....	75
6.10.2	relatedResource	76
6.11	resourceComponentType.....	76
7	Corpora.....	76
7.1	resourceInfo.....	76
7.2	corpusInfo.....	77

7.2.1	resourceType.....	77
7.2.2	corpusMediaType	78
7.3	corpusTextInfo.....	78
7.3.1	mediaType.....	80
7.3.2	lingualityInfo	80
7.3.2.1	lingualityType	81
7.3.2.2	multilingualityType.....	81
7.3.2.3	multilingualityTypeDetails	81
7.3.3	languageInfo	82
7.3.3.1	languageId	82
7.3.3.2	languageName	82
7.3.3.3	languageScript	83
7.3.3.4	sizePerLanguage	83
7.3.3.5	languageVarietyInfo.....	83
7.3.3.5.1	languageVarietyType	84
7.3.3.5.2	languageVarietyName.....	84
7.3.3.5.3	sizePerLanguageVariety.....	84
7.3.4	modalityInfo	84
7.3.4.1	modalityType	85
7.3.4.2	modalityTypeDetails.....	85
7.3.4.3	sizePerModality	86
7.3.5	textFormatInfo.....	86
7.3.5.1	mimeType	86
7.3.5.2	sizePerTextFormat.....	86
7.3.6	characterEncodingInfo	87
7.3.6.1	characterEncoding.....	87
7.3.6.2	sizePerCharacterEncoding.....	87
7.3.7	annotationInfo	87
7.3.7.1	annotationType.....	89
7.3.7.2	annotatedElements	91
7.3.7.3	annotationStandoff.....	91
7.3.7.4	segmentationLevel.....	92
7.3.7.5	annotationFormat.....	92
7.3.7.6	tagset.....	93
7.3.7.7	tagsetLanguageId.....	93
7.3.7.8	tagsetLanguageName.....	93
7.3.7.9	conformanceToStandardsBestPractices	93
7.3.7.10	theoreticModel.....	95
7.3.7.11	annotationManual	95

7.3.7.12	annotationMode.....	96
7.3.7.13	annotationModeDetails	96
7.3.7.14	annotationTool	96
7.3.7.15	annotationStartDate	97
7.3.7.16	annotationEndDate	97
7.3.7.17	sizePerAnnotation.....	97
7.3.7.18	interannotatorAgreement.....	97
7.3.7.19	intraannotatorAgreement.....	98
7.3.7.20	annotator.....	98
7.3.8	domainInfo	98
7.3.8.1	domain	99
7.3.8.2	sizePerDomain.....	99
7.3.8.3	conformanceToClassificationScheme.....	99
7.3.9	textClassificationInfo	100
7.3.9.1	textGenre	100
7.3.9.2	textType	100
7.3.9.3	register	101
7.3.9.4	subject_topic.....	101
7.3.9.5	sizePerTextClassification	101
7.3.10	timeCoverageInfo	101
7.3.10.1	timeCoverage	102
7.3.10.2	sizePerTimeCoverage.....	102
7.3.11	geographicCoverageInfo	102
7.3.11.1	geographicCoverage.....	102
7.3.11.2	sizePerGeographicCoverage	103
7.3.12	creationInfo	103
7.3.12.1	originalSource.....	103
7.3.12.2	creationMode.....	104
7.3.12.3	creationModeDetails.....	104
7.3.12.4	creationTool	104
7.3.13	linkToOtherMediaInfo.....	105
7.3.13.1	otherMedia.....	106
7.3.13.2	mediaTypeDetails	106
7.3.13.3	synchronizedWithText.....	106
7.3.13.4	synchronizedWithAudio	107
7.3.13.5	synchronizedWithVideo.....	107
7.3.13.6	synchronizedWithImage.....	107
7.3.13.7	synchronizedWithTextNumerical.....	107
7.4	corpusAudioInfo	108

7.4.1	audioSizeInfo	109
7.4.1.1	durationOfEffectiveSpeechInfo	110
7.4.1.1.1	durationUnit	110
7.4.1.2	durationOfAudioInfo	110
7.4.2	audioContentInfo.....	111
7.4.2.1	speechItems	111
7.4.2.2	nonSpeechItems	112
7.4.2.3	textualDescription	112
7.4.2.4	noiseLevel	113
7.4.3	settingInfo.....	113
7.4.3.1	naturality.....	114
7.4.3.2	conversationalType.....	114
7.4.3.3	scenarioType	114
7.4.3.4	audience	115
7.4.3.5	interactivity.....	115
7.4.3.6	interaction.....	116
7.4.4	audioFormatInfo.....	116
7.4.4.1	signalEncoding.....	117
7.4.4.2	samplingRate	118
7.4.4.3	quantization	118
7.4.4.4	byteOrder	118
7.4.4.5	signConvention	119
7.4.4.6	audioQualityMeasuresIncluded.....	119
7.4.4.7	numberOfTracks	119
7.4.4.8	recordingQuality	120
7.4.4.9	sizePerAudioFormat	120
7.4.4.10	compressionInfo	120
7.4.4.10.1	compression	121
7.4.4.10.2	compressionName	121
7.4.4.10.3	compressionLoss.....	121
7.4.5	audioClassificationInfo	122
7.4.5.1	audioGenre.....	122
7.4.5.2	speechGenre.....	123
7.4.5.3	sizePerAudioClassification	124
7.4.6	recordingInfo	124
7.4.6.1	recordingDeviceType	125
7.4.6.2	recordingDeviceTypeDetails.....	125
7.4.6.3	recordingPlatformSoftware	125
7.4.6.4	recordingEnvironment	126

7.4.6.5	sourceChannel	126
7.4.6.6	sourceChannelType	127
7.4.6.7	sourceChannelName	127
7.4.6.8	sourceChannelDetails	127
7.4.6.9	recorder	128
7.4.7	captureInfo	128
7.4.7.1	capturingDeviceType	129
7.4.7.2	capturingDeviceTypeDetails	129
7.4.7.3	capturingDetails	130
7.4.7.4	capturingEnvironment	130
7.4.7.5	sensorTechnology	130
7.4.7.6	sceneIllumination	130
7.4.7.7	personSourceSetInfo	131
7.4.7.7.1	numberOfPersons	132
7.4.7.7.2	ageOfPersons	132
7.4.7.7.3	ageRangeStart	133
7.4.7.7.4	ageRangeEnd	133
7.4.7.7.5	sexOfPersons	133
7.4.7.7.6	originOfPersons	134
7.4.7.7.7	dialectAccentOfPersons	134
7.4.7.7.8	geographicDistributionOfPersons	134
7.4.7.7.9	hearingImpairmentOfPersons	135
7.4.7.7.10	speakingImpairmentOfPersons	135
7.4.7.7.11	numberOfTrainedSpeakers	135
7.4.7.7.12	speechInfluences	136
7.4.7.7.13	participantInfo	136
7.4.7.7.13.1	alias	138
7.4.7.7.13.2	ageGroup	138
7.4.7.7.13.3	age	138
7.4.7.7.13.4	origin	139
7.4.7.7.13.5	placeOfLiving	139
7.4.7.7.13.6	placeOfBirth	139
7.4.7.7.13.7	placeOfChildhood	139
7.4.7.7.13.8	dialectAccent	140
7.4.7.7.13.9	speakingImpairment	140
7.4.7.7.13.10	hearingImpairment	140
7.4.7.7.13.11	smokingHabits	140
7.4.7.7.13.12	vocalTractConditions	141
7.4.7.7.13.13	profession	141
7.4.7.7.13.14	height	141

7.4.7.7.13.15	weight.....	142
7.4.7.7.13.16	trainedSpeaker.....	142
7.4.7.7.13.17	placeOfSecondEducation.....	142
7.4.7.7.13.18	educationLevel.....	142
7.5	corpusVideoInfo	143
7.5.1	videoContentInfo	144
7.5.1.1	typeOfVideoContent	145
7.5.1.2	textIncludedInVideoType	145
7.5.1.3	dynamicElementInfo	145
7.5.1.3.1	typeOfElement	147
7.5.1.3.2	bodyParts	147
7.5.1.3.3	distractors	147
7.5.1.3.4	interactiveMedia	148
7.5.1.3.5	faceViews.....	148
7.5.1.3.6	faceExpressions.....	148
7.5.1.3.7	bodyMovement	149
7.5.1.3.8	gestures	149
7.5.1.3.9	handArmMovement.....	149
7.5.1.3.10	handManipulation	149
7.5.1.3.11	headMovement	150
7.5.1.3.12	eyeMovement.....	150
7.5.1.3.13	posesPerSubject	150
7.5.2	videoFormatInfo	151
7.5.2.1	colourSpace.....	152
7.5.2.2	colourDepth	152
7.5.2.3	frameRate.....	152
7.5.2.4	resolutionInfo	153
7.5.2.4.1	sizeWidth	153
7.5.2.4.2	sizeHeight.....	153
7.5.2.4.3	resolutionStandard	154
7.5.2.5	visualModelling.....	154
7.5.2.6	fidelity	154
7.5.2.7	sizePerVideoFormat.....	155
7.5.3	videoClassificationInfo	155
7.5.3.1	videoGenre.....	155
7.5.3.2	sizePerVideoClassification.....	156
7.6	corpusImageInfo.....	156
7.6.1	imageContentInfo	157
7.6.1.1	typeOfImageContent.....	158

7.6.1.2	textIncludedInImage	158
7.6.1.3	StaticElementInfo	158
7.6.1.3.1	artifactParts.....	159
7.6.1.3.2	landscapeParts	160
7.6.1.3.3	personDescription.....	160
7.6.1.3.4	thingDescription	160
7.6.1.3.5	organizationDescription	160
7.6.1.3.6	eventDescription	161
7.6.2	imageFormatInfo	161
7.6.2.1	rasterOrVectorGraphics.....	162
7.6.2.2	quality	162
7.6.2.3	sizePerImageFormat.....	163
7.6.3	imageClassificationInfo	163
7.6.3.1	imageGenre.....	163
7.6.3.2	sizePerImageClassification	163
7.7	corpusTextNumericalInfo.....	164
7.7.1	textNumericalContentInfo.....	165
7.7.1.1	typeOfTextNumericalContent	165
7.7.2	textNumericalFormatInfo.....	165
7.7.2.1	sizePerTextNumericalFormat.....	165
7.8	corpusTextNgramInfo	166
7.8.1	ngramInfo	167
7.8.1.1	baseItem.....	168
7.8.1.2	order.....	168
7.8.1.3	perplexity	169
7.8.1.4	isFactored.....	169
7.8.1.5	factors	169
7.8.1.6	smoothing	170
7.8.1.7	interpolated.....	170
8	Tools/Services	170
8.1	resourceInfo.....	170
8.2	toolServiceInfo.....	171
8.2.1	toolServiceType.....	172
8.2.2	toolServiceSubtype	172
8.2.3	languageDependent	173
8.2.4	inputInfo	173
8.2.4.1	resourceType.....	174
8.2.5	outputInfo	175
8.2.6	toolServiceOperationInfo	176

8.2.6.1	operatingSystem	176
8.2.6.2	runningTime	177
8.2.6.3	runningEnvironmentInfo	177
8.2.6.3.1	requiredSoftware	177
8.2.6.3.2	requiredHardware.....	178
8.2.6.3.3	requiredLRs	178
8.2.6.3.4	runningEnvironmentDetails.....	178
8.2.7	toolServiceEvaluationInfo	179
8.2.7.1	evaluated.....	180
8.2.7.2	evaluationLevel.....	180
8.2.7.3	evaluationType.....	180
8.2.7.4	evaluationCriteria	180
8.2.7.5	evaluationMeasure.....	181
8.2.7.6	evaluationReport	181
8.2.7.7	evaluationTool	181
8.2.7.8	evaluationDetails	182
8.2.7.9	evaluator	182
8.2.8	toolServiceCreationInfo	182
8.2.8.1	implementationLanguage.....	183
8.2.8.2	formalism.....	183
8.2.8.3	creationDetails.....	183
9	Language Descriptions	183
9.1	resourceInfo.....	183
9.2	languageDescriptionInfo	184
9.2.1	languageDescriptionType	185
9.2.2	languageDescriptionEncodingInfo	185
9.2.2.1	encodingLevel	186
9.2.2.2	task.....	186
9.2.2.3	grammaticalPhenomenaCoverage	187
9.2.2.4	weightedGrammar	188
9.2.3	languageDescriptionOperationInfo	188
9.2.3.1	relatedLexiconInfo.....	188
9.2.3.1.1	relatedLexiconType.....	189
9.2.3.1.2	attachedLexiconPosition.....	189
9.2.3.1.3	compatibleLexiconType	189
9.2.4	languageDescriptionPerformanceInfo	190
9.2.4.1	robustness.....	190
9.2.4.2	shallowness	191
9.2.4.3	output.....	191

9.2.5	languageDescriptionResourceMediaType	191
9.2.5.1	languageDescriptionTextInfo	192
9.2.5.2	languageDescriptionVideoInfo	193
9.2.5.3	languageDescriptionImageInfo	194
10	Lexical/conceptual resources	195
10.1	resourceInfo	195
10.2	lexicalConceptualResourceInfo	196
10.2.1	lexicalConceptualResourceType	196
10.2.2	lexicalConceptualResourceEncodingInfo	197
10.2.2.1	encodingLevel	197
10.2.2.2	linguisticInformation	198
10.2.2.3	externalRef	199
10.2.2.4	extratextualInformation	200
10.2.2.5	extratextualInformationUnit	200
10.2.3	lexicalConceptualResourceMediaType	201
10.2.3.1	lexicalConceptualResourceTextInfo	201
10.2.3.2	lexicalConceptualResourceAudioInfo	202
10.2.3.3	lexicalConceptualResourceVideoInfo	203
10.2.3.4	lexicalConceptualResourceImageInfo	204
AppendixI-	Bibliography	206

1. Executive Summary

The current document presents the metadata schema developed for META-SHARE, and more specifically the updated version 2.0. N.B. The document is considered as a living document inline with the META-SHARE platform, reflecting progress and changes as requested and necessary due to developments in the field.

The META-SHARE metadata schema aims at the description of Language Resources (LRs), catering in particular for the needs of the HLT community. The schema covers the following resource/media types:

- corpora (text, audio, video, multimodal/multimedia corpora incl. sensorimotor resources, n-gram resources),
- lexical/conceptual resources (e.g. computational lexica, ontologies, machine-readable dictionaries, terminological resources, thesauri, multimodal/multimedia lexica and dictionaries, etc.)
- language descriptions (e.g. computational grammars)
- technologies (tools/services) that can be used for the processing of data resources.

The present document is meant to act as a user manual, providing explanations on the model contents for LR providers and LR curators that wish to describe their resources in accordance to it.

More specifically, after a short introduction, section 3 contains an overview of the general principles of the model and section 4 focuses on the conventions used for the presentation of the model in the current document; the following sections present in detail the model contents, grouped by resource type.

2. Introduction

The current document reports on 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, and more specifically the updated version 2.0 thereof; it has been implemented as an XML schema (Appendix II).

A more detailed account of the theoretical principles and a general introduction to the model can be found in [[Gavrilidou et al., 2011](#)]& [[Labropoulou et al., 2012](#)].

3. Model essentials

3.1 Basic concepts

In the context of META-SHARE, the term **metadata** refers to descriptions of LR^s, 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).

The mechanism we have adopted is the **component-based mechanism** (Component MetaData Infrastructure, CMDI), according to which semantically coherent elements are grouped together to form components [[Broeder et al., 2008](#)].

More specifically, **elements** are used to encode specific descriptive features of the LR^s. To cater for semantic consistency with other related schemas and models, links to the conceptually same or similar existing elements in the Dublin Core (DC, www.dublincore.org) and the ISO Data Category Registry (ISO DCR, [[ISO 12620, 2009](#)]) are provided¹; where necessary, the new elements will populate the ISO DCR.

In addition, the notion of **relations** has been introduced for the encoding of linking features between resources. Relations hold between the various forms of a LR (e.g. raw and annotated resource), different LR^s (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.). Relations are represented as elements in the current schema version.

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 LR^s, 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

¹ Links are provided to the ISO DCR data categories as of 2012/02/07; as work is ongoing, these links may have to be revised and updated.

WP7 (see [[Federmann et al., 2011](#)]) concerning which features are considered sufficient to give a sound "identity" to a resource; discussions taken up within the extended metadata group² concerning the need of specific features have also fed the specifications of the minimal schema.

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 LR (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 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 LR).

Thus, the META-SHARE model recognizes the following distinct entities:

- the *resource* itself, i.e. the LR being described, encompassing datasets and technologies,
- the *actor*, further distinguished into *person* and *organization*
- the *project*,
- the *document* and
- the *licence*.

From these, however, the main interest of META-SHARE lies in *resources*, which form the main entity to be described; *actors*, *projects*, *documents* etc. are described when the case arises, i.e. when they are linked to a specific resource. Therefore, it is not expected from META-SHARE, for instance, to provide a bibliographical list of all documents that are relevant to the HLT domain but only those documents that are related to specific resources (e.g. articles describing the creation and/or use of a resource, documentation manuals, annotation guidelines etc.).

²The extended metadata group includes the core metadata group consisting of the WP7 contributors, metadata contributors from the collaborating PSP projects, experts from the LR community at large and the WP8 implementation group (cf. [[Labropoulou et al. 2012](#)] for a list of the involved persons).

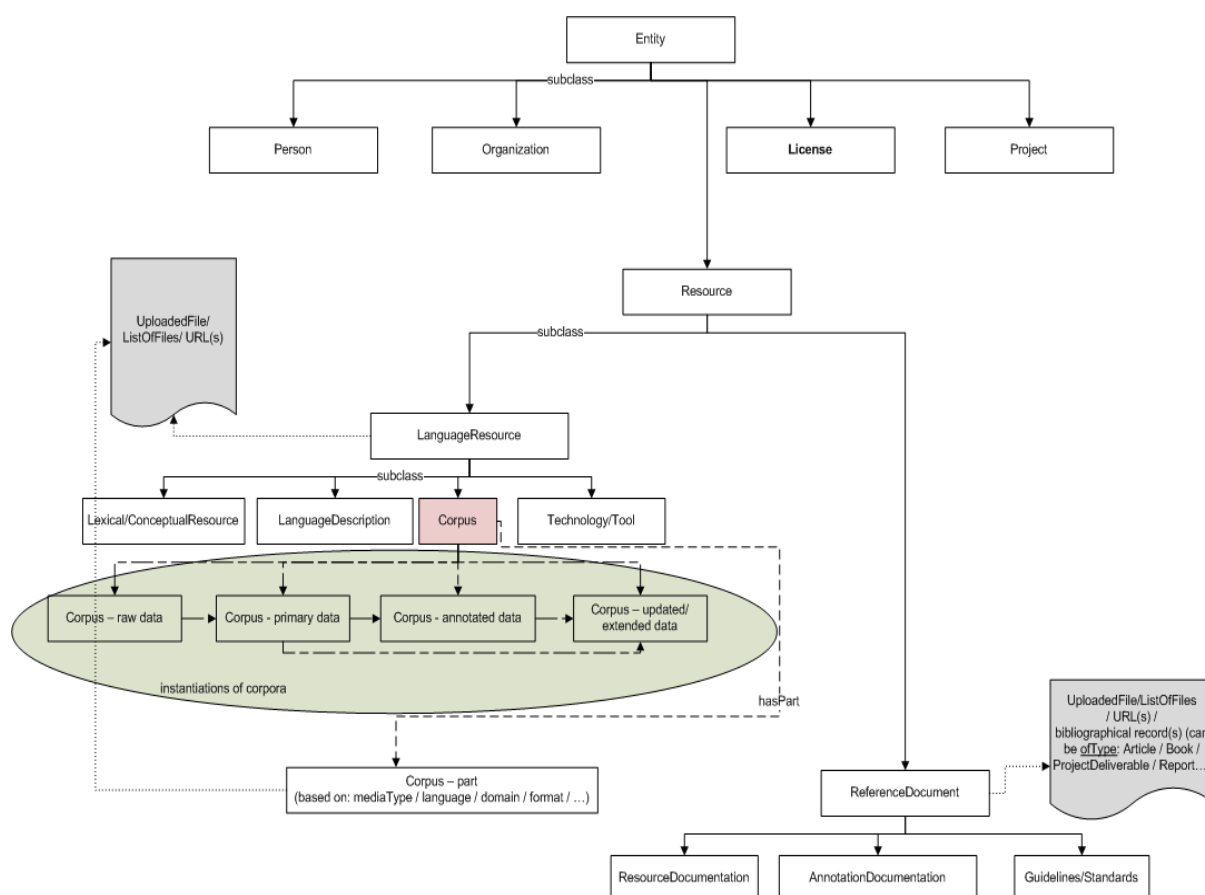


Figure 1: META-SHARE ontology excerpt

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)
- *tool / service* (including basic processing tools, applications, web services etc. required for processing data resources).

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 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*
- *video*
- *textNumerical* and
- *textNgram*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 may also include a video and/or an audio part; a sign language resource is also a resource with various media types (video, image, text). 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 *resourceInfo* 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 "content-wise" components that are idiosyncratic to a specific LR type.

The LR type-specific components are all located under the *resourceComponentType* component. Similarly, for each LR type, particular medium-dependent components are created to group together sets of features relevant to each LR_media type, given that media types and the recorded information for them differs across LR types; these are again grouped under an *xMediaType* component, where x stands for each of the LR type values.

³ It should be noted that the two values, *textNumerical* and *textNgram*, are not really distinct media types; they are actually subtypes of *text* but they present further particularities as regards their description due to their contents: numerical data (e.g. biometrical, geospatial data etc.) for the former and items with probability measures for the latter; this categorization allows us to better treat them in the metadata model.

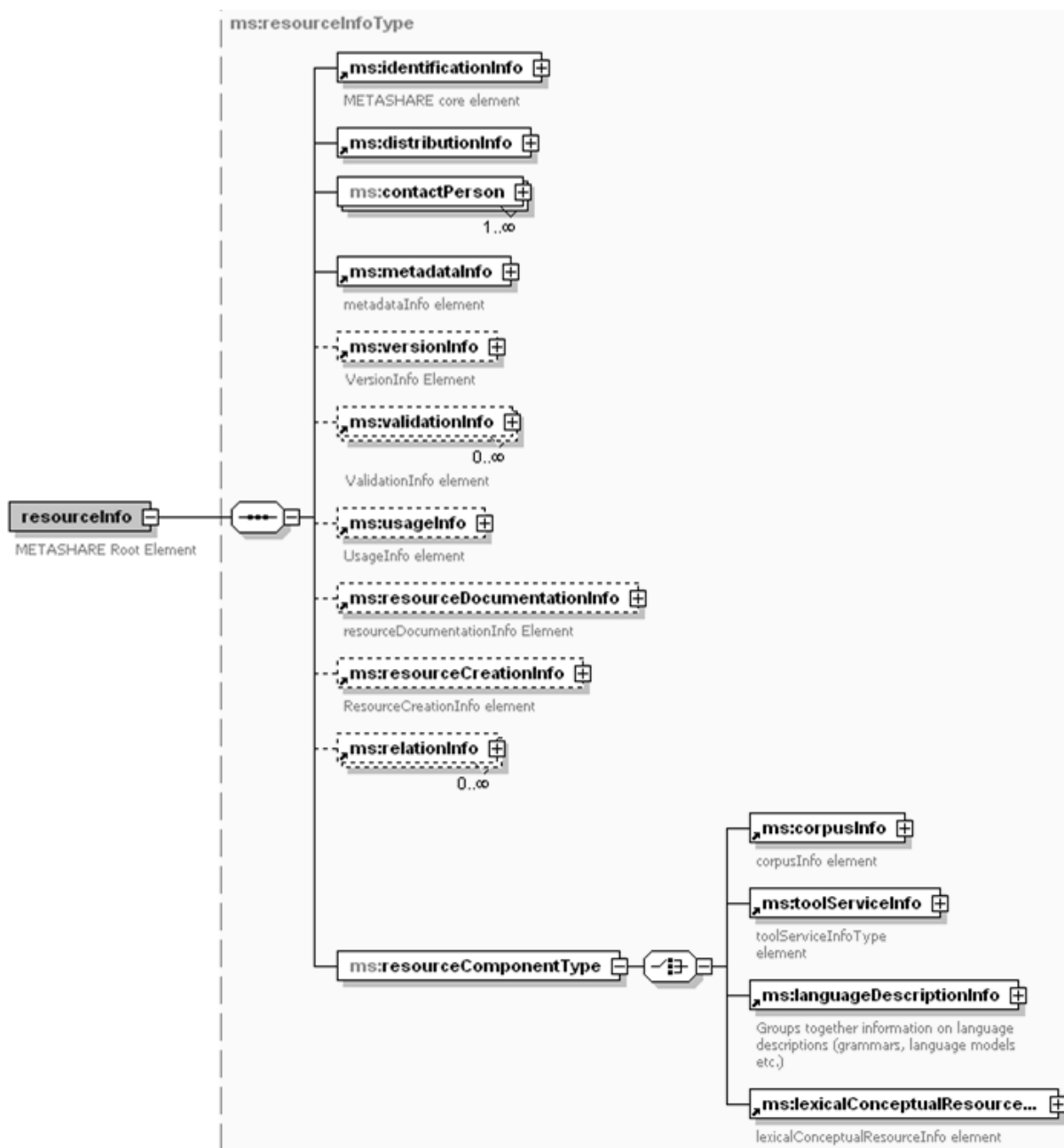


Figure 2: Common components for all LRs and resourceTypeInfo components

As aforesaid, the *resourceType* and *mediaType* elements encode the two classification axes of the schema, while each of the values of these two elements is associated with the appropriate component. Figure 3 shows how the resource-type and media-type components are put together, taking corpora as an example. The set of *resourceType* and *mediaType* components includes:

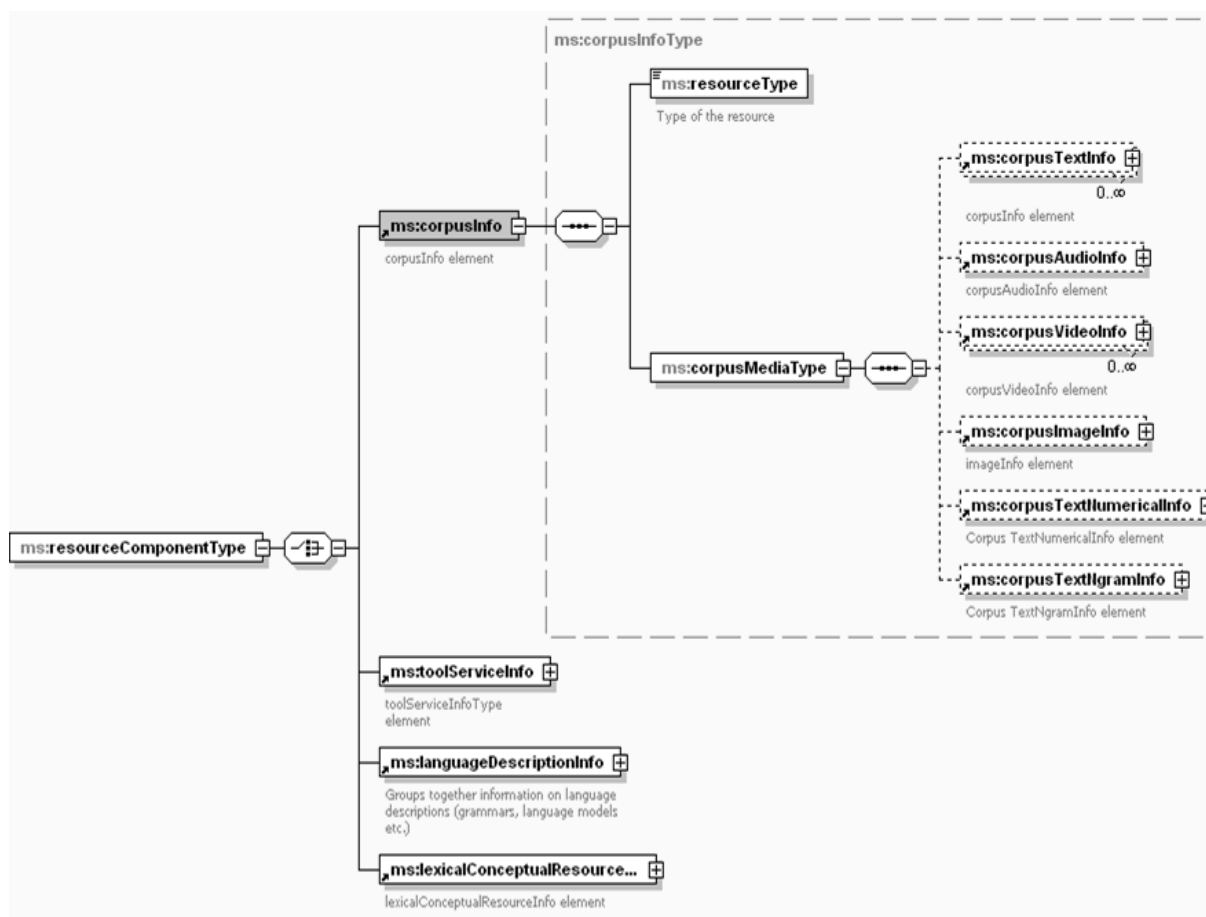


Figure 3: "Content-wise" components for corpora

- *corpusInfo*, *lexicalConceptualResourceInfo*, *languageDescriptionInfo*, *toolServiceInfo* include information specific to each LR type and take the values *corpus*, *lexical/conceptualResource*, *languageDescription* and *toolService* for the element *resourceType* respectively
- *corpusTextInfo*, *corpusAudioInfo*, *corpusVideoInfo*, *lexicalConceptualResourceTextInfo*, *lexicalConceptualResourceVideoInfo* etc. provide information depending on the media type of each LR type and include the *mediaType* element with the values *text*, *audio*, *video* etc. accordingly.

A set of six components enjoy a "special" status, in the sense that they can be attached to various components performing different roles, namely *personInfo*, *organizationInfo*, *communicationInfo*, *projectInfo*, *sizeInfo* and *documentInfo*. 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.

Finally, for the other entities of the model, i.e. those besides resources, special elements have been devised to allow for bulk encoding thereof independently of the resource to which these are linked. Thus, for instance, using such an element, a resource provider can upload the appropriate metadata for all persons at one go, and, then, while editing the metadata for the resources, create the appropriate links to the stored persons. These elements are not central to the model and are not explained in the current document.

4. Presentation structure and conventions

The following sections of the document present in detail the model. More specifically, the first section includes the "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, tools/services, language descriptions and, finally, lexical/conceptual resources.

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

- **definition:** a short statement explaining the semantics of the component in META-SHARE;
- **type:** it typically takes the value "component"; the value "special status component" is used for elements that are typed as such;
- **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*, *optional*) and repeatability (*1* for non-repeatable vs. *unbounded* for repeatable ones); if the element/component is condition-dependent⁴, 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
 - *xs:string*: free text
 - *xs:integer/int/double*: numerical fields
 - *xs:boolean*: yes/no

⁴ It should be noted that conditions are to be taken as guidelines for the metadata encoders and cannot all be automatically validated.

- *xs:date*: date, to be written according to the ISO-8601 standard
- *ms:myString*: free text in any language (the "lang" attribute must be used to specify the language of the text)
- *ms:emailAddress*: pattern of email addresses
- *ms:httpURI*: pattern of url's
- *ms:myStringURI*: either free text 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
- 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: a small list of possible values are provided for illustration purposes, especially in the case of free text elements
- DCLINK: the name of the corresponding element of the Dublin Core schema, provided for mapping purposes
- ISocatLINK: the name of the corresponding element of the ISocat DCR
- comments: used for notes where needed.

5 Special status components

5.1 personInfo

definition	Groups information relevant to persons related to the resource; to be used mainly for contact persons, resource creators, validators, annotators etc. for whom personal data can be provided
type	component
elements	surname <i>Status: mandatory</i>

	<i>Repeatability: unbounded</i> givenName <i>Status: recommended</i> <i>Repeatability: unbounded</i> sex <i>Status: recommended</i> <i>Repeatability: 1</i> position <i>Status: optional</i> <i>Repeatability: 1</i> communicationInfo <i>Status: mandatory</i> <i>Repeatability: 1</i> affiliation <i>Status: optional</i> <i>Repeatability: unbounded</i>
--	--

5.1.1 surname

definition	The surname (family name) of a person related to the resource
type	ms:myString
value space	
values	
examples	Smith; von Kamp; de Gruyter
DCLINK	
ISOcatLINK	http://www.isocat.org/datcat/DC-4195 ; ~ http://www.isocat.org/datcat/DC-2454 (the full name of a person)
comment	To be repeated only if provided in a different language

5.1.2 givenName

definition	The given name (first name) of a person related to the resource; initials can also be used
type	ms:myString
value space	
values	

examples	John; John Jr.; J.; John K.; J.K.
DCLINK	
ISocatLINK	http://www.isocat.org/datcat/DC-4194 ; ~ http://www.isocat.org/datecat/DC-2454 (the full name of a person)
comment	To be repeated only if provided in a different language

5.1.3 sex

definition	The gender of a person related to or participating in the resource	
type	closed controlled vocabulary	
value space	MS-sex	
values	value	definition
	male	
	female	
	unknown	
examples		
DCLINK		
ISocatLINK	http://www.isocat.org/datcat/DC-2560	

5.1.4 position

definition	The position or the title of a person if affiliated to an organization	
type	xs:string	
value space		
values		
examples	director; president of the organization; head of unit	
DCLINK		
ISocatLINK		

5.1.5 affiliation

definition	Groups information on organization to whom the person is affiliated	
type	special status component	
component	organizationInfo	

5.2 organizationInfo

definition	Groups information on organizations related to the resource
type	component
elements	<p>organizationName</p> <p><i>Status: mandatory</i></p> <p><i>Repeatability: unbounded</i></p> <p>organizationShortName</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>departmentName</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>communicationInfo</p> <p><i>Status: mandatory</i></p> <p><i>Repeatability: 1</i></p>

5.2.1 organizationName

definition	The full name of an organization
type	ms:myString
value space	
values	
examples	Charles University in Prague; Evaluations and Language Resources Distribution Agency; Deutsche Forschungszentrum für Künstliche Intelligenz; Jozef Stefan Institute; Fondazione Bruno Kessler
DCLINK	
ISOcatLINK	http://www.isocat.org/datcat/DC-2459
comment	To be repeated only if provided in a different language

5.2.2 organizationShortName

definition	The short name (abbreviation, acronym etc.) used for an organization
type	ms:myString
value space	
values	

examples	CUNI; ELDA; DFKI; JSI; FBK
DCLINK	
ISocatLINK	
comment	To be repeated only if provided in a different language.

5.2.3 departmentName

definition	The name of the department or unit (e.g. specific university faculty/department, department/unit of a research organization or private company etc.)
type	ms:myString
value space	
values	
examples	Institute of Formal and Applied Linguistics;Language Technology Lab
DCLINK	
ISocatLINK	
comment	To be repeated only if provided in a different language.

5.3 communicationInfo

definition	Groups information on communication details of a person or an organization
type	component
elements	<p>email</p> <p><i>Status: mandatory</i> <i>Repeatability: unbounded</i></p> <p>url</p> <p><i>Status: optional</i> <i>Repeatability: unbounded</i></p> <p>address</p> <p><i>Status: optional</i> <i>Repeatability: 1</i></p> <p>zipCode</p> <p><i>Status: optional</i> <i>Repeatability: 1</i></p> <p>city</p> <p><i>Status: optional</i></p>

	<i>Repeatability: 1</i> region <i>Status: optional</i> <i>Repeatability: 1</i> country <i>Status: optional</i> <i>Repeatability: 1</i> telephoneNumber <i>Status: optional</i> <i>Repeatability: unbounded</i> faxNumber <i>Status: optional</i> <i>Repeatability: unbounded</i>
--	--

5.3.1 email

definition	The email address of a person or an organization
type	ms:emailAddress
value space	
values	
examples	
DCLINK	
ISOcatLINK	http://www.isocat.org/datcat/DC-2521
comment	If it is unknown, please use the unknown@example.com invalid email address

5.3.2 url

definition	A URL used as homepage of an entity (e.g. of a person, organization, resource etc.) and/or where an entity (e.g.LR, document etc.) is located
type	ms:httpURI
value space	
values	
examples	http://www.meta-share.eu ; http://www.ilsp.gr/ ; http://www.natcorp.ox.ac.uk/ ; http://www.coli.uni-saarland.de/~hansu/ ; http://www.meta-net.eu/presentations/pdf/piperidis-et-al-may-2010-meta-share.pdf

DCLINK	
ISOcatLINK	http://www.isocat.org/datcat/DC-2546

5.3.3 address

definition	The street and the number of the postal address of a person or organization
type	xs:string
value space	
values	
examples	Wundtlaan 1; Artemidos 6 & Epidavrou
DCLINK	
ISOcatLINK	http://www.isocat.org/datcat/DC-2505

5.3.4 zipCode

definition	The zip code of the postal address of a person or organization
type	xs:string
value space	
values	
examples	
DCLINK	
ISOcatLINK	

5.3.5 city

definition	The name of the city, town or village as mentioned in the postal address of a person or organization
type	xs:string
value space	
values	
examples	
DCLINK	
ISOcatLINK	

5.3.6 region

definition	The name of the region, county or department as mentioned in the postal address of a person or organization
type	xs:string
value space	
values	
examples	
DCLINK	
ISocatLINK	
comment	~ http://www.isocat.org/datcat/DC-3814 (refers to the region or sub-region such as of a speech area, (federal) state, etc.)

5.3.7 country

definition	The name of the country mentioned in the postal address of a person or organization as defined in the list of values of ISO 3166
type	closed controlled vocabulary
value space	ISO 3166 , Codes for the representation of names of countries and their subdivisions
values	
examples	
DCLINK	
ISocatLINK	~ http://www.isocat.org/datcat/DC-3792
comment	

5.3.8 telephoneNumber

definition	The telephone number of a person or an organization; recommended format: +_international code_city code_number
type	xs:string
value space	
values	
examples	
DCLINK	
ISocatLINK	http://www.isocat.org/datcat/DC-2461

5.3.9 faxNumber

definition	The fax number of a person or an organization; recommended format: +_international code_city code_number
type	xs:string
value space	
values	
examples	
DCLINK	
ISOcatLINK	http://www.isocat.org/datcat/DC-2455

5.4 actorInfo

definition	Used to bring persons and organizations (in whatever role they may have with regard to the resource, e.g., resource creator, IPR holder, etc.)
type	link component
component	choice personInfo organizationInfo

5.5 projectInfo

definition	Groups information on a project related to the resource (e.g. a project the resource has been used in; a funded project that led to the resource creation etc.)
type	component
elements	projectName <i>Status: mandatory</i> <i>Repeatability: unbounded</i> projectShortName <i>Status: optional</i> <i>Repeatability: unbounded</i> projectID <i>Status: optional</i> <i>Repeatability: 1</i> url <i>Status: optional</i> <i>Repeatability: unbounded</i> fundingType

	<i>Status: mandatory</i> <i>Repeatability: unbounded</i> fundingCountry <i>Status: recommended</i> <i>Repeatability: unbounded</i> funder <i>Status: recommended</i> <i>Repeatability: unbounded</i> projectStartDate <i>Status: optional</i> <i>Repeatability: 1</i> projectEndDate <i>Status: optional</i> <i>Repeatability: 1</i>
--	---

5.5.1 projectName

definition	The full name of a project related to the resource
type	ms:myString
value space	
values	
examples	Technologies for the Multilingual European Information Society
DCLINK	
ISocatLINK	http://www.isocat.org/datcat/DC-2537
comment	To be repeated only if provided in different language.

5.5.2 projectShortName

definition	A short name or abbreviation of a project related to the resource
type	ms:myString
value space	
values	
examples	
DCLINK	
ISocatLINK	http://www.isocat.org/datcat/DC-2536
comment	To be repeated only if provided in different language.

5.5.3 projectID

definition	An unambiguous referent to a project related to the resource
type	xs:string
value space	
values	
examples	
DCLINK	
ISOcatLINK	http://www.isocat.org/datcat/DC-2535

5.5.4 fundingType

definition	Specifies the type of funding of the project	
type	open controlled vocabulary	
value space	MS-fundingType	
values	value	definition
	euFunds	Funding coming from EU sources
	nationalFunds	Funding coming from national sources
	ownFunds	Funding from the resource creators's own sources
	other ⁵	When none of the above is applicable
examples		
DCLINK		
ISOcatLINK		
comment	change the order of values in the xsd	

5.5.5 fundingCountry

definition	The name of the funding country, in case of national funding as mentioned in ISO3166
type	controlled vocabulary
value space	ISO 3166 , Codes for the representation of names of countries and their subdivisions
values	

⁵ Implementation pending

examples	
DCLINK	
ISocatLINK	
comment	to be used only when fundingType=nationalFunds

5.5.6 funder

definition	The full name of the funder of the project
type	xs:string
value space	
values	
examples	
DCLINK	
ISocatLINK	http://www.isocat.org/datcat/DC-2522
comment	to be used when the funder of the project refers to an organization, company etc.

5.5.7 projectStartDate

definition	The starting date of a project related to the resource
type	xs:date
value space	
values	
examples	
DCLINK	
ISocatLINK	

5.5.8 projectEndDate

definition	The end date of a project related to the resources
type	xs:date
value space	
values	
examples	
DCLINK	

ISocatLINK	
------------	--

5.6 sizeInfo

definition	Groups information on the size of the resource or of resource parts
type	component
elements	size <i>Status: mandatory</i> <i>Repeatability: 1</i> sizeUnit <i>Status: mandatory</i> <i>Repeatability: 1</i>

5.6.1 size

definition	Specifies the size of the resource with regard to the SizeUnit measurement in form of a number
type	xs:integer
value space	
values	
examples	
DCLINK	
ISocatLINK	http://www.isocat.org/datcat/DC-2580
comment	check type in xsd

5.6.2 sizeUnit

definition	Specifies the unit that is used when providing information on the size of the resource or of resource parts										
type	open controlled vocabulary										
value space	MS-sizeUnit										
values	<table> <tr> <td>value</td><td>definition</td></tr> <tr> <td>entries</td><td></td></tr> <tr> <td>turns</td><td></td></tr> <tr> <td>utterances</td><td></td></tr> <tr> <td>articles</td><td></td></tr> </table>	value	definition	entries		turns		utterances		articles	
value	definition										
entries											
turns											
utterances											
articles											

	files
	items
	elements
	units
	seconds
	minutes
	hours
	texts
	sentences
	tokens
	words
	keywords
	idiomaticExpressions
	neologisms
	multiWordUnits
	expressions
	synsets
	classes
	concepts
	lexicalTypes
	phoneticUnits
	syntacticUnits
	semanticUnits
	predicates
	phonemes
	diphones
	T-HPairs
	rules
	syllables
	frames
	images
	bytes
	kb
	mb
	gb
	rb
	shots

	unigram bigram trigram 4-gram 5-gram n-gram other
examples	
DCLINK	
ISOcatLINK	~ http://www.isocat.org/datcat/DC-2583
comment	

5.7 documentInfo

definition	Groups information on all the documents reporting on various aspects of the resource (creation, usage etc.), published or unpublished; it is used in various places of the metadata schema depending on its role (e.g. usage report, validation report, annotation manual etc.)
type	component
elements	documentType <i>Status: mandatory</i> <i>Repeatability: 1</i> title <i>Status: mandatory</i> <i>Repeatability: unbounded</i> author <i>Status: optional</i> <i>Repeatability: unbounded</i> editor <i>Status: optional</i> <i>Repeatability: unbounded</i> year <i>Status: optional</i> <i>Repeatability: 1</i> publisher <i>Status: optional</i> <i>Repeatability: unbounded</i>

	bookTitle <i>Status: optional</i> <i>Repeatability: 1</i>
	journal <i>Status: optional</i> <i>Repeatability: 1</i>
	volume <i>Status: optional</i> <i>Repeatability: 1</i>
	series <i>Status: optional</i> <i>Repeatability: 1</i>
	pages <i>Status: optional</i> <i>Repeatability: 1</i>
	edition <i>Status: optional</i> <i>Repeatability: 1</i>
	conference <i>Status: optional</i> <i>Repeatability: 1</i>
	doi <i>Status: optional</i> <i>Repeatability: 1</i>
	url <i>Status: optional</i> <i>Repeatability: 1</i>
	ISSN <i>Status: optional</i> <i>Repeatability: 1</i>
	ISBN <i>Status: optional</i> <i>Repeatability: 1</i>
	keywords <i>Status: optional</i> <i>Repeatability: 1</i>
	documentLanguageName

	<i>Status: optional</i> <i>Repeatability: 1</i> documentLanguageId <i>Status: optional</i> <i>Repeatability: 1</i>
--	--

5.7.1 documentType

definition	Specifies the type of the document provided with or related to the resource	
type	open controlled vocabulary	
value space	MS-documentationType	
values	value	definition
	article	An article from a journal or magazine
	book	A book with an explicit publisher
	booklet	A work that is printed and bound, but without a named publisher or sponsoring institution
	manual	Technical documentation
	techReport	A report published by a school or other institution
	mastersThesis	A Master's thesis
	phdThesis	A PhD thesis
	inBook	A part of a book, which may be a chapter (or section) and/or a range of pages.
	inCollection	A part of a book having its own title
	proceedings	The proceedings of a conference
	inProceedings	An article in a conference proceedings
	unpublished	A document having an author and title, but not formally published
	other	Use this type when nothing else fits
examples		
DCLINK		
ISocatLINK		

5.7.2 title

definition	The title of the document reporting on the resource
type	ms:myString

value space	
values	
examples	
DCLINK	
ISOcatLINK	
comment	To be repeated only if provided in different language

5.7.3 author

definition	The name(s) of the author(s), in the format described in the document
type	xs:string
value space	
values	
examples	
DCLINK	
ISOcatLINK	

5.7.4 editor

definition	The name of the editor as mentioned in the document
type	xs:string
value space	
values	
examples	
DCLINK	
ISOcatLINK	
comment	If there is also an author field, then the editor field gives the editor of the book or collection in which the reference appears.

5.7.5 year

definition	The year of publication or, for an unpublished work, the year it was written
type	xs:date
value space	
values	

examples	
DCLINK	
ISOcatLINK	

5.7.6 publisher

definition	The name of the publisher
type	xs:string
value space	
values	
examples	
DCLINK	
ISOcatLINK	

5.7.7 bookTitle

definition	The title of a book, part of which is being cited
type	xs:string
value space	
values	
examples	
DCLINK	
ISOcatLINK	
comment	For book entries, use the title field instead

5.7.8 journal

definition	A journal name. Abbreviations could also be provided
type	xs:string
value space	
values	
examples	
DCLINK	
ISOcatLINK	

5.7.9 volume

definition	Specifies the volume of a journal or multivolume book
type	xs:string
value space	
values	
examples	
DCLINK	
ISocatLINK	

5.7.10 series

definition	The name of a series or set of books. When citing an entire book, the title field gives its title and an optional series field gives the name of a series or multi-volume set in which the book is published
type	xs:string
value space	
values	
examples	
DCLINK	
ISocatLINK	

5.7.11 pages

definition	One or more page numbers or range of page numbers
type	xs:string
value space	
values	
examples	7; 41; 73-97
DCLINK	
ISocatLINK	

5.7.12 edition

definition	The edition of a book
type	xs:string

value space	
values	
examples	Second
DCLINK	
ISOcatLINK	

5.7.13 conference

definition	The name of the conference in which the document has been presented
type	xs:string
value space	
values	
examples	LREC 2010; EACL 2012 Workshop on Semantic Analysis in Social Networks
DCLINK	
ISOcatLINK	

5.7.14 doi

definition	A digital object identifier assigned to the document
type	xs:string
value space	
values	
examples	
DCLINK	
ISOcatLINK	

5.7.15 ISSN

definition	The International Standard Serial Number used to identify a journal
type	xs:string
value space	
values	
examples	
DCLINK	
ISOcatLINK	

5.7.16 ISBN

definition	The International Standard Book Number
type	xs:string
value space	
values	
examples	
DCLINK	
ISocatLINK	

5.7.17 keywords

definition	The keyword(s) for indexing and classification of the document
type	xs:string
value space	
values	
examples	
DCLINK	
ISocatLINK	

5.7.18 documentLanguageName

definition	The name of the language the document is written in, as mentioned in IETF BCP47 ⁶
type	controlled vocabulary
value space	IETF BCP47, Codes for the representation of names of languages (http://tools.ietf.org/html/bcp47)
values	
examples	
DCLINK	
ISocatLINK	

5.7.19 documentLanguageId

definition	The id of the language the document is written in, as mentioned in IETF
------------	---

⁶ To be implemented in next version

	BCP47 ⁷
type	controlled vocabulary
value space	IETF BCP47, Codes for the representation of names of languages (http://tools.ietf.org/html/bcp47)
values	
examples	
DCLINK	
ISOcatLINK	

5.8 documentationInfo

definition	Groups information on the documentation of the resource pointing to either structured and unstructured presentation of the relevant documents
type	component
components	choice documentUnstructured documentInfo

5.8.1 documentUnstructured

definition	Provides information on documents that report on the resource in form of free text
type	xs:string
value space	
values	
examples	
DCLINK	
ISOcatLINK	

5.9 targetResourceInfo

definition	Groups information on the resource related to the one being described; can be an identifier, a resource name or a URL
type	component
elements	targetResourceNameURI ⁸

⁷ To be implemented in next version

⁸ Implementation pending: reference to resource ID until identifier issues are resolved; for the time being, the relatedResourceNameURI can be used also for adding identifiers.

5.9.1 targetResourceNameURI

definition	The full name or a url to a resource related to the one being described; to be used for identifiers also for this version
type	ms:myStringURI
value space	
values	
examples	
DCLINK	
ISocatLINK	

6 resourceInfo – Common components

definition	Groups together all information required for the description of language resources
elements	<p>identificationInfo <i>Status: mandatory</i> <i>Repeatability: 1</i></p> <p>distributionInfo <i>Status: mandatory</i> <i>Repeatability: 1</i></p> <p>contactPerson <i>Status: mandatory</i> <i>Repeatability: unbounded</i></p> <p>metadataInfo <i>Status: mandatory</i> <i>Repeatability: 1</i></p> <p>versionInfo <i>Status: recommended</i> <i>Repeatability: 1</i></p> <p>validationInfo <i>Status: recommended</i> <i>Repeatability: unbounded</i></p> <p>usageInfo <i>Status: recommended</i> <i>Repeatability: 1</i></p>

	resourceDocumentationInfo <i>Status: recommended</i> <i>Repeatability: 1</i> resourceCreationInfo <i>Status: recommended</i> <i>Repeatability: 1</i> relationInfo <i>Status: recommended</i> <i>Repeatability: unbounded</i>
--	--

6.1 identificationInfo

definition	Groups together information needed to identify the resource
type	component
elements	resourceName <i>Status: mandatory</i> <i>Repeatability: unbounded</i> resourceShortName <i>Status: optional</i> <i>Repeatability: unbounded</i> description <i>Status: mandatory</i> <i>Repeatability: unbounded</i> url <i>Status: recommended</i> <i>Repeatability: unbounded</i> metaShareId <i>Status: mandatory</i> <i>Repeatability: 1</i> identifier <i>Status: optional</i> <i>Repeatability: unbounded</i>

6.1.1 resourceName

definition	The full name by which the resource is known
type	ms:myString

value space	
values	
examples	British National Corpus; Penn Treebank; EXMARaLDA
DCLINK	dc:title
ISOcatLINK	http://www.isocat.org/datcat/DC-2545
comment	If there are two forms (a full and a short name), please use this for the full name and "resourceShortName" for the short alternative; The element could be repeated only if provided in different language.

6.1.2 resourceShortName

definition	The short form (abbreviation, acronym etc.) used to identify the resource
type	ms:myString
value space	
values	
examples	BNC; PTB; HNC
DCLINK	
ISOcatLINK	http://www.isocat.org/datcat/DC-2544
comment	To be repeated only if provided in different language

6.1.3 description

definition	Provides the description of the resource in prose
type	ms:myString
value space	
values	
examples	
DCLINK	dc:description
ISOcatLINK	~ http://www.isocat.org/datcat/DC-2520 (A description in general prose text of the issues that are indicated by the context; it can occur at many different places in a component and profile)
comment	To be repeated only if provided in different language

6.1.4 metaShareId

definition	An unambiguous referent to the resource within META-SHARE
type	xs:string
value space	
values	
examples	
DCLINK	
ISocatLINK	

6.1.5 identifier

definition	A reference to the resource like a pid or an internal identifier used by the resource provider; the attribute "type" is obligatorily used for further specification ⁹
type	xs:string
value space	
examples	
DCLINK	dc:identifier
ISocatLINK	~ http://www.isocat.org/datcat/DC-2573 (refers to pids only)

6.2 distributionInfo

definition	Groups information on the distribution of the resource
type	component
elements	<p>availability</p> <p><i>Status: mandatory</i></p> <p><i>Repeatability: 1</i></p> <p>availabilityStartDate</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p>availabilityEndDate</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p>licenceInfo</p> <p><i>Status: condition-dependent mandatory</i></p>

⁹ The attribute "type" is not included in the current XSD's and the editor.

	<p><i>Condition:</i> <i>availability="available-restrictedUse" or "available-unrestrictedUse"</i></p> <p><i>Repeatability:</i> <i>unbounded</i></p> <p>iprHolder</p> <p><i>Status:</i> <i>optional</i></p> <p><i>Repeatability:</i> <i>unbounded</i></p>
--	--

6.2.1 availability

definition	Specifies the 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	For resources available with no restrictions of use
	available-restrictedUse	For resources available but with restrictions of uses imposed by the type of licence
	notAvailableThroughMetaShare	For resources that are not available through META-SHARE
	underNegotiation	For resources that their availability is still pending
examples		
DCLINK	dc:rights	
ISOcatLINK	~ http://www.isocat.org/datcat/DC-2453 (different values)	

6.2.1.1 availabilityStartDate

definition	Specifies the start date of availability of a resource
type	xs:date
value space	
values	
examples	
DCLINK	dc:date
ISOcatLINK	

6.2.1.2 availabilityEndDate

definition	Specifies the end date of availability of a resource
type	xs:date
value space	
values	
examples	
DCLINK	dc:date
ISocatLINK	

6.2.2 licenceInfo

definition	Groups information on licences 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	<p>licence</p> <p><i>Status: mandatory</i></p> <p><i>Repeatability: unbounded</i></p> <p>restrictionsOfUse</p> <p><i>Status: condition-dependent mandatory</i></p> <p><i>Condition: availability="available-restrictedUse"</i></p> <p><i>Repeatability: unbounded</i></p> <p>userNature</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>membershipInfo</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>distributionAccessMedium</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>downloadLocation</p> <p><i>Status: condition-dependent mandatory</i></p> <p><i>Condition: disrtibutionAccessMedium="downloadable"</i></p> <p><i>Repeatability: unbounded</i></p> <p>executionLocation</p>

	<p><i>Status: condition-dependent mandatory</i></p> <p><i>Condition: distributionAccessMedium="accessibleThroughInterface" or "webExecutable"</i></p> <p><i>Repeatability: unbounded</i></p> <p>price</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p>attributionText</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>licensor</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>distributionRightsHolder</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p>
--	--

6.2.2.1 licence

definition	The licence of use for the resource	
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 underNegotiation other
examples	
DCLINK	dc:rights
ISOcatLINK	http://www.isocat.org/datcat/DC-2457

6.2.2.2 restrictionsOfUse

definition	Specifies the restrictions imposed by the type of the licence	
type	open controlled vocabulary	
value space	MS-restrictionsOfUse	
values	value informLicensor redeposit onlyMSmembers academic-nonCommercialUse evaluationUse commercialUse attribution shareAlike noDerivatives noRedistribution other	definition
examples		
DCLINK	dc:rights	

ISOcatLINK	
------------	--

6.2.2.3 userNature

definition	The conditions imposed by the nature of the user (for instance, a research use may have different implications depending on this)	
type	closed controlled vocabulary	
value space	MS-userNature	
values	value	definition
	academic	For academic institutions
	commercial	For commercial institutions
examples		
DCLINK		
ISOcatLINK		

6.2.2.4 membershipInfo

definition	The conditions imposed by the user being member of some association/institution (e.g., ELRA, LDC) distributing the resource. This indicates the availability conditions (and prices) for users who are members or not.	
type	component	
elements	member <i>Status: mandatory</i> <i>Repeatability: 1</i> membershipInstitution <i>Status: mandatory</i> <i>Repeatability: unbounded</i>	

6.2.2.4.1 member

definition	Whether the user is a member or not
type	xs:boolean
value space	
values	
examples	
DCLINK	
ISOcatLINK	

6.2.2.4.2 membershipInstitution

definition	This lists the different institutions releasing the resources and establishing membership conditions										
type	open controlled vocabulary										
value space	MS-membershipInstitution										
values	<table> <tr> <th>value</th><th>definition</th></tr> <tr> <td>ELRA</td><td></td></tr> <tr> <td>LDC</td><td></td></tr> <tr> <td>TST-Centrale</td><td></td></tr> <tr> <td>other</td><td></td></tr> </table>	value	definition	ELRA		LDC		TST-Centrale		other	
value	definition										
ELRA											
LDC											
TST-Centrale											
other											
examples											
DCLINK											
ISocatLINK											

6.2.2.5 distributionAccessMedium

definition	Specifies the medium (channel) used for delivery or providing access to the resource																				
type	open controlled vocabulary																				
value space	MS-distributionAccessMedium																				
values	<table> <tr> <th>value</th><th>definition</th></tr> <tr> <td>webExecutable</td><td></td></tr> <tr> <td>paperCopy</td><td></td></tr> <tr> <td>hardDisk</td><td></td></tr> <tr> <td>bluRay</td><td></td></tr> <tr> <td>DVD-R</td><td></td></tr> <tr> <td>CD-ROM</td><td></td></tr> <tr> <td>downloadable</td><td></td></tr> <tr> <td>accessibleThroughInterface</td><td></td></tr> <tr> <td>other</td><td></td></tr> </table>	value	definition	webExecutable		paperCopy		hardDisk		bluRay		DVD-R		CD-ROM		downloadable		accessibleThroughInterface		other	
value	definition																				
webExecutable																					
paperCopy																					
hardDisk																					
bluRay																					
DVD-R																					
CD-ROM																					
downloadable																					
accessibleThroughInterface																					
other																					
examples																					
DCLINK																					
ISocatLINK	http://www.isocat.org/datcat/DC-2458																				

6.2.2.6downloadLocation

definition	Any url where the resource can be downloaded from
type	ms:httpURI
value space	
values	
examples	
DCLINK	
ISocatLINK	

6.2.2.7executionLocation

definition	Any url where the service providing access to a resource is being executed
type	ms:httpURI
value space	
values	
examples	
DCLINK	
ISocatLINK	http://www.isocat.org/datcat/DC-2569

6.2.2.8price

definition	Specifies the costs that are required to access the resource, a fragment of the resource or to use a tool or service
type	xs:string
value space	
values	
examples	
DCLINK	
ISocatLINK	http://www.isocat.org/datcat/DC-2460

6.2.2.9attributionText

definition	The text that must be quoted for attribution purposes when using a resource
type	ms:myString
value space	
values	
examples	

DCLINK	
ISOcatLINK	

6.2.2.10 licensor

definition	Groups information on person who is legally eligible to licence and actually licenses the resource. The licensor could be different from the creator, the distributor or the IP rightsholder. The licensor has the necessary rights or licences to license the work and is the party that actually licenses the resource that enters the META-SHARE network. She will have obtained the necessary rights or licences from the IPR holder and she may have a distribution agreement with a distributor that disseminates the work under a set of conditions defined in the specific licence and collects revenue on the licensor's behalf. The attribution of the creator, separately from the attribution of the licensor, may be part of the licence under which the resource is distributed (as e.g. is the case with Creative Commons Licences).
type	special status component
component	personInfo

6.2.2.11 distributionRightsHolder

definition	Groups information on a person or an organization that holds the distribution rights. The range and scope of distribution rights is defined in the distribution agreement. The distributor in most cases only has a limited licence to distribute the work and collect royalties on behalf of the licensor or the IPR holder and cannot give to any recipient of the work permissions that exceed the scope of the distribution agreement (e.g. to allow uses of the work that are not defined in the distribution agreement)
type	special status component
component	actorInfo
DCLINK	~dc:publisher
ISOcatLINK	

6.2.3 iprHolder

definition	Groups information on a person or an organization who holds the full Intellectual Property Rights (Copyright, trademark etc) that subsist in the
------------	--

	resource. The IPR holder could be different from the creator that may have assigned the rights to the IPR holder (e.g. an author as a creator assigns her rights to the publisher who is the IPR holder) and the distributor that holds a specific licence (i.e. a permission) to distribute the work within the META-SHARE network.
type	linked component
component	actorInfo

6.3 contactPerson

definition	Groups information on the person(s) that is responsible for giving information for the resource
type	linked component
component	personInfo
DCLINK	
ISOcatLINK	

6.4 metadataInfo

definition	Groups information on the metadata record itself
type	component
elements	<p>metadataCreationDate <i>Status: mandatory</i> <i>Repeatability: 1</i></p> <p>metadataCreator <i>Status: optional</i> <i>Repeatability: unbounded</i></p> <p>source <i>Status: optional</i> <i>Repeatability: 1</i></p> <p>originalMetadataSchema <i>Status: optional</i> <i>Repeatability: 1</i></p> <p>originalMetadataLink <i>Status: optional</i> <i>Repeatability: 1</i></p> <p>metadataLanguageName</p>

	<i>Status: optional</i> <i>Repeatability: unbounded</i> metadataLanguageId <i>Status: optional</i> <i>Repeatability: unbounded</i> metadataLastDateUpdated <i>Status: optional</i> <i>Repeatability: 1</i> revision <i>Status: optional</i> <i>Repeatability: 1</i>
--	---

6.4.1 metadataCreationDate

definition	The date of creation of this metadata description
type	xs:date
value space	
values	
examples	
DCLINK	dc:date
ISocatLINK	http://www.isocat.org/datcat/DC-2541
comment	if the metadata record has resulted from conversion after harvesting, this is also considered the metadataCreationDate

6.4.2 metadataCreator

definition	Groups information on the person that created the metadata in META-SHARE editor
type	linked component
component	personInfo
DCLINK	~dc:contributor
ISocatLINK	~http://www.isocat.org/datcat/DC-2542 (it refers to the name of the metadata creator)

6.4.3 source

definition	Refers to the catalogue or repository from which the metadata has been
------------	--

	originated.
type	ms:myStringURI
value space	
values	
examples	OLAC; CLARIN; ELDA; LDC
DCLINK	
ISocatLINK	
comment	open issue: value to be assigned automatically depending on where the metadata has been harvested from

6.4.4 originalMetadataSchema

definition	Refers to the metadata schema originally used for the description of the resource
type	xs:string
value space	
values	
examples	
DCLINK	
ISocatLINK	

6.4.5 originalMetadataLink

definition	A link to the metadata of the original source
type	ms:httpURI
value space	
values	
examples	
DCLINK	
ISocatLINK	
comment	to be automatically assigned in harvesting process

6.4.6 metadataLanguageName

definition	The name of the language in which the metadata description is written,
------------	--

	according to IETF BCP47 ¹⁰
type	controlled vocabulary
value space	IETF BCP47, Codes for the representation of names of languages (http://tools.ietf.org/html/bcp47)
values	
examples	
DCLINK	
ISOcatLINK	http://www.isocat.org/datcat/DC-2543
comment	for current version, default value should be English

6.4.7 metadataLanguageId

definition	The identifier of the language in which the metadata description is written according to IETF BCP47 ¹¹
type	controlled vocabulary
value space	IETF BCP47, Codes for the representation of names of languages (http://tools.ietf.org/html/bcp47)
values	
examples	
DCLINK	
ISOcatLINK	http://www.isocat.org/datcat/DC-2483
comment	for current version, default value should be English

6.4.8 metadataLastDateUpdated

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

¹⁰ To be implemented in next version

¹¹ To be implemented in next version

6.4.9 revision

definition	Provides an account of the revisions in free text or a link to a document with revisions
type	ms:myStringURI
value space	
values	
examples	
DCLINK	
ISOcatLINK	

6.5 versionInfo

definition	Groups information on a specific version or release of the resource
type	component
elements	version <i>Status: mandatory</i> <i>Repeatability: 1</i> revision <i>Status: optional</i> <i>Repeatability: 1</i> lastDateUpdated <i>Status: optional</i> <i>Repeatability: 1</i> updateFrequency <i>Status: optional</i> <i>Repeatability: 1</i>

6.5.1 version

definition	Any string, usually a number, that identifies the version of a resource
type	xs:string
value space	
values	
examples	
DCLINK	
ISOcatLINK	http://www.isocat.org/datcat/DC-2547

6.5.2 lastDateUpdated

definition	Date of the last update of the version of the resource
type	xs:date
value space	
values	
examples	
DCLINK	dc:date
ISOcatLINK	~http://www.isocat.org/datcat/DC-2526

6.5.3 updateFrequency

definition	Specifies the frequency with which the resource is updated
type	xs:string
value space	
values	
examples	yearly updated; updated each month; no update forecast
DCLINK	
ISOcatLINK	http://www.isocat.org/datcat/DC-2540

6.6 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.).
type	component
elements	validated <i>Status: mandatory</i> <i>Repeatability: 1</i> validationType <i>Status: optional</i> <i>Repeatability: 1</i> validationMode <i>Status: optional</i> <i>Repeatability: 1</i>

	validationModeDetails <i>Status: optional</i> <i>Repeatability: 1</i> validationExtent <i>Status: optional</i> <i>Repeatability: 1</i> validationExtentDetails <i>Status: optional</i> <i>Repeatability: 1</i> sizePerValidation <i>Status: optional</i> <i>Repeatability: 1</i> validationReport <i>Status: optional</i> <i>Repeatability: 1</i> validationTool <i>Status: optional</i> <i>Repeatability: 1</i> validator <i>Status: optional</i> <i>Repeatability: unbounded</i>
--	--

6.6.1 validated

definition	Specifies the validation status of the resource
type	xs:boolean
value space	
values	
examples	
DCLINK	
ISOcatLINK	http://www.isocat.org/datcat/DC-2584
comment	please, use "yes" even for partially validated resources

6.6.2 validationType

definition	Specifies the type of the validation that have been performed
------------	---

type	closed controlled vocabulary
value space	MS-validationType
values	Value definition formal content
examples	
DCLINK	
ISOcatLINK	http://www.isocat.org/datcat/DC-2587

6.6.3 validationMode

definition	Specifies the validation methodology applied
type	closed controlled vocabulary
value space	MS-validationMode
values	value definition manual automatic mixed interactive
examples	
DCLINK	
ISOcatLINK	http://www.isocat.org/datcat/DC-2586

6.6.4 validationModeDetails

definition	Textual field for additional information on validation
type	xs:string
value space	
values	
examples	
DCLINK	
ISOcatLINK	

6.6.5 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	http://www.isocat.org/datcat/DC-2586

6.6.6 validationExtentDetails

definition	Provides 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	xs:string
value space	
values	
examples	
DCLINK	
ISocatLINK	
comment	to be used if only part of the resource has been validated and as an alternative to sizePerValidationType if the validated part cannot be counted otherwise

6.6.7 sizePerValidation

definition	Specifies the size of the validated part of a resource
type	linked component
component	sizeInfo

6.6.8 validationReport

definition	A short account of the validation details or a link to the validation report
type	linked component

component	documentationInfo
DCLINK	
ISocatLINK	

6.6.9 validationTool

definition	The name, the identifier or the url of the tool used for the validation of the resource
type	linked component
component	targetResourceInfo
DCLINK	
ISocatLINK	

6.6.10 validator

definition	Groups information on the person(s) or the organization(s) that validated the resource
type	linked component
component	actorInfo
DCLINK	
ISocatLINK	participant role (http://www.isocat.org/datcat/DC-2559) is used for all the functional roles of persons participating in the content of the resource.

6.7 usageInfo

definition	Groups information on usage of the resource (both intended and actual use)
type	component
elements	accessTool <i>Status: optional</i> <i>Repeatability: unbounded</i> resourceAssociatedWith <i>Status: optional</i> <i>Repeatability: unbounded</i> foreseenUseInfo <i>Status: recommended</i> <i>Repeatability: unbounded</i>

	actualUseInfo <i>Status: recommended</i> <i>Repeatability: unbounded</i>
--	--

6.7.1 accessTool

definition	The name or the identifier or the url of the tool used to access a resource (e.g. a corpus workbench)
type	linked component
component	targetResourceInfo
examples	a workbench used for accessing a corpus
DCLINK	
ISOcatLINK	~ http://www.isocat.org/datcat/DC-2514

6.7.2 resourceAssociatedWith

definition	Refers to another resource that the resource described uses for its operation
type	linked component
component	targetResourceInfo
examples	a tagger using a lexicon as one of its components; a corpus used for training a tool
DCLINK	
ISOcatLINK	

6.7.3 foreseenUseInfo

definition	Groups information on the use for which the resource is created
type	component
elements	foreseenUse <i>Status: mandatory</i> <i>Repeatability: 1</i> useNLPSpecific <i>Status: recommended</i> <i>Repeatability: unbounded</i>

6.7.3.1 foreseenUse

definition	Classification of the intended use of the resource						
type	closed controlled vocabulary						
value space	MS-foreseenUse						
values	<table> <tr> <th>value</th><th>definition</th></tr> <tr> <td>humanUse</td><td></td></tr> <tr> <td>nlpApplications</td><td></td></tr> </table>	value	definition	humanUse		nlpApplications	
value	definition						
humanUse							
nlpApplications							
examples							
DCLINK							
ISOcatLINK							
comment	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						

6.7.3.2 useNLPSpecific

definition	Specifies the NLP application for which the resource is created or the application in which it has actually been used.																																		
type	open controlled vocabulary																																		
value space	MS-useNLPSpecific																																		
values	<table> <tr> <th>value</th><th>definition</th></tr> <tr> <td>parsing</td><td></td></tr> <tr> <td>contradictionDetection</td><td></td></tr> <tr> <td>opinionMining</td><td></td></tr> <tr> <td>wordSenseDisambiguation</td><td></td></tr> <tr> <td>voiceControl</td><td></td></tr> <tr> <td>topicDetection_Tracking</td><td></td></tr> <tr> <td>textualEntailment</td><td></td></tr> <tr> <td>textMining</td><td></td></tr> <tr> <td>textCategorisation</td><td></td></tr> <tr> <td>terminologyExtraction</td><td></td></tr> <tr> <td>summarisation</td><td></td></tr> <tr> <td>spellChecking</td><td></td></tr> <tr> <td>speechUnderstanding</td><td></td></tr> <tr> <td>speechToSpeechTranslation</td><td></td></tr> <tr> <td>speechSynthesis</td><td></td></tr> <tr> <td>speechRecognition</td><td></td></tr> </table>	value	definition	parsing		contradictionDetection		opinionMining		wordSenseDisambiguation		voiceControl		topicDetection_Tracking		textualEntailment		textMining		textCategorisation		terminologyExtraction		summarisation		spellChecking		speechUnderstanding		speechToSpeechTranslation		speechSynthesis		speechRecognition	
value	definition																																		
parsing																																			
contradictionDetection																																			
opinionMining																																			
wordSenseDisambiguation																																			
voiceControl																																			
topicDetection_Tracking																																			
textualEntailment																																			
textMining																																			
textCategorisation																																			
terminologyExtraction																																			
summarisation																																			
spellChecking																																			
speechUnderstanding																																			
speechToSpeechTranslation																																			
speechSynthesis																																			
speechRecognition																																			

	signLanguageRecognition signLanguageGeneration semanticWeb questionAnswering informationExtraction posTagging personIdentification naturalLanguageUnderstanding naturalLanguageGeneration namedEntityRecognition multimediaDocumentProcessing morphosyntacticTagging morphologicalAnalysis linguisticResearch lexiconEnhancement lemmatization languageModelsTraining languageModelling languageIdentification knowledgeRepresentation knowledgeDiscovery emotionRecognition emotionGeneration documentClassification derivationalMorphologicalAnalysis coreferenceResolution bilingualLexiconInduction annotation webServices eventExtraction semanticRoleLabelling readingAndWritingAidApplications temporalExpressionRecognition intra-documentCoreferenceResolution visualSceneUnderstanding entityMentionRecognition sentimentAnalysis
--	---

	<p>machineTranslation</p> <p>persuasiveExpressionMining</p> <p>qualitativeAnalysis</p> <p>texToSpeechSynthesis</p> <p>personRecognition</p> <p>textGeneration</p> <p>avatarSynthesis</p> <p>discourseAnalysis</p> <p>expressionRecognition</p> <p>faceRecognition</p> <p>faceVerification</p> <p>humanoidAgentSynthesis</p> <p>informationRetrieval</p> <p>lexiconAccess</p> <p>lexiconAcquisitionFromCorpora</p> <p>lexiconExtractionFromLexica</p> <p>lexiconFormatConversion</p> <p>lexiconMerging</p> <p>lexiconVisualization</p> <p>lipTrackingAnalysis</p> <p>multimediaDevelopment</p> <p>speakerIdentification</p> <p>speakerVerification</p> <p>speechLipsCorrelationAnalysis</p> <p>speechAnalysis</p> <p>speechAssistedVideoControl</p> <p>speechVerification</p> <p>spokenDialogueSystems</p> <p>talkingHeadSynthesis</p> <p>userAuthentication</p> <p>other</p>
examples	
DCLINK	
ISOcatLINK	
comments	

6.7.4 actualUseInfo

definition	Groups information on how the resource has already been used
type	component
elements	<p>actualUse</p> <p><i>Status: mandatory</i></p> <p><i>Repeatability: 1</i></p> <p>useNLPSpecific</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>usageReport</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>derivedResource</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>usageProject</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>actualUseDetails</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p>

6.7.4.1 actualUse

definition	Classification of the actual use of the resource						
type	closed controlled vocabulary						
value space	MS-actualUse						
values	<table> <tr> <th>value</th><th>definition</th></tr> <tr> <td>humanUse</td><td></td></tr> <tr> <td>nlpApplications</td><td></td></tr> </table>	value	definition	humanUse		nlpApplications	
value	definition						
humanUse							
nlpApplications							
examples							
DCLINK							
ISOcatLINK							

6.7.4.2 usageReport

definition	Reports on the research papers documenting the usage of a resource, either in
------------	---

	a structured form or in free text
type	linked component
component	documentationInfo

6.7.4.3 derivedResource

definition	The name, the identifier or the url of the outcome or product of the resource.
type	linked component
component	targetResourceInfo
examples	terminological list as the result of term extraction corpus

6.7.4.4 usageProject

definition	Groups information on the project in which the resource has been used
type	linked component
component	projectInfo

6.7.4.5 actualUseDetails

definition	Reports on the usage of the resource in free text
type	xs:string
value space	
values	
examples	
DCLINK	
ISocatLINK	

6.8 resourceDocumentationInfo

definition	Groups together information on any document describing the resource
type	component
elements	documentation <i>Status: recommended</i> <i>Repeatability: unbounded</i> samplesLocation <i>Status: recommended</i> <i>Repeatability: unbounded</i> toolDocumentationType

	<i>Status: condition-dependent mandatory</i> <i>Condition: resourceType=toolService</i> <i>Repeatability: unbounded</i>
--	---

6.8.1 documentation

definition	Refers to papers, manuals, reports etc. describing the resource
type	linked component
value space	documentationInfo

6.8.2 samplesLocation

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

6.8.3 toolDocumentationType

definition	Specifies the type of documentation for tool or service
type	open controlled vocabulary
value space	MS-toolDocumentationType
values	value definition online manual helpFunctions none other
examples	
DCLINK	
ISOcatLINK	

comment	to be used only for tools
---------	---------------------------

6.9 resourceCreationInfo

definition	Groups information on the creation procedure of a resource
type	component
elements	<p>resourceCreator</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>fundingProject</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>creationStartDate</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p> <p>creationEndDate</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p>

6.9.1 resourceCreator

definition	Groups information on the person or the organization that has created the resource
type	special status component
component	actorInfo
DCLINK	dc:creator
ISocatLINK	http://www.isocat.org/datcat/DC-2513

6.9.2 fundingProject

definition	Groups information on the project that has funded the creation of the resource
type	special status component
component	projectInfo
comment	to be repeated for each project; internal funding of a resource is also thought of as a kind of project

6.9.3 creationStartDate

definition	The date in which the creation process was started
type	xs:date
value space	
values	
examples	
DCLINK	dc:date
ISOcatLINK	~ http://www.isocat.org/datcat/DC-2510

6.9.4 creationEndDate

definition	The date in which the creation process was completed
type	xs:date
value space	
values	
examples	
DCLINK	dc:date
ISOcatLINK	http://www.isocat.org/datcat/DC-2509

6.10 relationInfo

definition	Groups information on the relations of the resource being described with other resources
type	component
elements	relationType <i>Status: mandatory</i> <i>Repeatability: 1</i> relatedResource <i>Status: mandatory</i> <i>Repeatability: 1</i>

6.10.1 relationType

definition	Specifies the type of relation not covered by the ones proposed by META-SHARE
type	xs:string

value space	
values	
examples	
DCLINK	
ISocatLINK	http://www.isocat.org/datcat/DC-2549

6.10.2 relatedResource

definition	The full name, the identifier or the url of the related resource
type	linked component
component	targetResourceInfo

6.11 resourceComponentType

definition	Used for distinguishing between resource types
type	component
components	choice corpusInfo toolServiceInfo languageDescriptionInfo lexicalConceptualResourceInfo

7 Corpora

7.1 resourceInfo

definition	The profile for corpora; extension of the resourceInfo for corpora
elements	resourceInfo – common components identificationInfo distributionInfo contactPerson metadataInfo versionInfo validationInfo usageInfo resourceDocumentationInfo

	resourceCreationInfo relationInfo resourceComponentType corpusInfo
--	---

7.2 corpusInfo

definition	Groups together information on corpora of all media types
type	component
elements	resourceType <i>Status: mandatory with value "corpus"</i> <i>Repeatability: 1</i> corpusMediaType <i>Status: mandatory</i> <i>Repeatability:1</i>

7.2.1 resourceType

definition	Specifies the type of the resource being described	
type	closed controlled vocabulary	
value space	MS-resourceType	
values	value corpus lexicalConceptualResource languageDescription toolService	definition for text, speech, video, multimodal/multimedia corpora, corpora with sensorimotor information, n-gram resources includes lexica, ontologies, dictionaries, word lists etc. covers grammars etc. used for tools, systems, system components etc.
examples		
DCLINK	dc:type	
ISocatLINK		

7.2.2 corpusMediaType

definition	Used to specify the media type specific to corpora and group together the relevant information
type	component
component	<p>choice</p> <p>corpusTextInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>corpusAudioInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p> <p>corpusVideoInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>corpusImageInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p> <p>corpusTextNumericalInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p> <p>corpusTextNgramInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p>

7.3 corpusTextInfo

definition	Groups together information on the text component of a resource
type	component
elements	<p>mediaType</p> <p><i>Status: mandatory with value "text"</i></p> <p><i>Repeatability: 1</i></p> <p>lingualityInfo</p> <p><i>Status: mandatory</i></p> <p><i>Repeatability: 1</i></p> <p>languageInfo</p> <p><i>Status: mandatory</i></p> <p><i>Repeatability: unbounded</i></p>

	<p><u>modalityInfo</u></p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p><u>sizeInfo</u></p> <p><i>Status: mandatory</i></p> <p><i>Repeatability: unbounded</i></p> <p><u>textFormatInfo</u></p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p><u>characterEncodingInfo</u></p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p><u>annotationInfo</u></p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p><u>domainInfo</u></p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p><u>textClassificationInfo</u></p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p><u>timeCoverageInfo</u></p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p><u>geographicCoverageInfo</u></p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p><u>creationInfo</u></p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p> <p><u>linkToOtherMediaInfo</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p>
--	---

7.3.1 mediaType

definition	Specifies the media type of the resource and basically corresponds to the physical medium of the content representation. Each media type is described through a distinctive set of features. A resource may consist of parts attributed to different types of media. A tool/service may take as input/output more than one different media types.	
type	closed controlled vocabulary	
value space	MS-mediaType	
values	value	definition
	text	only written medium
	audio	for a whole resource or part of a resource that is recorded as an audio file
	video	for moving image; used for multi-media corpora, with video for the moving image part
	image	for still image like photographs, drawings etc.
	textNumerical	for the description of sensorimotor data (e.g. biological measurements, geospatial data etc.)
	textNgram	for n-gram resources
examples		
DCLINK	dc:type	
ISocatLINK	~ http://www.isocat.org/datcat/DC-2570 (different values)	

7.3.2 lingualityInfo

definition	Groups information on the number of languages of the resource part and of the way they are combined to each other	
type	component	
element	lingualityType <i>Status: mandatory</i> <i>Repeatability: 1</i> multilingualityType <i>Status: optional</i> <i>Repeatability: 1</i> multilingualityTypeDetails <i>Status: optional</i> <i>Repeatability: 1</i>	

7.3.2.1 lingualityType

definition	Indicates whether the resource includes one, two or more languages								
type	closed controlled vocabulary								
value space	MS-lingualityType								
values	<table> <tr> <th>value</th><th>definition</th></tr> <tr> <td>monolingual</td><td></td></tr> <tr> <td>bilingual</td><td></td></tr> <tr> <td>multilingual</td><td></td></tr> </table>	value	definition	monolingual		bilingual		multilingual	
value	definition								
monolingual									
bilingual									
multilingual									
examples									
DCLINK									
ISOcatLINK	http://www.isocat.org/datcat/DC-2491								

7.3.2.2 multilingualityType

definition	Indicates whether the corpus is parallel, comparable or mixed										
type	open controlled vocabulary										
values	MS-multilingualityType										
value space	<table> <tr> <th>Value</th><th>definition</th></tr> <tr> <td>comparable</td><td></td></tr> <tr> <td>parallel</td><td></td></tr> <tr> <td>multilingualSingleText</td><td></td></tr> <tr> <td>other</td><td></td></tr> </table>	Value	definition	comparable		parallel		multilingualSingleText		other	
Value	definition										
comparable											
parallel											
multilingualSingleText											
other											
examples											
DCLINK											
ISOcatLINK											

7.3.2.3 multilingualityTypeDetails

definition	Provides further information on multilinguality of a resource in free text
type	xs:string
value space	
values	
examples	direction of translation
DCLINK	
ISOcatLINK	

7.3.3 languageInfo

definition	Groups information on the languages represented in the resource
type	component
elements	<p>languageId</p> <p><i>Status: mandatory</i></p> <p><i>Repeatability: 1</i></p> <p>languageName</p> <p><i>Status: mandatory</i></p> <p><i>Repeatability: 1</i></p> <p>languageScript</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p>sizePerLanguage</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p>languageVarietyInfo</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p>

7.3.3.1 languageId

definition	The identifier of the language that is included in the resource or supported by the tool/service according to the IETF BCP47 standard ¹²
type	controlled vocabulary
value space	IETF BCP47, Codes for the representation of names of languages
values	
examples	
DCLINK	dc:language
ISOcatLINK	http://www.isocat.org/datcat/DC-2482 (ISO 6939)

7.3.3.2 languageName

definition	A human understandable name of the language that is used in the resource or supported by the tool/service according to the IETF BCP47 standard ¹³
type	controlled vocabulary

¹² To be implemented in next version

¹³ To be implemented in next version

value space	IETF BCP47, Codes for the representation of names of languages
values	
examples	
DCLINK	dc:language
ISocatLINK	

7.3.3.3 languageScript

definition	Specifies the writing system used to represent the language of the resource. Please use the 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	http://www.isocat.org/datcat/DC-2485

7.3.3.4 sizePerLanguage

definition	Provides information on the size per language component
type	linked component
component	sizeInfo
ISocatLINK	http://www.isocat.org/datcat/DC-2581

7.3.3.5 languageVarietyInfo

definition	Groups information on language varieties occurred in the resource (e.g. dialects)
type	component
elements	languageVarietyType <i>Status: mandatory</i> <i>Repeatability: 1</i> languageVarietyName <i>Status: mandatory</i> <i>Repeatability: 1</i> sizePerLanguageVariety <i>Status: optional</i> <i>Repeatability: 1</i>
comment	to be repeated for different language varieties

7.3.3.5.1 languageVarietyType

definition	Specifies the type of the language variety that occurs in the resource or is supported by a tool/service
type	open controlled vocabulary
value space	MS-languageVarietyType
values	<div>value</div> <div>definition</div> <div>dialect</div> <div>jargon</div> <div>other</div>
examples	
DCLINK	
ISocatLINK	

7.3.3.5.2 languageVarietyName

definition	The name of the language variety that occurs in the resource or is supported by a tool/service
type	xs:string
value space	
values	
examples	
DCLINK	
ISocatLINK	http://www.isocat.org/datcat/DC-2466

7.3.3.5.3 sizePerLanguageVariety

definition	Provides information on the size per language variety component
type	linked component
component	sizeInfo

7.3.4 modalityInfo

definition	Groups information on the modalities represented in the resource
type	component
elements	modalityType

	<i>Status: mandatory</i> <i>Repeatability: 1</i> modalityTypeDetails <i>Status: optional</i> <i>Repeatability: 1</i> sizePerModality <i>Status: optional</i> <i>Repeatability: 1</i>
--	---

7.3.4.1 modalityType

definition	Specifies the type of the modality represented in the resource or processed by a tool/service																		
type	open controlled vocabulary																		
value space	MS-modalityType																		
values	<table> <tr> <td>value</td><td>definition</td></tr> <tr> <td>bodyGesture</td><td></td></tr> <tr> <td>facialExpression</td><td></td></tr> <tr> <td>voice</td><td></td></tr> <tr> <td>combinationOfModalities</td><td></td></tr> <tr> <td>signLanguage</td><td></td></tr> <tr> <td>spokenLanguage</td><td></td></tr> <tr> <td>writtenLanguage</td><td></td></tr> <tr> <td>other</td><td></td></tr> </table>	value	definition	bodyGesture		facialExpression		voice		combinationOfModalities		signLanguage		spokenLanguage		writtenLanguage		other	
value	definition																		
bodyGesture																			
facialExpression																			
voice																			
combinationOfModalities																			
signLanguage																			
spokenLanguage																			
writtenLanguage																			
other																			
examples																			
DCLINK																			
ISOcatLINK	~ http://www.isocat.org/datcat/DC-2490																		

7.3.4.2 modalityTypeDetails

definition	Provides further information on modalities
type	xs:string
value space	
values	
examples	
DCLINK	
ISOcatLINK	

7.3.4.3 sizePerModality

definition	Provides information on the size per modality component
type	linked component
component	sizeInfo

7.3.5 textFormatInfo

definition	Groups information on the text format(s) of a resource
type	component
elements	mimeType <i>Status: mandatory</i> <i>Repeatability: 1</i> sizePerTextFormat <i>Status: optional</i> <i>Repeatability: 1</i>
comment	to be repeated if parts of the resource are in different formats

7.3.5.1 mimeType

definition	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 of the Internet Assigned Numbers Authority (http://www.iana.org/)
type	xs:string
value space	
values	
examples	
DCLINK	dc:format
ISOcatLINK	http://www.isocat.org/datcat/DC-2571

7.3.5.2 sizePerTextFormat

definition	Provides information on the size of the resource parts with different format
type	linked component
component	sizeInfo

7.3.6 characterEncodingInfo

definition	Groups together information on character encoding of the resource
type	component
elements	characterEncoding <i>Status: mandatory</i> <i>Repeatability: 1</i> sizePerCharacterEncoding <i>Status: optional</i> <i>Repeatability: 1</i>
comment	to be repeated if parts of the resource have different character encodings

7.3.6.1 characterEncoding

definition	The name of the character encoding used in the resource or accepted by the tool/service
type	closed controlled vocabulary
value space	MS-characterEncoding
values	
examples	US-ASCII; ISO8859-1; UTF-8
DCLINK	
ISocatLINK	http://www.isocat.org/datcat/DC-2564
comment	values are included in the xsd; too long catalogue to be presented here

7.3.6.2 sizePerCharacterEncoding

annotation	Provides information on the size of the resource parts with different character encoding
type	linked component
component	sizeInfo

7.3.7 annotationInfo

definition	Groups information on the annotated part(s) of a resource
type	component
elements	annotationType <i>Status: mandatory</i> <i>Repeatability: 1</i>

	<p><u>annotatedElements</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p><u>annotationStandoff</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p><u>segmentationLevel</u></p> <p><i>Status: optional</i></p> <p><i>Condition: annotationType=segmentation</i></p> <p><i>Repeatability: unbounded</i></p> <p><u>annotationFormat</u></p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p> <p><u>tagset</u></p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p> <p><u>tagsetLanguageId</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p><u>tagsetLanguageName</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p><u>conformanceToStandardsBestPractices</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p><u>theoreticModel</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p><u>annotationManual</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p><u>annotationMode</u></p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p> <p><u>annotationModeDetails</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p>
--	---

	annotationTool <i>Status: recommended</i> <i>Repeatability: unbounded</i> annotationStartDate <i>Status: optional</i> <i>Repeatability: 1</i> annotationEndDate <i>Status: optional</i> <i>Repeatability: 1</i> sizePerAnnotation <i>Status: optional</i> <i>Repeatability: 1</i> interannotatorAgreement <i>Status: optional</i> <i>Repeatability: 1</i> intraannotatorAgreement <i>Status: optional</i> <i>Repeatability: 1</i> annotator <i>Status: optional</i> <i>Repeatability: unbounded</i>
comment	to be repeated for different annotation types

7.3.7.1 annotationType

definition	Specifies the annotation level of the resource or the annotation type a tool/service requires or produces as an output
type	open controlled vocabulary
value space	MS-annotationType
values	<div>value</div> <div>alignment</div> <div>discourseAnnotation</div> <div>discourseAnnotation-audienceReactions</div> <div>discourseAnnotation-coreference</div> <div>discourseAnnotation-dialogueActs</div> <div>discourseAnnotation-discourseRelations</div> <div>lemmatization</div> <div>definition</div>

<p> modalityAnnotation-bodyMovements modalityAnnotation-facialExpressions modalityAnnotation-gazeEyeMovements modalityAnnotation-handArmGestures modalityAnnotation-handManipulationOfObjects modalityAnnotation-headMovements modalityAnnotation-lipMovements morphosyntacticAnnotation morphosyntacticAnnotation-bPosTagging morphosyntacticAnnotation-PosTagging 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 </p>
--

	syntacticAnnotation-shallowParsing syntacticAnnotation-subcategorizationFrames syntacticAnnotation-treebanks syntacticosemanticAnnotation-links translation transliteration other
examples	
DCLINK	
ISOcatLINK	http://www.isocat.org/datcat/DC-2462

7.3.7.2 annotatedElements

definition	Specifies the elements annotated in each annotation level
type	open controlled vocabulary
value space	MS-annotatedElements
values	value definition speakerNoise backgroundNoise mispronunciations truncation discourseMarkers other
examples	
DCLINK	
ISOcatLINK	

7.3.7.3 annotationStandoff

definition	Indicates whether the annotation is created inline or in a stand-off fashion
type	xs:boolean
value space	
values	
examples	
DCLINK	
ISOcatLINK	http://www.isocat.org/datcat/DC-2507

7.3.7.4 segmentationLevel

definition	Specifies the segmentation unit in terms of which the resource has been segmented or the level of segmentation a tool/service requires/outputs
type	closed controlled vocabulary
value space	MS-segmentationLevel
values	<div>value</div> <div>definition</div> <p> paragraph sentence clause word wordGroup utterance topic signal phoneme syllable phrase diphone prosodicBoundaries frame scene shot other </p>
examples	
DCLINK	
ISocatLINK	
comment	to be used when annotationType=segmentation

7.3.7.5 annotationFormat

definition	Specifies the format that is used in the annotation process since often the mime type will not be sufficient for machine processing
type	xs:string
value space	
values	
examples	
DCLINK	dc:format

ISocatLINK	http://www.isocat.org/datcat/DC-2562
------------	---

7.3.7.6 tagset

definition	A name or a url reference to the tagset used in the annotation of the resource or used by the tool/service
type	ms:myStringURI
value space	
values	
examples	
DCLINK	
ISocatLINK	http://www.isocat.org/datcat/DC-2497

7.3.7.7 tagsetLanguageId

definition	The identifier of the tagset language as expressed in the values of IETF BP47 ¹⁴
type	controlled vocabulary
value space	IETF BP47, Codes for the representation of names of languages
values	
examples	
DCLINK	
ISocatLINK	http://www.isocat.org/datcat/DC-2498

7.3.7.8 tagsetLanguageName

definition	The name of the tagset language expressed in the values of IETF BP47 ¹⁵
type	controlled vocabulary
value space	IETF BP47, Codes for the representation of names of languages
values	
examples	
DCLINK	
ISocatLINK	http://www.isocat.org/datcat/DC-2499

7.3.7.9 conformanceToStandardsBestPractices

definition	Specifies the standards or the best practices to which the tagset used for the
------------	--

¹⁴ To be implemented in next version

¹⁵ To be implemented in next version

	annotation conforms	
type	open controlled vocabulary	
value space	MS-standards	
values	value	definition
	CES	http://www.cs.vassar.edu/CES/
	EML	http://www.w3.org/TR/2009/WD-emotionml-20091029/
	EMMA	http://www.w3.org/TR/emma/
	GMX	
	HamNoSys	http://www.sign-lang.uni-hamburg.de/projects/hamnosys.html
	InkML	http://www.w3.org/TR/InkML/
	ISO12620	http://www.ttt.org/clsframe/datcats.html
	ISO16642	
	ISO1987	
	ISO26162	
	ISO30042	
	ISO704	
	LMF	http://www.lexicalmarkupframework.org/
	MAF	
	MLIF	http://mlif.loria.fr/
	MULTEXT	
	multimodalInteractionFramework	http://www.w3.org/TR/mmi-framework/
	OAXAL	http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=oaxal
	OWL	http://www.w3.org/TR/owl-features/
	pennTreeBank	http://www.cis.upenn.edu/~treebank/
	pragueTreeBanks	http://ufal.mff.cuni.cz/pdt2.o/
	RDF	http://www.w3.org/RDF/
	SemAF	
	SemAF_DA	
	SemAF_NE	
	SemAF_SRL	
	SemAF_DS	

	SKOS SRX SynAF TBX TMX TEI http://www.tei-c.org/index.xml TEI_P3 http://www.tei-c.org/Vault/GL/P3/index.htm TEI_P4 http://www.tei-c.org/Guidelines/P4/ TEI_P5 http://www.tei-c.org/Guidelines/P5/ TimeML http://timeml.org/site/index.html XCES http://www.xces.org/ XLIFF http://docs.oasis-open.org/xliff/v1.2/os/xliff-core.html MUMIN http://www.cst.dk/mumin/ BML http://www.mindmakers.org/projects/BML other
examples	
DCLINK	
ISOcatLI NK	

7.3.7.10 theoreticModel

definition	Name of the theoretic model applied for the creation or enrichment of the resource, and/or reference (URL or bibliographic reference) to informative material about the theoretic model used
type	ms:myStringURI
value space	
values	
examples	LFG; HPSG; wordnet; framenet
DCLINK	
ISOcatLINK	~ http://www.isocat.org/datcat/DC-2501 (defintion restricts to annotation task); ~ http://www.isocat.org/datcat/DC-2636

7.3.7.11 annotationManual

definition	A bibliographic reference or ms:httpURI link to the annotation manual
------------	---

type	linked component
value space	documentationInfo
values	
examples	
DCLINK	
ISocatLINK	

7.3.7.12 annotationMode

definition	Indicates whether the resource is annotated manually or by automatic processes										
type	closed controlled vocabulary										
value space	MS-annotationMode										
values	<table> <tr> <th>Value</th><th>definition</th></tr> <tr> <td>automatic</td><td></td></tr> <tr> <td>manual</td><td></td></tr> <tr> <td>mixed</td><td></td></tr> <tr> <td>interactive</td><td></td></tr> </table>	Value	definition	automatic		manual		mixed		interactive	
Value	definition										
automatic											
manual											
mixed											
interactive											
examples											
DCLINK											
ISocatLINK	http://www.isocat.org/datcat/DC-2506										

7.3.7.13 annotationModeDetails

definition	Provides further information on annotation process
type	xs:string
value space	
values	
examples	
DCLINK	
ISocatLINK	

7.3.7.14 annotationTool

definition	The name, the identifier or the url of the tool used for the annotation of the resource
type	linked component

component	targetResourceInfo
ISocatLINK	http://www.isocat.org/datcat/DC-3785

7.3.7.15 annotationStartDate

definition	The date in which the annotation process has started
type	date
value space	
values	
examples	
DCLINK	dc:date
ISocatLINK	

7.3.7.16 annotationEndDate

definition	The date in which the annotation process has ended
type	date
value space	
values	
examples	
DCLINK	dc:date
ISocatLINK	

7.3.7.17 sizePerAnnotation

definition	Provides information on size for the annotated parts of the resource
type	linked component
component	sizeInfo

7.3.7.18 interannotatorAgreement

definition	Provides information on the interannotator agreement and the methods/metrics applied
type	xs:string
value space	
values	
examples	
DCLINK	

ISocatLINK	http://www.isocat.org/datcat/DC-2525
------------	---

7.3.7.19 intraannotatorAgreement

definition	Provides information on the intra-annotator agreement and the methods/metrics applied
type	xs:string
value space	
values	
examples	
DCLINK	
ISocatLINK	

7.3.7.20 annotator

definition	Groups information on the annotators of the specific annotation type
type	linked component
component	actorInfo

7.3.8 domainInfo

definition	Groups together information on domains represented in the resource; can be repeated for parts of the resource with distinct domain
type	component
elements	domain <i>Status: mandatory</i> <i>Repeatability: 1</i> sizePerDomain <i>Status: optional</i> <i>Repeatability: 1</i> conformanceToClassificationScheme <i>Status: optional</i> <i>Repeatability: 1</i>
comment	To be repeated for different domains in terms of size and classification scheme

7.3.8.1 domain

definition	Specifies the application domain of the resource or the tool/service
type	xs:string
value space	
values	
examples	
DCLINK	dc:subject
ISOcatLINK	http://www.isocat.org/datcat/DC-2467

7.3.8.2 sizePerDomain

definition	Specifies the size of resource parts per domain
type	linked component
component	sizeInfo

7.3.8.3 conformanceToClassificationScheme

definition	Specifies the external classification schemes	
type	open controlled vocabulary	
value space	MS-conformanceToClassificationScheme	
values	value	definition
	ANC_domainClassification	
	ANC_genreClassification	
	BNC_domainClassification	
	BNC_textTypeClassification	
	DDC_classification	
	libraryOfCongress_domainClassification	
	libraryOfCongressSubjectHeading_classification	
	MESH_classification	
	NLK_classification	
	PAROLE_topicClassification	
	PAROLE_genreClassification	
	UDC_classification	
	other	
examples		
DCLINK		
ISOcatLINK		

7.3.9 textClassificationInfo

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

7.3.9.1 textGenre

definition	Genre: The conventionalized discourse or text types of the content of the resource, based on extra-linguistic and internal linguistic criteria
type	xs:string
value space	
values	
examples	
DCLINK	dc:subject
ISocatLINK	http://www.isocat.org/datcat/DC-2470

7.3.9.2 textType

definition	Specifies the type of the text according to a text type classification
type	xs:string

value space	
values	
examples	
DCLINK	dc:subject
ISocatLINK	
comment	For text corpora that have already been using text type classification

7.3.9.3 register

definition	For corpora that have already been using register classification
type	xs:string
value space	
values	
examples	
DCLINK	
ISocatLINK	

7.3.9.4 subject_topic

definition	For corpora that have already been using subject classification
type	xs:string
value space	
values	
examples	
DCLINK	dc:subject
ISocatLINK	

7.3.9.5 sizePerTextClassification

definition	Provides information on size of resource parts with different text classification
type	linked component
component	sizeInfo

7.3.10 timeCoverageInfo

definition	Groups together information on time classification of the resource
type	component

elements	timeCoverage <i>Status: mandatory</i> <i>Repeatability: 1</i> sizePerTimeCoverage <i>Status: optional</i> <i>Repeatability: 1</i>
----------	--

7.3.10.1 timeCoverage

definition	The time period that the content of a resource is about
type	xs:string
value space	
values	
examples	
DCLINK	dc:coverage
ISOcatLINK	http://www.isocat.org/datcat/DC-2502

7.3.10.2 sizePerTimeCoverage

definition	Provides information on size per time period represented in the resource
type	linked component
component	sizeInfo

7.3.11 geographicCoverageInfo

definition	Groups information on geographic classification of the resource
type	component
elements	geographicCoverage <i>Status: mandatory</i> <i>Repeatability: 1</i> sizePerGeographicCoverage <i>Status: optional</i> <i>Repeatability: 1</i>

7.3.11.1 geographicCoverage

definition	The geographic region that the content of a resource is about; for countries,
------------	---

	recommended use of ISO-3166
type	xs:string
value space	
values	
examples	
DCLINK	dc:coverage
ISocatLINK	http://www.isocat.org/datcat/DC-2471

7.3.11.2 sizePerGeographicCoverage

definition	Provides information on size per geographically distinct section of the resource
type	linked component
component	sizeInfo

7.3.12 creationInfo

definition	Groups together information on the resource creation (e.g. for corpora, selection of texts/audio files/ video files etc. and structural encoding thereof; for lexica, construction of lemma list etc.)
type	component
elements	<p>originalSource</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>creationMode</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p> <p>creationModeDetails</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p>creationTool</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p>
comment	for annotation, use the annotationInfo component

7.3.12.1 originalSource

definition	The name, the identifier or the url of the original resources that were at the
------------	--

	base of the creation process of the resource
type	linked component
value space	targetResourceInfo
values	
examples	
DCLINK	dc:source
ISOcatLINK	http://www.isocat.org/datcat/DC-2534
comment	alternative to the relation hasOriginalSource

7.3.12.2 creationMode

definition	Specifies whether the resource is created automatically or in manual or interactive mode										
type	closed controlled vocabulary										
value space	MS-creationMode										
values	<table> <tr> <th>value</th><th>definition</th></tr> <tr> <td>automatic</td><td></td></tr> <tr> <td>manual</td><td></td></tr> <tr> <td>mixed</td><td></td></tr> <tr> <td>interactive</td><td></td></tr> </table>	value	definition	automatic		manual		mixed		interactive	
value	definition										
automatic											
manual											
mixed											
interactive											
examples											
DCLINK											
ISOcatLINK	~ http://www.isocat.org/datcat/DC-2516										

7.3.12.3 creationModeDetails

definition	Provides further information on the creation methods and processes
type	xs:string
value space	
values	
examples	
DCLINK	
ISOcatLINK	

7.3.12.4 creationTool

definition	The name, the identifier or the url of the tool used in the creation process
------------	--

type	linked component
value space	targetResourceInfo
values	
examples	
DCLINK	
ISocatLINK	http://www.isocat.org/datcat/DC-2511

7.3.13 linkToOtherMediaInfo

definition	Groups information on the way different media of the resource interact with or link to each other. To be used for multimodal resources or for resources representing sensorimotor data
type	component
elements	<p>otherMedia</p> <p><i>Status: mandatory</i></p> <p><i>Repeatability: 1</i></p> <p>mediaTypeDetails</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p>synchronizedWithText</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p>synchronizedWithAudio</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p>synchronizedWithVideo</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p>synchronizedWithImage</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p>synchronizedWithTextNumerical</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p>

7.3.13.1 otherMedia

definition	Specifies the media types that are linked to the media type described within the same resource
type	closed controlled vocabulary
value space	MS-otherMedia
values	<div>value</div> <div>definition</div> <div>text</div> <div>textNumerical</div> <div>video</div> <div>audio</div> <div>image</div>
examples	
DCLINK	
ISOcatLINK	

7.3.13.2 mediaTypeDetails

definition	Provides further information on the way the media types are linked and/or synchronized with each other within the same resource
type	xs:string
value space	
values	
examples	
DCLINK	
ISOcatLINK	

7.3.13.3 synchronizedWithText

definition	Whether video, text and textNumerical media type is synchronized with text within the same resource
type	xs:boolean
value space	
values	
examples	
DCLINK	
ISOcatLINK	
comment	it is possible to have more than one corpusTextInfo within the same resource

	e.g. the script and the subtitles of a movie
--	--

7.3.13.4 synchronizedWithAudio

definition	Whether text, video or textNumerical media type is synchronized with audio within the same resource
type	xs:boolean
value space	
values	
examples	
DCLINK	
ISocatLINK	

7.3.13.5 synchronizedWithVideo

definition	Whether text or textNumerical media type is synchronized with video within the same resource
type	xs:boolean
value space	
values	
examples	
DCLINK	
ISocatLINK	

7.3.13.6 synchronizedWithImage

definition	Whether text or textNumerical media type is synchronized with image within the same resource
type	xs:boolean
value space	
values	
examples	
DCLINK	
ISocatLINK	

7.3.13.7 synchronizedWithTextNumerical

definition	Whether video or audio media type is synchronized with the textNumerical
------------	--

	(representation of sensorimotor measurements) within the same resource
type	xs:boolean
value space	
values	
examples	
DCLINK	
ISocatLINK	

7.4 corpusAudioInfo

definition	Groups together information on the audio module of a corpus
type	component
elements	<p>mediaType</p> <p><i>Status: mandatory with value "audio"</i></p> <p><i>Repeatability: 1</i></p> <p>lingualityInfo</p> <p><i>Status: mandatory</i></p> <p><i>Repeatability: 1</i></p> <p>languageInfo</p> <p><i>Status: mandatory</i></p> <p><i>Repeatability: unbounded</i></p> <p>modalityInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p> <p>audioSizeInfo</p> <p><i>Status: mandatory</i></p> <p><i>Repeatability: unbounded</i></p> <p>audioContentInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p> <p>settingInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p> <p>audioFormatInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>annotationInfo</p>

	<p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>domainInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>timeCoverageInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>geographicCoverageInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>audioClassificationInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>recordingInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p> <p>captureInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p> <p>creationInfo</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p>linkToOtherMediaInfo</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p>
--	---

7.4.1 audioSizeInfo

definition	SizeInfo Element for Audio parts of a resource
type	component
elements	<p>sizeInfo</p> <p><i>Status: mandatory</i></p> <p><i>Repeatability: unbounded</i></p> <p>durationOfEffectiveSpeechInfo</p> <p><i>Status: optional</i></p>

	<i>Repeatability: unbounded</i> durationOfAudioInfo <i>Status: optional</i> <i>Repeatability: unbounded</i>
--	--

7.4.1.1 durationOfEffectiveSpeechInfo

definition	Groups together information on the duration of effective speech
type	component
elements	size <i>Status: mandatory</i> <i>Repeatability: 1</i> durationUnit <i>Status: mandatory</i> <i>Repeatability: 1</i>
ISOcatLINK	~ http://www.isocat.org/datcat/DC-2691 (element)

7.4.1.1.1 durationUnit

definition	Specification of the unit of size that is used when providing information on the size of a resource								
type	closed controlled vocabulary								
value space	MS-durationUnit								
values	<table> <tr> <th>value</th><th>definition</th></tr> <tr> <td>hours</td><td></td></tr> <tr> <td>minutes</td><td></td></tr> <tr> <td>seconds</td><td></td></tr> </table>	value	definition	hours		minutes		seconds	
value	definition								
hours									
minutes									
seconds									
examples									
DCLINK									
ISOcatLINK	http://www.isocat.org/datcat/DC-2567								

7.4.1.2 durationOfAudioInfo

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

	durationUnit <i>Status: mandatory</i> <i>Repeatability: 1</i>
--	---

7.4.2 audioContentInfo

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

7.4.2.1 speechItems

definition	Specifies the distinct elements that are pronounced and annotated as such																						
type	open controlled vocabulary																						
value space	MS-speechItems																						
values	<table> <tr> <th>value</th><th>definition</th></tr> <tr><td>applicationWords</td><td></td></tr> <tr><td>creditCardNumbers</td><td></td></tr> <tr><td>freeSpeech</td><td></td></tr> <tr><td>isolatedWords</td><td></td></tr> <tr><td>isolatedDigits</td><td></td></tr> <tr><td>moneyAmounts</td><td></td></tr> <tr><td>naturalNumbers</td><td></td></tr> <tr><td>phoneticallyBalancedSentences</td><td></td></tr> <tr><td>phoneticallyRichSentences</td><td></td></tr> <tr><td>phoneticallyRichWords</td><td></td></tr> </table>	value	definition	applicationWords		creditCardNumbers		freeSpeech		isolatedWords		isolatedDigits		moneyAmounts		naturalNumbers		phoneticallyBalancedSentences		phoneticallyRichSentences		phoneticallyRichWords	
value	definition																						
applicationWords																							
creditCardNumbers																							
freeSpeech																							
isolatedWords																							
isolatedDigits																							
moneyAmounts																							
naturalNumbers																							
phoneticallyBalancedSentences																							
phoneticallyRichSentences																							
phoneticallyRichWords																							

	properNouns telephoneNumbers yesNoQuestions vcvSequences other
examples	
DCLINK	
ISocatLINK	

7.4.2.2 nonSpeechItems

definition	Specifies the distinct elements that maybe included in the audio corpus
type	open controlled vocabulary
value space	MS-nonSpeechItems
values	Value definition commercial music noise notes sounds tempo other
examples	
DCLINK	
ISocatLINK	

7.4.2.3 textualDescription

definition	The legend of the soundtrack
type	xs:string
value space	
values	
examples	
DCLINK	
ISocatLINK	

7.4.2.4 noiseLevel

definition	Specifies the level of background noise								
type	closed controlled vocabulary								
value space	MS-noiseLevel								
values	<table> <tr> <th>value</th><th>definition</th></tr> <tr> <td>low</td><td></td></tr> <tr> <td>medium</td><td></td></tr> <tr> <td>high</td><td></td></tr> </table>	value	definition	low		medium		high	
value	definition								
low									
medium									
high									
examples									
DCLINK									
ISocatLINK									

7.4.3 settingInfo

definition	Groups together information on the setting of the audio and/or video part of a resource
type	component
elements	<p>naturality</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p>conversationalType</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p>scenarioType</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p>audience</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p>interactivity</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p>interaction</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p>

7.4.3.1 naturality

definition	Specifies the level of naturality for multimodal/multimedia resources		
type	open controlled vocabulary		
value space	MS-naturality		
value	value	definition	ISOcatLink
	assisted		
	elicited		
	natural		
	planned		http://www.isocat.org/datcat/DC-2665
	prompted		
	readSpeech		
	semiPlanned		http://www.isocat.org/datcat/DC-2664
	spontaneous		http://www.isocat.org/datcat/DC-2663
	other		
examples			
DCLINK			
ISOcatLINK	http://www.isocat.org/datcat/DC-2492		

7.4.3.2 conversationalType

definition	Specifies the conversational type of the resource		
type	closed controlled vocabulary		
value space	MS-speechSetting		
values	value	definition	ISOcatLink
	monologue		http://www.isocat.org/datcat/DC-2659
	dialogue		http://www.isocat.org/datcat/DC-2660
	multilogue		
examples			
DCLINK			
ISOcatLINK			

7.4.3.3 scenarioType

definition	Indicates the task defined for the conversation or the interaction of
------------	---

	participants
type	open controlled vocabulary
value space	MS-scenarioType
values	<div>value</div> <div>definition</div> <div>frogStory</div> <div>pearStory</div> <div>mapTask</div> <div>onlineEducationalGame</div> <div>rolePlay</div> <div>wordGame</div> <div>wizardOfOz</div> <div>other</div>
examples	
DCLINK	
ISOcatLINK	

7.4.3.4 audience

definition	Indication of the intended audience size
type	closed controlled vocabulary
value space	MS-audience
values	<div>value</div> <div>definition</div> <div>ISOcatLink</div> <div>no</div> <div>http://www.isocat.org/datecat/DC-2618</div> <div>few</div> <div>some</div> <div>largePublic</div>
examples	
DCLINK	
ISOcatLINK	

7.4.3.5 interactivity

definition	Indicates the level of conversational interaction between speakers (for audio component) or participants (for video component)
type	open controlled vocabulary
value space	MS-interactivity
values	<div>value</div> <div>definition</div> <div>ISOcatLink</div>

	<p>interactive http://www.isocat.org/datcat/DC-2613</p> <p>nonInteractive http://www.isocat.org/datcat/DC-2614</p> <p>semiInteractive http://www.isocat.org/datcat/DC-2615</p> <p>overlapping</p> <p>other</p>
examples	
DCLINK	
ISocatLINK	http://www.isocat.org/datcat/DC-2476

7.4.3.6 interaction

definition	Specifies the parts that interact in an audio or video component
type	xs:string
value space	
values	
examples	human-machine, human-human
DCLINK	
ISocatLINK	

7.4.4 audioFormatInfo

definition	Groups together information on the format of the audio part of a resource
type	component
elements	<p>mimeType</p> <p><i>Status: mandatory</i></p> <p><i>Repeatability: 1</i></p> <p>signalEncoding</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>samplingRate</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p> <p>quantization</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p>byteOrder</p> <p><i>Status: optional</i></p>

	<p><i>Repeatability: 1</i></p> <p><u>signConvention</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p><u>audioQualityMeasuresIncluded</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p><u>numberOfTracks</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p><u>recordingQuality</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p><u>sizePerAudioFormat</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p><u>compressionInfo</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p>
--	---

7.4.4.1 signalEncoding

definition	Specifies the encoding the audio type uses												
type	open controlled vocabulary												
value space	MS-signalEncoding												
values	<table> <tr> <th>value</th><th>definition</th></tr> <tr> <td>ADPCM</td><td></td></tr> <tr> <td>aLaw</td><td></td></tr> <tr> <td>linearPCM</td><td></td></tr> <tr> <td>μ-law</td><td></td></tr> <tr> <td>other</td><td></td></tr> </table>	value	definition	ADPCM		aLaw		linearPCM		μ-law		other	
value	definition												
ADPCM													
aLaw													
linearPCM													
μ-law													
other													
examples													
DCLINK													
ISOcatLINK													

7.4.4.2 samplingRate

definition	Specifies the format of files contained in the resource in Hertz												
type	closed controlled vocabulary												
value space	MS-samplingRate												
values	<table> <tr> <th>value</th><th>definition</th></tr> <tr> <td>8000</td><td></td></tr> <tr> <td>16000</td><td></td></tr> <tr> <td>44100</td><td></td></tr> <tr> <td>48000</td><td></td></tr> <tr> <td>96000</td><td></td></tr> </table>	value	definition	8000		16000		44100		48000		96000	
value	definition												
8000													
16000													
44100													
48000													
96000													
examples													
DCLINK													
ISOcatLINK	~ http://www.isocat.org/datcat/DC-2684												

7.4.4.3 quantization

definition	The number of bits for each audio sample										
type	closed controlled vocabulary										
value space	MS-quantization										
values	<table> <tr> <th>Value</th><th>definition</th></tr> <tr> <td>8</td><td></td></tr> <tr> <td>16</td><td></td></tr> <tr> <td>32</td><td></td></tr> <tr> <td>64</td><td></td></tr> </table>	Value	definition	8		16		32		64	
Value	definition										
8											
16											
32											
64											
examples											
DCLINK											
ISOcatLINK											

7.4.4.4 byteOrder

definition	The byte order of 2 or more bytes sample		
type	closed controlled vocabulary		
value space	MS-byteOrder		
vare	value	definition	ISOcatLink
	littleEndian		http://www.isocat.org/datcat/DC-2688
	bigEndian		http://www.isocat.org/datcat/DC-2687
examples			

DCLINK	
ISocatLINK	http://www.isocat.org/datcat/DC-2686

7.4.4.5 signConvention

definition	Binary representation of numbers								
type	closed controlled vocabulary								
value space	MS-signConvention								
value	<table> <tr> <th>value</th><th>definition</th></tr> <tr> <td>signedInteger</td><td></td></tr> <tr> <td>unsignedInteger</td><td></td></tr> <tr> <td>floatingPoint</td><td></td></tr> </table>	value	definition	signedInteger		unsignedInteger		floatingPoint	
value	definition								
signedInteger									
unsignedInteger									
floatingPoint									
examples									
DCLINK									
ISocatLINK									

7.4.4.6 audioQualityMeasuresIncluded

definition	Specifies the audio quality measures												
type	open controlled vocabulary												
value space	MS-audioQualityMeasuresIncluded												
values	<table> <tr> <th>value</th><th>definition</th></tr> <tr> <td>SNR</td><td></td></tr> <tr> <td>crossTalk</td><td></td></tr> <tr> <td>clippingRate</td><td></td></tr> <tr> <td>backgroundNoise</td><td></td></tr> <tr> <td>other</td><td></td></tr> </table>	value	definition	SNR		crossTalk		clippingRate		backgroundNoise		other	
value	definition												
SNR													
crossTalk													
clippingRate													
backgroundNoise													
other													
examples													
DCLINK													
ISocatLINK													

7.4.4.7 numberOfTracks

definition	Specifies the number of audio channels				
type	open controlled vocabulary				
value space	MS-numberOfTracks				
values	<table> <tr> <th>value</th><th>definition</th></tr> <tr> <td>1</td><td>mono</td></tr> </table>	value	definition	1	mono
value	definition				
1	mono				

	2 stereo 4 8
examples	
DCLINK	
ISocatLINK	

7.4.4.8 recordingQuality

definition	Indication of the audio or video recording quality												
type	open controlled vocabulary												
value space	MS-recordingQuality												
values	<table> <tr> <th>value</th><th>definition</th></tr> <tr> <td>veryLow</td><td></td></tr> <tr> <td>low</td><td></td></tr> <tr> <td>medium</td><td></td></tr> <tr> <td>high</td><td></td></tr> <tr> <td>veryHigh</td><td></td></tr> </table>	value	definition	veryLow		low		medium		high		veryHigh	
value	definition												
veryLow													
low													
medium													
high													
veryHigh													
examples													
DCLINK													
ISocatLINK													

7.4.4.9 sizePerAudioFormat

definition	Used to give info on size of parts of a resource that differ as to the format
type	linked component
component	audioSizeInfo

7.4.4.10 compressionInfo

definition	Groups together information on the compression status and method of a resource
type	component
elements	compression <i>Status: optional</i> <i>Repeatability: 1</i> compressionName <i>Status: optional</i>

	<i>Repeatability: unbounded</i> compressionLoss <i>Status: optional</i> <i>Repeatability: 1</i>
--	--

7.4.4.10.1 compression

definition	Whether the audio, video or image is compressed or not
type	xs:boolean
value space	
values	
examples	
DCLINK	
ISocatLINK	

7.4.4.10.2 compressionName

definition	The name of the compression applied																				
type	open controlled vocabulary																				
value space	MS-compressionName																				
values	<table> <tr> <th>value</th><th>definition</th></tr> <tr><td>flac</td><td></td></tr> <tr><td>shorten</td><td></td></tr> <tr><td>mp3</td><td></td></tr> <tr><td>oggVorbis</td><td></td></tr> <tr><td>atrac</td><td></td></tr> <tr><td>aac</td><td></td></tr> <tr><td>mpeg</td><td></td></tr> <tr><td>realAudio</td><td></td></tr> <tr><td>other</td><td></td></tr> </table>	value	definition	flac		shorten		mp3		oggVorbis		atrac		aac		mpeg		realAudio		other	
value	definition																				
flac																					
shorten																					
mp3																					
oggVorbis																					
atrac																					
aac																					
mpeg																					
realAudio																					
other																					
examples																					
DCLINK																					
ISocatLINK	~ http://www.isocat.org/datcat/DC-2685																				

7.4.4.10.3 compressionLoss

definition	Whether there is loss due to compression
type	xs:boolean

value space	
values	
examples	
DCLINK	
ISocatLINK	

7.4.5 audioClassificationInfo

definition	Groups together information on audio type/genre of the resource
type	component
elements	<p>audioGenre <i>Status: mandatory</i> <i>Repeatability: 1</i></p> <p>speechGenre <i>Status: recommended</i> <i>Repeatability: 1</i></p> <p>subject topic <i>Status: optional</i> <i>Repeatability: 1</i></p> <p>register <i>Status: optional</i> <i>Repeatability: 1</i></p> <p>conformanceToClassificationScheme <i>Status: optional</i> <i>Repeatability: 1</i></p> <p>sizePerAudioClassification <i>Status: optional</i> <i>Repeatability: unbounded</i></p>

7.4.5.1 audioGenre

definition	A first indication of type of sounds recorded		
type	open controlled vocabulary		
value space	MS-audioGenre		
values	value	definition	ISocatLink
	speech		
	humanNonSpeech		

	song instrumental noise animalVocalizations other	http://www.isocat.org/datecat/DC-2611
examples		
DCLINK		
ISocatLINK		

7.4.5.2 speechGenre

definition	The conventionalized discourse 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 broadcastNews conversation meeting roundtable interview debate call-in lecture emotionalExpressive airTrafficControl questionAnswer presentation narrative other	definition
examples		
DCLINK		
ISocatLINK		

7.4.5.3 sizePerAudioClassification

definition	The size of the audio subparts of the resource in terms of classification criteria
type	linked component
component	sizeInfo

7.4.6 recordingInfo

definition	Groups together information on the recording of the audio or video part of a resource
type	component
elements	<p>recordingDeviceType</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>recordingDeviceTypeDetails</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p>recordingPlatformSoftware</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>recordingEnvironment</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>sourceChannel</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>sourceChannelType</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>sourceChannelName</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>sourceChannelDetails</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p>recorder</p> <p><i>Status: optional</i></p>

	<i>Repeatability: unbounded</i>
--	---------------------------------

7.4.6.1 recordingDeviceType

definition	The nature of the recording platform hardware and the storage medium																
type	open controlled vocabulary																
value space	MS-recordingDeviceType																
values	<table> <tr> <th>value</th><th>definition</th></tr> <tr> <td>hardDisk</td><td></td></tr> <tr> <td>dv</td><td></td></tr> <tr> <td>tapeVHS</td><td></td></tr> <tr> <td>DAT</td><td></td></tr> <tr> <td>flash</td><td></td></tr> <tr> <td>soundBlasterCard</td><td></td></tr> <tr> <td>other</td><td></td></tr> </table>	value	definition	hardDisk		dv		tapeVHS		DAT		flash		soundBlasterCard		other	
value	definition																
hardDisk																	
dv																	
tapeVHS																	
DAT																	
flash																	
soundBlasterCard																	
other																	
examples																	
DCLINK																	
ISOcatLINK	http://www.isocat.org/datcat/DC-2693																

7.4.6.2 recordingDeviceTypeDetails

definition	Free text description of the recoding device
type	xs:string
value space	
values	
examples	
DCLINK	
ISOcatLINK	

7.4.6.3 recordingPlatformSoftware

definition	The software used for the recording platform
type	xs:string
value space	
values	
examples	cubase; audition_adobe; soundforge_sony
DCLINK	

ISOcatLINK	http://www.isocat.org/datcat/DC-2694
------------	---

7.4.6.4 recordingEnvironment

definition	Where the recording took place		
type	open controlled vocabulary		
value space	MS-recordingEnvironment		
values	value	definition	ISOcatLink
	office		~ http://www.isocat.org/datcat/DC-2697
	inCar		
	studio		http://www.isocat.org/datcat/DC-2698
		bus; train	
	transport		
	openPublicPlace		http://www.isocat.org/datcat/DC-2700
	closedPublicPlace		
	anechoicChamber		
	conferenceRoom		
	lectureRoom		
	industrial		
	other		
examples			
DCLINK			
ISOcatLINK	http://www.isocat.org/datcat/DC-2568		

7.4.6.5 sourceChannel

definition	Information on the source channel		
type	closed controlled vocabulary		
value space	MS-sourceChannel		
values	value	definition	
	internet		
	radio		
	tv		
	telephone		
	laryngograph		
	airflow		
	EMA		

	other
examples	
DCLINK	
ISOcatLINK	

7.4.6.6 sourceChannelType

definition	Type of the source channel																				
type	open controlled vocabulary																				
value space	MS-sourceChannelType																				
values	<table> <tr> <th>value</th><th>definition</th></tr> <tr> <td>ISDN</td><td></td></tr> <tr> <td>GSM</td><td></td></tr> <tr> <td>3G</td><td></td></tr> <tr> <td>CDMA</td><td></td></tr> <tr> <td>DVB-T</td><td></td></tr> <tr> <td>DVB-S</td><td></td></tr> <tr> <td>DVB-C</td><td></td></tr> <tr> <td>VOIP</td><td></td></tr> <tr> <td>other</td><td></td></tr> </table>	value	definition	ISDN		GSM		3G		CDMA		DVB-T		DVB-S		DVB-C		VOIP		other	
value	definition																				
ISDN																					
GSM																					
3G																					
CDMA																					
DVB-T																					
DVB-S																					
DVB-C																					
VOIP																					
other																					
examples																					
DCLINK																					
ISOcatLINK																					

7.4.6.7 sourceChannelName

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

7.4.6.8 sourceChannelDetails

definition	The details of the channel equipment used (brand, type etc.)
type	xs:string

value space	
values	
examples	
DCLINK	
ISocatLINK	

7.4.6.9 recorder

definition	Information on the recorder(s) of the audio or video part of the resource
type	linked component
component	actorInfo

7.4.7 captureInfo

definition	Groups together information on the capture of the audio or video part of a corpus
type	component
elements	<p>capturingDeviceType</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>capturingDeviceTypeDetails</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p>capturingDetails</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p>capturingEnvironment</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p>sensorTechnology</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>sceneIllumination</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p>personSourceSetInfo</p> <p><i>Status: recommended</i></p>

	<i>Repeatability: 1</i>
--	-------------------------

7.4.7.1 capturingDeviceType

definition	The transducers through which the data is captured																																
type	open controlled vocabulary																																
value space	MS-capturingDeviceType																																
values	<table> <tr> <th>Value</th><th>definition</th></tr> <tr><td>studioEquipment</td><td></td></tr> <tr><td>microphone</td><td></td></tr> <tr><td>closeTalkMicrophone</td><td></td></tr> <tr><td>farfieldMicrophone</td><td></td></tr> <tr><td>lavalierMicrophone</td><td></td></tr> <tr><td>microphoneArray</td><td></td></tr> <tr><td>embeddedMicrophone</td><td></td></tr> <tr><td>largeMembraneMicrophone</td><td></td></tr> <tr><td>laryngograph</td><td></td></tr> <tr><td>telephoneFixed</td><td></td></tr> <tr><td>telephoneMobile</td><td></td></tr> <tr><td>telephoneIP</td><td></td></tr> <tr><td>camera</td><td></td></tr> <tr><td>webcam</td><td></td></tr> <tr><td>other</td><td></td></tr> </table>	Value	definition	studioEquipment		microphone		closeTalkMicrophone		farfieldMicrophone		lavalierMicrophone		microphoneArray		embeddedMicrophone		largeMembraneMicrophone		laryngograph		telephoneFixed		telephoneMobile		telephoneIP		camera		webcam		other	
Value	definition																																
studioEquipment																																	
microphone																																	
closeTalkMicrophone																																	
farfieldMicrophone																																	
lavalierMicrophone																																	
microphoneArray																																	
embeddedMicrophone																																	
largeMembraneMicrophone																																	
laryngograph																																	
telephoneFixed																																	
telephoneMobile																																	
telephoneIP																																	
camera																																	
webcam																																	
other																																	
examples																																	
DCLINK																																	
ISocatLINK																																	

7.4.7.2 capturingDeviceTypeDetails

definition	Provides further information on the capturing device
type	xs:string
value space	
values	
examples	
DCLINK	
ISocatLINK	

7.4.7.3 capturingDetails

definition	Provides further information on the capturing method and procedure
type	xs:string
value space	
values	
examples	
DCLINK	
ISocatLINK	~ http://www.isocat.org/datcat/DC-2563

7.4.7.4 capturingEnvironment

definition	Type of capturing environment
type	closed controlled vocabulary
value space	MS-capturingEnvironment
values	Value definition complex plain
examples	
DCLINK	
ISocatLINK	

7.4.7.5 sensorTechnology

definition	Specifies either the type of image sensor or the sensing method used in the camera or the image-capture device
type	xs:string
value space	
values	
examples	
DCLINK	
ISocatLINK	

7.4.7.6 sceneIllumination

definition	Information on the illumination of the scene
type	closed controlled vocabulary
value space	MS-sceneIllumination

values	Value definition daylight fix multipleSources singleSource variable other
examples	
DCLINK	
ISocatLINK	

7.4.7.7 personSourceSetInfo

definition	Groups information on the persons (speakers, video participants, etc.) in the audio and video parts of the resource
type	component
elements	numberOfPersons <i>Status: optional</i> <i>Repeatability: 1</i> ageOfPersons <i>Status: optional</i> <i>Repeatability: unbounded</i> ageRangeStart <i>Status: optional</i> <i>Repeatability: 1</i> ageRangeEnd <i>Status: optional</i> <i>Repeatability: 1</i> sexOfPersons <i>Status: optional</i> <i>Repeatability: 1</i> originOfPersons <i>Status: optional</i> <i>Repeatability: 1</i> dialectAccentOfPersons <i>Status: optional</i> <i>Repeatability: unbounded</i>

	geographicDistributionOfPersons <i>Status: optional</i> <i>Repeatability: 1</i> hearingImpairmentOfPersons <i>Status: optional</i> <i>Repeatability: 1</i> speakingImpairmentOfPersons <i>Status: optional</i> <i>Repeatability: 1</i> numberOfTrainedSpeakers <i>Status: optional</i> <i>Repeatability: 1</i> speechInfluences <i>Status: optional</i> <i>Repeatability: unbounded</i> participantInfo <i>Status: optional</i> <i>Repeatability: unbounded</i>
--	--

7.4.7.7.1 numberOfPersons

definition	The number of persons participating in the audio or video part of the resource
type	xs:integer
value space	
values	
examples	
DCLINK	
ISOcatLINK	~ http://www.isocat.org/datcat/DC-2692 (only for speakers)

7.4.7.7.2 ageOfPersons

definition	The age range of the group of participants						
type	closed controlled vocabulary						
value space	MS-ageOfPersons						
values	<table> <tr> <td>value</td><td>definition</td></tr> <tr> <td>child</td><td></td></tr> <tr> <td>teenager</td><td></td></tr> </table>	value	definition	child		teenager	
value	definition						
child							
teenager							

	adult elderly
examples	
DCLINK	
ISocatLINK	

7.4.7.7.3 ageRangeStart

definition	Start of age range of the group of participants
type	xs:integer
value space	
values	
examples	
DCLINK	
ISocatLINK	

7.4.7.7.4 ageRangeEnd

definition	End of age range of the group of participants
type	xs:integer
value space	
values	
examples	
DCLINK	
ISocatLINK	

7.4.7.7.5 sexOfPersons

definition	The gender of the group of participants										
type	closed controlled vocabulary										
value space	MS-sexOfPersons										
values	<table> <tr> <th>value</th><th>definition</th></tr> <tr> <td>male</td><td></td></tr> <tr> <td>female</td><td></td></tr> <tr> <td>mixed</td><td></td></tr> <tr> <td>unknown</td><td></td></tr> </table>	value	definition	male		female		mixed		unknown	
value	definition										
male											
female											
mixed											
unknown											
examples											
DCLINK											

ISOcatLINK	~ http://www.isocat.org/datcat/DC-2560 (refers to the sex of each one of the participants)
------------	---

7.4.7.7.6 originOfPersons

definition	The language origin of the group of participants										
type	closed controlled vocabulary										
value space	MS-originOfPersons										
values	<table> <tr> <th>value</th><th>definition</th></tr> <tr> <td>native</td><td></td></tr> <tr> <td>nonNative</td><td></td></tr> <tr> <td>mixed</td><td></td></tr> <tr> <td>unknown</td><td></td></tr> </table>	value	definition	native		nonNative		mixed		unknown	
value	definition										
native											
nonNative											
mixed											
unknown											
examples											
DCLINK											
ISOcatLINK											

7.4.7.7.7 dialectAccentOfPersons

definition	Provides information on the dialect of the group of participants
type	xs:string
value space	
values	
examples	
DCLINK	
ISOcatLINK	

7.4.7.7.8 geographicDistributionOfPersons

definition	Gives information on the geographic distribution of the participants
type	xs:string
value space	
values	
examples	
DCLINK	
ISOcatLINK	

7.4.7.7.9 hearingImpairmentOfPersons

definition	Whether the group of participants contains persons with hearing impairments
type	closed controlled vocabulary
value space	MS-hearingImpairmentOfPersons
values	<div>value definition</div> <div>yes</div> <div>no</div> <div>mixed</div>
examples	
DCLINK	
ISOcatLINK	

7.4.7.7.10 speakingImpairmentOfPersons

definition	Whether the group of participants contains persons withwith speakingimpairments
type	closed controlled vocabulary
value space	MS-hearingImpairmentOfPersons
values	<div>value definition</div> <div>yes</div> <div>no</div> <div>mixed</div>
examples	
DCLINK	
ISOcatLINK	

7.4.7.7.11numberOfTrainedSpeakers

definition	The number of participants that have been trained for the specific task
type	xs:integer
value space	
values	
examples	
DCLINK	
ISOcatLINK	

7.4.7.7.12 **speechInfluences**

definition	Specifies the factors influencing speech												
type	open controlled vocabulary												
value space	MS-speechInfluences												
values	<table> <tr> <th>value</th><th>definition</th></tr> <tr> <td>alcohol</td><td></td></tr> <tr> <td>sleepDeprivation</td><td></td></tr> <tr> <td>hyperbaric</td><td></td></tr> <tr> <td>medication</td><td></td></tr> <tr> <td>other</td><td></td></tr> </table>	value	definition	alcohol		sleepDeprivation		hyperbaric		medication		other	
value	definition												
alcohol													
sleepDeprivation													
hyperbaric													
medication													
other													
examples													
DCLINK													
ISOcatLINK													
comment	restricted to audio parts of the resource												

7.4.7.7.13 **participantInfo**

definition	Groups information on the individual person(s) participating in the audio, video, sensorimotor (textNumerical) part of the resource
type	component
elements	<p>alias</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p>ageGroup</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p>age</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p>sex</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p>origin</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p>placeOfLiving</p> <p><i>Status: optional</i></p>

	<p><i>Repeatability: 1</i></p> <p><u>placeOfBirth</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p><u>placeOfChildhood</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p><u>dialectAccent</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p><u>speakingImpairment</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p><u>hearingImpairment</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p><u>smokingHabits</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p><u>vocalTractConditions</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p><u>profession</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p><u>height</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p><u>weight</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p><u>trainedSpeaker</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p><u>placeOfSecondEducation</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p>
--	---

	educationLevel <i>Status: optional</i> <i>Repeatability: 1</i>
--	--

7.4.7.7.13.1 alias

definition	The name of the person used instead of the real one
type	xs:string
value space	
values	
examples	
DCLINK	
ISocatLINK	~ http://www.isocat.org/datcat/DC-2552

7.4.7.7.13.2 ageGroup

definition	The age group to which the participant belongs										
type	closed controlled vocabulary										
value space	MS-ageGroup										
values	<table border="1"> <thead> <tr> <th>value</th><th>definition</th></tr> </thead> <tbody> <tr> <td>child</td><td></td></tr> <tr> <td>teenager</td><td></td></tr> <tr> <td>adult</td><td></td></tr> <tr> <td>elderly</td><td></td></tr> </tbody> </table>	value	definition	child		teenager		adult		elderly	
value	definition										
child											
teenager											
adult											
elderly											
examples											
DCLINK											
ISocatLINK											

7.4.7.7.13.3 age

definition	The age of the participant
type	free text
value space	
values	
examples	
DCLINK	
ISocatLINK	http://www.isocat.org/datcat/DC-2550

7.4.7.7.13.4 origin

definition	The language origin of the participant
type	closed controlled vocabulary
value space	MS-origin
values	<div>value definition</div> <div>native</div> <div>nonNative</div> <div>unknown</div>
examples	
DCLINK	
ISocatLINK	

7.4.7.7.13.5 placeOfLiving

definition	The participant's place of living
type	xs:string
value space	
values	
examples	
DCLINK	
ISocatLINK	

7.4.7.7.13.6 placeOfBirth

definition	The place in which the participant has been born
type	xs:string
value space	
values	
examples	
DCLINK	
ISocatLINK	

7.4.7.7.13.7 placeOfChildhood

definition	The place in which the participant lived as a child
type	xs:string
value space	
values	

examples	
DCLINK	
ISocatLINK	

7.4.7.7.13.8dialectAccent

definition	Provides information on the dialect of the participant
type	xs:string
value space	
values	
examples	
DCLINK	
ISocatLINK	~ http://www.isocat.org/datcat/DC-2553 (dominant language)

7.4.7.7.13.9speakingImpairment

definition	Provides information on any speaking impairment the participant may have
type	xs:string
value space	
values	
examples	
DCLINK	
ISocatLINK	

7.4.7.7.13.10 hearingImpairment

definition	Provides information on any hearing impairment the participant may have
type	xs:string
value space	
values	
examples	
DCLINK	
ISocatLINK	

7.4.7.7.13.11 smokingHabits

definition	Provides information on whether the participant smokes and on his/her smoking habits in general
------------	---

type	xs:string
value space	
values	
examples	
DCLINK	
ISOcatLINK	

7.4.7.7.13.12 vocalTractConditions

definition	Provides information on the vocal tract conditions that may influence the speech of the participant	
type	open controlled vocabulary	
value space	MS-vocalTractConditions	
values	value	definition
	dentalProsthesis	
	other	
examples		
DCLINK		
ISOcatLINK		

7.4.7.7.13.13 profession

definition	Provides information on the participant's profession	
type	xs:string	
value space		
values		
examples		
DCLINK		
ISOcatLINK	http://www.isocat.org/datcat/DC-2558	

7.4.7.7.13.14 height

definition	Provides information on the height of the participant in cm	
type	xs:integer	
value space		
values		
examples		
DCLINK		

ISOcatLINK	
------------	--

7.4.7.7.13.15 weight

definition	Provides information on the weight of the participant
type	xs:integer
value space	
values	
examples	
DCLINK	
ISOcatLINK	

7.4.7.7.13.16 trainedSpeaker

definition	Provides information on whether the participant is trained in a specific task
type	xs:boolean
value space	
values	
examples	
DCLINK	
ISOcatLINK	

7.4.7.7.13.17 placeOfSecondEducation

definition	Specifies the place of the secondary education of the participant
type	xs:string
value space	
values	
examples	
DCLINK	
ISOcatLINK	

7.4.7.7.13.18 educationLevel

definition	Provides information on the education level of the participant
type	xs:string
value space	
values	

examples	
DCLINK	
ISocatLINK	http://www.isocat.org/datcat/DC-2554

7.5 corpusVideoInfo

definition	Groups together information on the video component of a corpus
type	component
elements	<p>mediaType</p> <p><i>Status: mandatory with value "video"</i></p> <p><i>Repeatability: 1</i></p> <p>lingualityInfo</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p>languageInfo</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>modalityInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p> <p>sizeInfo</p> <p><i>Status: mandatory</i></p> <p><i>Repeatability: unbounded</i></p> <p>videoContentInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p> <p>settingInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p> <p>videoFormatInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>annotationInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>domainInfo</p> <p><i>Status: recommended</i></p>

	<p><i>Repeatability: unbounded</i></p> <p>timeCoverageInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>geographicCoverageInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>videoClassificationInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>recordingInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p> <p>captureInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p> <p>creationInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p> <p>linkToOtherMediaInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p>
--	--

7.5.1 videoContentInfo

definition	Groups together information on the contents of the video part of a resource
type	component
elements	<p>typeOfVideoContent</p> <p><i>Status: mandatory</i></p> <p><i>Repeatability: unbounded</i></p> <p>textIncludedInVideoType</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>dynamicElementInfo</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p>

comments	
----------	--

7.5.1.1 typeOfVideoContent

definition	Main type of object or people represented in the video
type	xs:string
value space	
values	
examples	animated; true life
DCLINK	
ISOcatLINK	
comments	

7.5.1.2 textIncludedInVideoType

definition	Indicates if text is present in or in conjunction with the video								
type	closed controlled vocabulary								
value space	MS-textIncludedInVideoType								
values	<table> <tr> <td>value</td><td>definition</td></tr> <tr> <td></td><td>captions</td></tr> <tr> <td></td><td>subtitles</td></tr> <tr> <td></td><td>none</td></tr> </table>	value	definition		captions		subtitles		none
value	definition								
	captions								
	subtitles								
	none								
examples									
DCLINK									
ISOcatLINK									

7.5.1.3 dynamicElementInfo

definition	Groups information on the dynamic elements that are represented in the video part of the resource
type	component
elements	typeOfElement <i>Status: optional</i> <i>Repeatability: unbounded</i> bodyParts <i>Status: optional</i> <i>Repeatability: unbounded</i>

	<p><u>distractors</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p><u>interactiveMedia</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p><u>faceViews</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p><u>faceExpressions</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p><u>bodyMovement</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p><u>gestures</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p><u>handArmMovement</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p><u>handManipulation</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p><u>headMovement</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p><u>eyeMovement</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p><u>posesPerSubject</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p>
comments	

7.5.1.3.1 typeOfElement

definition	The type of objects or people that represented in the video or image part of the resource
type	xs:string
value space	
values	
examples	
DCLINK	
ISocatLINK	
comments	

7.5.1.3.2 bodyParts

definition	The body parts visible in the video or image part of the resource	
type	closed controlled vocabulary	
value space	MS-bodyParts	
values	value	definition
	arms	
	face	
	feet	
	hands	
	head	
	legs	
	mouth	
	wholeBody	
	none	
examples		
DCLINK		
ISocatLINK		
comments		

7.5.1.3.3 distractors

definition	Any distractors visible in the resource
type	xs:string
value space	
values	

examples	glasses; hat; markers; microphone; mobilePhone; pen; paper; notepad; none
DCLINK	
ISOcatLINK	
comments	

7.5.1.3.4 interactiveMedia

definition	Any interactive media visible in the resource
type	xs:string
value space	
values	
examples	computerPen; dataGlove; desktopPC; graphicalScreen; laptop; mousePDA; tactileScreen
DCLINK	
ISOcatLINK	
comments	

7.5.1.3.5 faceViews

definition	Indicates the view of the face(s) that appear in the video or on the image part of the resource
type	xs:string
value space	
values	
examples	frontal; profile
DCLINK	
ISOcatLINK	
comments	

7.5.1.3.6 faceExpressions

definition	Indicates the facial expressions visible in the resource
type	xs:string
value space	
values	
examples	anger, disgust, fear, happiness, sadness, surprise, boredom; faceNotVisible; neutral
DCLINK	

ISocatLINK	
------------	--

7.5.1.3.7 bodyMovement

definition	Indicates the body parts that move in the video part of the resource
type	xs:string
value space	
values	
examples	head; arms; legs; lowerBody; upperBody; wholeBody
DCLINK	
ISocatLINK	
comments	

7.5.1.3.8 gestures

definition	Indicates the type of gestures visible in the resource
type	xs:string
value space	
values	
examples	emotional; interactional; nonInteractional; unidentified
DCLINK	
ISocatLINK	
comments	

7.5.1.3.9 handArmMovement

definition	Indicates the movement of hands and/or arms visible in the resource
type	xs:string
value space	
values	
examples	right hand; left hand; right arm; left arm; pair; complex; function; spatial
DCLINK	
ISocatLINK	
comments	

7.5.1.3.10 handManipulation

definition	Gives information on the manipulation of objects by hand
------------	--

type	xs:string
value space	
values	
examples	changingPosition; direct; indirect; joining; splitting; modifyingObjects
DCLINK	
ISOcatLINK	
comments	

7.5.1.3.11 headMovement

definition	Indicates the movements of the head visible in the resource
type	xs:string
value space	
values	
examples	backward; forward; rotationLeft; rotationRight; sidewardLeft; sidewardRight
DCLINK	
ISOcatLINK	
comments	

7.5.1.3.12 eyeMovement

definition	Indicates the movement of the eyes visible in the resource
type	xs:string
value space	
values	
examples	convergence; horizontal; pursuit; saccades; vertical
DCLINK	
ISOcatLINK	
comments	

7.5.1.3.13 posesPerSubject

definition	The number of poses per subject that participates in the video part of the resource
type	xs:int
value space	
values	

examples	
DCLINK	
ISOCatLINK	
comments	

7.5.2 videoFormatInfo

definition	Groups information on the format(s) of a resource; repeated if parts of the resource are in different formats
type	component
elements	<p>mimeType</p> <p><i>Status: mandatory</i></p> <p><i>Repeatability: unbounded</i></p> <p>colourSpace</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>colourDepth</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p>frameRate</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p>resolutionInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>visualModelling</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p>fidelity</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p>compressionInfo</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p>recordingQuality</p> <p><i>Status: optional</i></p>

	<i>Repeatability: 1</i> sizePerVideoFormat <i>Status: optional</i> <i>Repeatability: 1</i>
--	---

7.5.2.1 colourSpace

definition	Defines the colour space for the video										
type	closed controlled vocabulary										
value space	MS-colourSpace										
values	<table> <tr> <th>value</th><th>definition</th></tr> <tr> <td>RGB</td><td></td></tr> <tr> <td>CMYK</td><td></td></tr> <tr> <td>4:2:2</td><td></td></tr> <tr> <td>YUV</td><td></td></tr> </table>	value	definition	RGB		CMYK		4:2:2		YUV	
value	definition										
RGB											
CMYK											
4:2:2											
YUV											
examples											
DCLINK											
ISocatLINK											

7.5.2.2 colourDepth

definition	The number of bits used to represent the colour of a single pixel
type	xs:int
value space	
values	
examples	8; 16; 24; 30; 36
DCLINK	
ISocatLINK	
comments	

7.5.2.3 frameRate

definition	The number of frames per second
type	xs:int
value space	
values	
examples	25; 30

DCLINK	
ISOcatLINK	
comments	Hz, fps ?

7.5.2.4 resolutionInfo

definition	Groups together information on the image resolution
type	component
elements	sizeWidth <i>Status: optional</i> <i>Repeatability: 1</i> sizeHeight <i>Status: optional</i> <i>Repeatability: 1</i> resolutionStandard <i>Status: optional</i> <i>Repeatability: 1</i>
comments	

7.5.2.4.1 sizeWidth

definition	The frame width in pixels
type	xs:int
value space	
values	
examples	
DCLINK	
ISOcatLINK	
comments	

7.5.2.4.2 sizeHeight

definition	The frame height in pixels
type	xs:int
value space	
values	
examples	
DCLINK	

ISOcatLINK	
comments	

7.5.2.4.3 resolutionStandard

definition	The standard to which the resolution conforms										
type	open controlled vocabulary										
value space	MS-resolutionStandard										
values	<table> <tr> <th>value</th><th>definition</th></tr> <tr> <td>VGA</td><td></td></tr> <tr> <td>HD.720</td><td></td></tr> <tr> <td>HD.1080</td><td></td></tr> <tr> <td>other</td><td></td></tr> </table>	value	definition	VGA		HD.720		HD.1080		other	
value	definition										
VGA											
HD.720											
HD.1080											
other											
examples											
DCLINK											
ISOcatLINK											
comments											

7.5.2.5 visualModelling

definition	The dimensional form applied on video or image corpus						
type	closed controlled vocabulary						
value space	MS-visualModelling						
values	<table> <tr> <th>value</th><th>definition</th></tr> <tr> <td>2D</td><td></td></tr> <tr> <td>3D</td><td></td></tr> </table>	value	definition	2D		3D	
value	definition						
2D							
3D							
examples							
DCLINK							
ISOcatLINK							
comments							

7.5.2.6 fidelity

definition	Defines whether blur is present in the moving sequences
type	xs:boolean
value space	
values	

examples	
DCLINK	
ISOcatLINK	
comments	

7.5.2.7 sizePerVideoFormat

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

7.5.3 videoClassificationInfo

definition	Groups together information on video genre of the resource
type	component
elements	videoGenre <i>Status: recommended</i> <i>Repeatability: 1</i> subject topic <i>Status: optional</i> <i>Repeatability: 1</i> conformanceToClassificationScheme <i>Status: optional</i> <i>Repeatability: 1</i> sizePerVideoClassification <i>Status: optional</i> <i>Repeatability: 1</i>
comments	

7.5.3.1 videoGenre

definition	A first indication of type of video recorded
type	xs:string
value space	
values	
examples	news; sports; documentary; tvSeries; videotapedEvent
DCLINK	

ISOcatLINK	
comments	

7.5.3.2 sizePerVideoClassification

definition	Used to give info on size of parts with different video classification
type	special status component
component	sizeInfo

7.6 corpusImageInfo

definition	Groups together information on the image component of a resource
type	component
elements	<p>mediaType</p> <p><i>Status: mandatory with value "image"</i></p> <p><i>Repeatability: 1</i></p> <p><i>Repeatability: 1</i></p> <p>lingualityInfo</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p>languageInfo</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>modalityInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p> <p>sizeInfo</p> <p><i>Status: mandatory</i></p> <p><i>Repeatability: unbounded</i></p> <p>imageContentInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p> <p>imageFormatInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>annotationInfo</p> <p><i>Status: recommended</i></p>

	<p><i>Repeatability: unbounded</i></p> <p>domainInfo</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>timeCoverageInfo</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>geographicCoverageInfo</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>imageClassificationInfo</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>captureInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p> <p>creationInfo</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p>linkToOtherMediaInfo</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p>
--	---

7.6.1 imageContentInfo

definition	Groups together information on the contents of the image part of a resource
type	component
element	<p>typeOfImageContent</p> <p><i>Status: mandatory</i></p> <p><i>Repeatability: unbounded</i></p> <p>textIncludedInImage</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>staticElementInfo</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p>

comments	
----------	--

7.6.1.1 typeOfImageContent

definition	The main types of object or people represented in the image corpus
type	xs:string
value space	
values	
examples	car; people; mountain; scenery; objects; buildings; people; faces; events; biometrical measurements; diagrams etc.
DCLINK	
ISOcatLINK	
comments	

7.6.1.2 textIncludedInImage

definition	Provides information on the type of text that may be on the image										
type	closedcontrolled vocabulary										
value space	MS-textIncludedInImage										
values	<table> <tr> <th>value</th><th>definition</th></tr> <tr> <td>captions</td><td></td></tr> <tr> <td>subtitles</td><td></td></tr> <tr> <td>captureTime</td><td></td></tr> <tr> <td>none</td><td></td></tr> </table>	value	definition	captions		subtitles		captureTime		none	
value	definition										
captions											
subtitles											
captureTime											
none											
examples											
DCLINK											
ISOcatLINK											
comments											

7.6.1.3 StaticElementInfo

definition	Groups information on the static element visible on the images
type	component
element	typeOfElement <i>Status: mandatory</i> <i>Repeatability: unbounded</i> bodyParts <i>Status: optional</i>

	<p><i>Repeatability: unbounded</i></p> <p><u>faceExpressions</u></p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p><u>faceViews</u></p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p><u>artifactParts</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded¹</i></p> <p><u>landscapeParts</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p><u>personDescription</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p><u>thingDescription</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p><u>organizationDescription</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p><u>eventDescription</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p>
--	---

7.6.1.3.1 artifactParts

definition	Indicates the parts of the artifacts represented in the image corpus
type	xs:string
value space	
values	
examples	
DCLINK	
ISocatLINK	

7.6.1.3.2 landscapeParts

definition	Indicates the landscape parts represented in the image corpus
type	xs:string
value space	
values	
examples	
DCLINK	
ISOcatLINK	

7.6.1.3.3 personDescription

definition	Provides descriptive features for the persons represented in the image corpus
type	xs:string
value space	
values	
examples	
DCLINK	
ISOcatLINK	

7.6.1.3.4 thingDescription

definition	Provides description of the things represented in the image corpus
type	xs:string
value space	
values	
examples	
DCLINK	
ISOcatLINK	

7.6.1.3.5 organizationDescription

definition	Provides description of the organizations that may appear in the image corpus
type	xs:string
value space	
values	
examples	

DCLINK	
ISOcatLINK	

7.6.1.3.6 eventDescription

definition	Provides description of any events represented in the image corpus
type	xs:string
value space	
values	
examples	
DCLINK	
ISOcatLINK	

7.6.2 imageFormatInfo

definition	Groups information on the format of the image component of the resource
type	component
element	<p>mimeType</p> <p><i>Status: mandatory</i> <i>Repeatability: unbounded</i></p> <p>colourSpace</p> <p><i>Status: recommended</i> <i>Repeatability: unbounded</i></p> <p>colourDepth</p> <p><i>Status: optional</i> <i>Repeatability: unbounded</i></p> <p>compressionInfo</p> <p><i>Status: optional</i> <i>Repeatability: 1</i></p> <p>resolutionInfo</p> <p><i>Status: optional</i> <i>Repeatability: unbounded</i></p> <p>visualModelling</p> <p><i>Status: optional</i> <i>Repeatability: 1</i></p> <p>rasterOrVectorGraphics</p> <p><i>Status: optional</i></p>

	<i>Repeatability: 1</i> quality <i>Status: optional</i> <i>Repeatability: 1</i> sizePerImageFormat <i>Status: optional</i> <i>Repeatability: 1</i>
comments	

7.6.2.1 rasterOrVectorGraphics

definition	Indicates if the image is stored as raster or vector graphics
type	closed controlled vocabulary
value space	MS-rasterOrVectorGraphics
values	value definition raster vector
examples	
DCLINK	
ISocatLINK	

7.6.2.2 quality

definition	Specifies the quality level of image resource
type	closed controlled vocabulary
value space	MS-quality
values	value definition veryLow low medium high veryHigh
examples	veryLow; low; medium; high; veryHigh
DCLINK	
ISocatLINK	http://www.isocat.org/datcat/DC-2574

7.6.2.3 sizePerImageFormat

definition	Used to give info on size of parts of a resource that differ as to the format
type	linked component
component	sizeInfo

7.6.3 imageClassificationInfo

definition	Groups information on the classification of the image corpus
type	component
element	imageGenre <i>Status: optional</i> <i>Repeatability: 1</i> subject topic <i>Status: optional</i> <i>Repeatability: 1</i> conformanceToClassificationScheme <i>Status: optional</i> <i>Repeatability: 1</i> sizePerImageClassification <i>Status: optional</i> <i>Repeatability: unbounded</i>
comments	

7.6.3.1 imageGenre

definition	A first indication of the genre of images
type	xs:string
value space	
values	
examples	portraits, still images, etc.
DCLINK	
ISOcatLINK	

7.6.3.2 sizePerImageClassification

definition	Provides information on size of parts with different image classification
type	special status component

component	sizeInfo
-----------	--------------------------

7.7 corpusTextNumericalInfo

definition	Groups together information on the textNumerical component of a corpus. It is used basically for the textual representation of measurements and observations linked to sensorimotor recordings
type	component
elements	<p>mediaType</p> <p><i>Status: mandatory with value "textNumerical"</i></p> <p><i>Repeatability: 1</i></p> <p>modalityInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p> <p>sizeInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>textNumericalContentInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p> <p>textNumericalFormatInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>recordingInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p> <p>captureInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p> <p>creationInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p> <p>annotationInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p> <p>linkToOtherMediaInfo</p> <p><i>Status: recommended</i></p>

	<i>Repeatability: unbounded</i>
comments	

7.7.1 textNumericalContentInfo

definition	Groups information on the content of the textNumerical part of the resource
type	component
elements	typeOfTextNumericalContent <i>Status: mandatory</i> <i>Repeatability: unbounded</i>

7.7.1.1 typeOfTextNumericalContent

definition	Specifies the content that is represented in the textNumerical part of the resource
type	xs:string
value space	
values	
examples	temperature, coordinates etc.
DCLINK	
ISocatLINK	

7.7.2 textNumericalFormatInfo

definition	Groups information on the format(s) of the textNumerical part of the resource
type	component
elements	mimeType <i>Status: mandatory</i> <i>Repeatability: 1</i> sizePerTextNumericalFormat <i>Status: optional</i> <i>Repeatability: 1</i>

7.7.2.1 sizePerTextNumericalFormat

definition	Gives information on the size of textNumerical resource parts with different format
------------	---

type	linked component
component	sizeInfo

7.8 corpusTextNgramInfo

definition	Groups together information required for n-gram resources; information can be provided both as regards features drawn from the source corpus (e.g. language coverage, size, format, domains etc.) and features pertaining to the n-gram output itself (e.g. range of n-grams, type of item included, etc.)
type	component
elements	<p>mediaType <i>Status: mandatory with value textNgram</i> <i>Repeatability: 1</i></p> <p>ngramInfo <i>Status: mandatory</i> <i>Repeatability: 1</i></p> <p>lingualityInfo <i>Status: mandatory</i> <i>Repeatability: 1</i></p> <p>languageInfo <i>Status: mandatory</i> <i>Repeatability: unbounded</i></p> <p>modalityInfo <i>Status: recommended</i> <i>Repeatability: 1</i></p> <p>sizeInfo <i>Status: mandatory</i> <i>Repeatability: unbounded</i></p> <p>textFormatInfo <i>Status: recommended</i> <i>Repeatability: unbounded</i></p> <p>characterEncodingInfo <i>Status: recommended</i> <i>Repeatability: unbounded</i></p> <p>annotationInfo <i>Status: recommended</i> <i>Repeatability: unbounded</i></p>

	domainInfo <i>Status: recommended</i> <i>Repeatability: unbounded</i> textClassificationInfo <i>Status: recommended</i> <i>Repeatability: unbounded</i> timeCoverageInfo <i>Status: recommended</i> <i>Repeatability: unbounded</i> geographicCoverageInfo <i>Status: recommended</i> <i>Repeatability: unbounded</i> creationInfo <i>Status: recommended</i> <i>Repeatability: 1</i>
comment	<p>it should be clear that this is intended to describe the n-gram corpus and not the source corpus, i.e. annotation and creation components must include information on the procedure, methods and tools used to create and enrich the n-gram corpus itself rather than the annotation and creation of the source corpus; the source corpus can be described as a separate resource and the element "originalSource" (in the creationInfo component) can be used to provide the appropriate link</p>

7.8.1 ngramInfo

definition	Groups information specific to n-gram resources (e.g. range of n-grams, base item etc.)
type	component
elements	baseItem <i>Status: mandatory</i> <i>Repeatability: unbounded</i> order <i>Status: mandatory</i> <i>Repeatability: 1</i> perplexity <i>Status: recommended</i>

	<i>Repeatability: 1</i> isFactored <i>Status: recommended</i> <i>Repeatability: 1</i> factors <i>Status: recommended</i> <i>Repeatability: unbounded</i> smoothing <i>Status: recommended</i> <i>Repeatability: 1</i> interpolated <i>Status: optional</i> <i>Repeatability: 1</i>
--	--

7.8.1.1 baseItem

definition	Type of item that is represented in the n-gram resource												
type	open controlled vocabulary												
value space	MS-baseItem												
values	<table> <tr> <th>value</th><th>definition</th></tr> <tr> <td>word</td><td></td></tr> <tr> <td>syllable</td><td></td></tr> <tr> <td>letter</td><td></td></tr> <tr> <td>phoneme</td><td></td></tr> <tr> <td>other</td><td></td></tr> </table>	value	definition	word		syllable		letter		phoneme		other	
value	definition												
word													
syllable													
letter													
phoneme													
other													
examples													
DCLINK													
ISOcatLINK													
comments													

7.8.1.2 order

definition	The maximum number of items in the sequence
type	xs:int
value space	
values	
examples	

DCLINK	
ISocatLINK	
comments	

7.8.1.3 perplexity

definition	Derived from running on test set taken from the same corpus
type	xs:double
value space	
values	
examples	
DCLINK	
ISocatLINK	
comments	

7.8.1.4 isFactored

definition	Specifies whether the model is factored or not
type	xs:boolean
value space	
values	
examples	
DCLINK	
ISocatLINK	
comments	

7.8.1.5 factors

definition	The list of factors that have been used for the n-gram model
type	xs:string
value space	
values	
examples	POS tags in Penn Treebank: NN, DT, JJ
DCLINK	
ISocatLINK	
comments	These can be POS tags, NER tags, etc.

7.8.1.6 smoothing

definition	The technique used for giving unseen items some probability
type	xs:string
value space	
values	
examples	Knesser-Ney, Good-Turing, ...
DCLINK	
ISocatLINK	
comments	

7.8.1.7 interpolated

definition	Interpolated language models are constructed by 2 or more corpora. Each corpus is represented in the model according to a predefined weight.
type	xs:string
value space	
values	
examples	
DCLINK	
ISocatLINK	
comments	

8 Tools/Services

8.1 resourceInfo

definition	Extension of Resource for Tools and Services
elements	<p>resourceInfo – common components</p> <ul style="list-style-type: none"> identificationInfo distributionInfo contactPerson metadataInfo versionInfo validationInfo usageInfo resourceDocumentationInfo resourceCreationInfo

	relationInfo resourceComponentType toolServiceInfo
--	--

8.2 toolServiceInfo

definition	Groups together elements required for the description of tools and services
type	component
elements	resourceType <i>Status: mandatory with value "toolService"</i> <i>Repeatability: 1</i> toolServiceType <i>Status: mandatory</i> <i>Repeatability: 1</i> toolServiceSubtype <i>Status: optional</i> <i>Repeatability: unbounded</i> languageDependent <i>Status: mandatory</i> <i>Repeatability: 1</i> inputInfo <i>Status: recommended</i> <i>Repeatability: 1</i> outputInfo <i>Status: recommended</i> <i>Repeatability: 1</i> toolServiceOperationInfo <i>Status: recommended</i> <i>Repeatability: 1</i> toolServiceEvaluationInfo <i>Status: recommended</i> <i>Repeatability: 1</i> toolServiceCreationInfo <i>Status: optional</i> <i>Repeatability: 1</i>

8.2.1 toolServiceType

definition	Specifies the type of the tool or service	
type	open controlled vocabulary	
value space	MS-toolServiceType	
values	<div>value</div> <div>tool</div> <div>service</div> <div>platform</div> <div>suiteOfTools</div> <div>infrastructure</div> <div>architecture</div> <div>nlpDevelopmentEnvironment</div> <div>other</div>	<div>definition</div> <div>A device that performs one or more of the tasks listed in the ForeseenUseInfo component</div> <div>A form in which NLP tasks are realized and delivered to a user, without the need of acquiring and installing the corresponding tools</div> <div>A technology that eases the development of new tools and services in the NLP field</div> <div>A more or less fixed integrated organisation of tools applied for solving a combination of tasks</div> <div>A technology that supports the flexible development of NLP applications, together with all the requested resources</div> <div>A technology that supports the development of data resources, like lexicons, grammars, corpora, etc. Can be included in an Architecture or in a Platform</div>
examples		
DCLINK		
ISocatLINK		

8.2.2 toolServiceSubtype

definition	Specifies the subtype of tool or service
type	xs:string
value space	

values	
examples	parser, tagger, annotator, corpus workbench etc.
DCLINK	
ISOcatLINK	

8.2.3 languageDependent

definition	Indicates whether the operation of the tool or service is language dependent or not
type	xs:boolean
value space	
values	
examples	
DCLINK	
ISOcatLINK	

8.2.4 inputInfo

definition	Groups together information on the requirements set on the input resource of a tool or service
type	component
elements	<p>mediaType</p> <p><i>Status: mandatory</i></p> <p><i>Repeatability: unbounded</i></p> <p>resourceType</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>modalityType</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>languageName</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>languageId</p> <p><i>Status: optional</i></p>

	<p><i>Repeatability: unbounded</i></p> <p>languageVarietyName</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>mimeType</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>characterEncoding</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>annotationType</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>annotationFormat</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>tagset</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>segmentationLevel</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>conformanceToStandardsBestPractices</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p>
--	---

8.2.4.1 resourceType

definition	The type of the resource that a tool or service takes as input or produces as output
type	closed controlled vocabulary
value space	MS-resourceType
values	<div>value</div> <div>definition</div> <div>corpus</div> <div>lexicalConceptualResource</div> <div>languageDescription</div>
examples	

DCLINK	
ISOcatLINK	
comments	

8.2.5 outputInfo

definition	Groups together information on the requirements set on the output of a tool or service
type	component
elements	<p>mediaType</p> <p><i>Status: mandatory</i></p> <p><i>Repeatability: unbounded</i></p> <p>resourceType</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>modalityType</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>languageName</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>languageId</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>languageVarietyName</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>mimeType</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>characterEncoding</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>annotationType</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p>

	annotationFormat <i>Status: optional</i> <i>Repeatability: unbounded</i> tagset <i>Status: optional</i> <i>Repeatability: unbounded</i> segmentationLevel <i>Status: optional</i> <i>Repeatability: unbounded</i> conformanceToStandardsBestPractices <i>Status: optional</i> <i>Repeatability: unbounded</i>
--	--

8.2.6 toolServiceOperationInfo

definition	Groups together information on the operation of a tool or service
type	component
elements	operatingSystem <i>Status: mandatory</i> <i>Repeatability: unbounded</i> runningEnvironmentInfo <i>Status: recommended</i> <i>Repeatability: 1</i> runningTime <i>Status: optional</i> <i>Repeatability: 1</i>

8.2.6.1 operatingSystem

definition	The operating system on which the tool will be running										
type	open controlled vocabulary										
value space	MS-operatingSystem										
values	<table> <tr> <td>value</td><td>definition</td></tr> <tr> <td>os-independent</td><td></td></tr> <tr> <td>windows</td><td></td></tr> <tr> <td>linux</td><td></td></tr> <tr> <td>unix</td><td></td></tr> </table>	value	definition	os-independent		windows		linux		unix	
value	definition										
os-independent											
windows											
linux											
unix											

	mac-OS other
examples	
DCLINK	
ISocatLINK	http://www.isocat.org/datcat/DC-2572

8.2.6.2 runningTime

definition	Gives information on the running time of a tool or service
type	xs:string
value space	
values	
examples	
DCLINK	
ISocatLINK	
comment	recommended format: expressed in seconds

8.2.6.3 runningEnvironmentInfo

definition	Groups together information on the running environment of a tool or a language description
type	component
elements	requiredSoftware <i>Status: recommended</i> <i>Repeatability: unbounded</i> requiredHardware <i>Status: optional</i> <i>Repeatability: unbounded</i> requiredLRs <i>Status: optional</i> <i>Repeatability: unbounded</i> runningEnvironmentDetails <i>Status: optional</i> <i>Repeatability: 1</i>

8.2.6.3.1 requiredSoftware

definition	Additional software required for running a tool and/or computational
------------	--

	grammar
type	linked component
component	targetResourceInfo
examples	
DCLINK	
ISOcatLINK	~ http://www.isocat.org/datcat/DC-2576

8.2.6.3.2requiredHardware

definition	Hardware required for running a tool and/or computational grammar
type	open controlled vocabulary
value space	MS-requiredHardware
values	Value annotation graphicCard microphone ocrSystem specialHardwareEquipment none other
examples	
DCLINK	
ISOcatLINK	~ http://www.isocat.org/datcat/DC-2576

8.2.6.3.3requiredLRs

definition	If for running a tool and/or computational grammar, specific LRs (e.g. a grammar, a list of words etc.) are required
type	linked component
value space	targetResourceInfo
examples	
DCLINK	
ISOcatLINK	

8.2.6.3.4runningEnvironmentDetails

definition	Provides further information on the running environment
type	xs:string
value space	

values	
examples	
DCLINK	
ISocatLINK	

8.2.7 toolServiceEvaluationInfo

definition	Groups together information on the evaluation status of a tool or service
type	component
elements	<p>evaluated</p> <p><i>Status: mandatory</i></p> <p><i>Repeatability: 1</i></p> <p>evaluationLevel</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>evaluationType</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>evaluationCriteria</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>evaluationMeasure</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>evaluationReport</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>evaluationTool</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>evaluationDetails</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p> <p>evaluator</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p>

8.2.7.1 evaluated

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

8.2.7.2 evaluationLevel

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

8.2.7.3 evaluationType

definition	Indicates the evaluation type
type	closed controlled vocabulary
value space	MS-evaluationType
values	value defintion glassBox blackBox
examples	
DCLINK	
ISocatLINK	

8.2.7.4 evaluationCriteria

definition	Defines the criteria of the evaluation of a tool
------------	--

type	closed controlled vocabulary
value space	MS-evaluationCriteria
values	value definition extrinsic intrinsic
examples	
DCLINK	
ISocatLINK	

8.2.7.5 evaluationMeasure

definition	Defines whether the evaluation measure is human or automatic
type	closed controlled vocabulary
value space	MS-evaluationMeasure
values	value definition human automatic
examples	
DCLINK	
ISocatLINK	

8.2.7.6 evaluationReport

definition	A bibliographical record of or link to a report describing the evaluation process, tool, method etc. of the tool or service
type	linked component
value space	documentationInfo
examples	
DCLINK	
ISocatLINK	

8.2.7.7 evaluationTool

definition	The name or id or url of the tool used for the evaluation of the tool or service
type	linked component
value space	targetResourceInfo
examples	

DCLINK	
ISOcatLINK	

8.2.7.8 evaluationDetails

definition	Provides further information on the evaluation process of a tool or service
type	xs:string
value space	
values	
examples	
DCLINK	
ISOcatLINK	

8.2.7.9 evaluator

definition	Groups information on person or organization that evaluated the tool or service
type	linked component
component	actorInfo

8.2.8 toolServiceCreationInfo

definition	Groups together information on the creation of a tool or service
type	component
elements	<p>implementationLanguage</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>formalism</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>originalSource</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>creationDetails</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p>

8.2.8.1 implementationLanguage

definition	The programming languages needed for allowing user contributions, or for running the tools, in case no executables are available
type	xs:string
value space	
values	
examples	
DCLINK	
ISocatLINK	

8.2.8.2 formalism

definition	Reference (name, bibliographic reference or link to url) for the formalism used for the creation/enrichment of the resource (grammar or tool/service)
type	ms:myStringURI
value space	
values	
examples	FUG; LBNF; ATN; GATE
DCLINK	
ISocatLINK	

8.2.8.3 creationDetails

definition	Provides additional information on the creation of a tool or service
type	xs:string
value space	
values	
examples	
DCLINK	
ISocatLINK	

9 Language Descriptions

9.1 resourceInfo

definition	The profile for Language Descriptions (e.g. grammars), i.e. the resourceInfo component as intended for them
------------	---

elements	<p>resourceInfo – common components</p> <p>identificationInfo</p> <p>contactPerson</p> <p>versionInfo</p> <p>distributionInfo</p> <p>validationInfo</p> <p>resourceCreationInfo</p> <p>usageInfo</p> <p>metadataInfo</p> <p>resourceDocumentationInfo</p> <p>relationInfo</p> <p>resourceComponentType</p> <p>languageDescriptionInfo</p>
----------	---

9.2 languageDescriptionInfo

definition	Groups together information on language descriptions (grammars)
type	component
elements	<p>resourceType</p> <p><i>Status: Mandatory with value "languageDescription"</i></p> <p><i>Repeatability: 1</i></p> <p>languageDescriptionType</p> <p><i>Status: Mandatory</i></p> <p><i>Repeatability: 1</i></p> <p>languageDescriptionEncodingInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p> <p>languageDescriptionOperationInfo</p> <p><i>Status: Optional</i></p> <p><i>Repeatability: 1</i></p> <p>languageDescriptionPerformanceInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p> <p>creationInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p> <p>languageDescriptionMediaType</p>

	<i>Status: mandatory</i> <i>Repeatability: 1</i>
--	---

9.2.1 languageDescriptionType

definition	The type of the language description		
type	open controlled vocabulary		
value space	MS-languageDescriptionType		
values	value	definition	ISOcatLINK
	grammar	Value for languageDescriptionType; to be used for resources describing the morphological, syntactic etc. system (or part thereof) of a language	http://www.isocat.org/datcat/DC-4363
	other		
examples			
DCLINK			
ISOcatLINK	~ http://www.isocat.org/datcat/DC-3806 (ISOcat concept is more general)		

9.2.2 languageDescriptionEncodingInfo

definition	Groups together information on the contents of the LanguageDescriptions
type	component
elements	encodingLevel <i>Status: mandatory</i> <i>Repeatability: unbounded</i> conformanceToStandardsBestPractices <i>Status: recommended</i> <i>Repeatability: unbounded</i> theoreticModel <i>Status: recommended</i> <i>Repeatability: unbounded</i> formalism <i>Status: optional</i>

	<i>Repeatability: 1</i> task <i>Status: recommended</i> <i>Repeatability: unbounded</i> grammaticalPhenomenaCoverage <i>Status: recommended</i> <i>Repeatability: unbounded</i> weightedGrammar <i>Status: optional</i> <i>Repeatability: 1</i>
--	--

9.2.2.1 encodingLevel

definition	Information on the linguistic levels covered by the resource (grammar or lexical/conceptual resource)		
type	open controlled vocabulary		
value space	MS-encodingLevel		
values	value	definition	ISOcatLINK
	phonetics		
	phonology		
	semantics		
	morphology		
	syntax		http://www.isocat.org/datcat/DC-2647
	pragmatics		http://www.isocat.org/datcat/DC-2643
	other		
examples			
DCLINK			
ISOcatLINK			

9.2.2.2 task

definition	An indication of the task performed by the grammar				
type	open controlled vocabulary				
value space	MS-task				
values	<table> <tr> <th>value</th><th>definition</th></tr> <tr> <td>anaphoraResolution</td><td>Value for task; for grammars used for anaphora resolution</td></tr> </table>	value	definition	anaphoraResolution	Value for task; for grammars used for anaphora resolution
value	definition				
anaphoraResolution	Value for task; for grammars used for anaphora resolution				

	chunking	Value for task; for grammars used for chunking (shallow syntactic parsing)
	parsing	Value for task; for grammars used for parsing
	npRecognition	Value for task; for grammars used for identifying noun phrases
	titlesParsing	Value for task; for grammars that are restricted to parsing titles
	definitionsParsing	Value for task; for grammars that are restricted to parsing definitions
	analysis	Value for task; for grammars used for analysis tasks
	generation	Value for task; for grammars used for language generation
	other	
examples		
DCLINK		
ISOcatLINK	~ http://www.isocat.org/datcat/DC-2500 (more general)	

9.2.2.3 grammaticalPhenomenaCoverage

definition	An indication of the grammatical phenomena covered by the grammar	
type	open controlled vocabulary	
value space	MS-grammaticalPhenomenaCoverage	
values	Value	definition
	clauseStructure	Value for grammaticalPheonomenaCoverage; for grammars that can recognise the structure of clauses
	ppAttachment	Value for grammaticalPheonomenaCoverage; for grammars that cover pp-attachment (attachment of prepositional phrases)
	npStructure	value for grammaticalPheonomenaCoverage; for grammars that cover structure of noun phrases
	coordination	value for grammaticalPheonomenaCoverage; for grammars that cover coordination
	anaphora	value for grammaticalPheonomenaCoverage; for grammars that cover anaphora problems
	other	
examples		
DCLINK		

ISocatLINK	
------------	--

9.2.2.4 weightedGrammar

definition	Indicates whether the grammar contains numerical weights (incl. probabilities)
type	xs:boolean
value space	
values	
examples	
DCLINK	
ISocatLINK	

9.2.3 languageDescriptionOperationInfo

definition	Groups together information on the operation requirements of the LanguageDescriptions
type	component
elements	runningEnvironmentInfo <i>Status: recommended</i> <i>Repeatability: 1</i> relatedLexiconInfo <i>Status: recommended</i> <i>Repeatability: 1</i>

9.2.3.1 relatedLexiconInfo

definition	Groups together information on requirements for lexica set by the LanguageDescriptions
type	component
elements	relatedLexiconType <i>Status: mandatory</i> <i>Repeatability: 1</i> attachedLexiconPosition <i>Status: optional</i> <i>Repeatability: 1</i> compatibleLexiconType

	<i>Status: optional</i> <i>Repeatability: unbounded</i>
--	--

9.2.3.1.1 relatedLexiconType

definition	Indicates the 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	For lexica that are included in the grammar file
	attached	For lexica attached to a resource (e.g. as a different file in the same directory, or downloadable from another position); the element attachedLexiconPosition should be used for further information
	compatible	For grammars that can use any lexicon as long as they are of a compatible type; the element compatibleLexiconType must be used for further specification
	none	
examples		
DCLINK		
ISOcatLINK		

9.2.3.1.2 attachedLexiconPosition

definition	Indicates the position of the lexicon, if attached to the grammar
type	ms:myStringURI
value space	
values	
examples	
DCLINK	
ISOcatLINK	

9.2.3.1.3 compatibleLexiconType

definition	Type of (external) lexicon that can be used with the grammar	
type	open controlled vocabulary	
value space	MS-compatibleLexiconType	
value	value	definition

	wordnet	Value for compatibleLexiconType; to be used for grammars that are compatible with wordnets
	wordlist	Value for compatibleLexiconType; to be used for grammars that need wordlists for their operation
	morphologicalLexicon	Value for compatibleLexiconType; to be used for grammars that need lexica with morphological information for their operation
	other	
examples		
DCLINK		
ISocatLINK		

9.2.4 languageDescriptionPerformanceInfo

definition	Groups together information on the performance of the LanguageDescriptions
type	component
elements	robustness <i>Status: recommended</i> <i>Repeatability: 1</i> shallowness <i>Status: recommended</i> <i>Repeatability: 1</i> output <i>Status: recommended</i> <i>Repeatability: 1</i>

9.2.4.1 robustness

definition	Free text statement on the robustness of the grammar (how well the grammar can cope with misspelt/unknown etc. input, i.e. whether it can produce even partial interpretations of the input)
type	xs:string
value space	
values	
examples	
DCLINK	
ISocatLINK	

9.2.4.2 shallowness

definition	Free text statement on the shallowness of the grammar (how deep the syntactic analysis performed by the grammar can be)
type	xs:string
value space	
values	
examples	
DCLINK	
ISocatLINK	

9.2.4.3 output

definition	Indicates whether the output of the operation of the grammar is a statement of grammaticality (grammatical/ungrammatical) or structures (interpretation of the input)
type	xs:string
value space	
values	
examples	grammaticality statement; tree structures; chunks
DCLINK	
ISocatLINK	

9.2.5 languageDescriptionResourceMediaType

definition	Groups information on the media type-specific components for language descriptions
type	component
elements	languageDescriptionTextInfo <i>Status: recommended</i> <i>Repeatability: 1</i> languageDescriptionVideoInfo <i>Status: optional</i> <i>Repeatability: 1</i> languageDescriptionImageInfo <i>Status: optional</i>

	<i>Repeatability: 1</i>
--	-------------------------

9.2.5.1 languageDescriptionTextInfo

definition	Groups together all information relevant to the text module of a language description (e.g. format, languages, size etc.); it is obligatory for all language descriptions
type	component
elements	<p>mediaType</p> <p><i>Status: mandatory, with value "text"</i></p> <p><i>Repeatability: 1</i></p> <p>lingualityInfo</p> <p><i>Status: mandatory</i></p> <p><i>Repeatability: 1</i></p> <p>languageInfo</p> <p><i>Status: mandatory</i></p> <p><i>Repeatability: unbounded</i></p> <p>modalityInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p> <p>sizeInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>textFormatInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>characterEncodingInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>domainInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>timeCoverageInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>geographicCoverageInfo</p> <p><i>Status: recommended</i></p>

	<i>Repeatability: unbounded</i> creationInfo <i>Status: recommended</i> <i>Repeatability: 1</i> linkToOtherMediaInfo <i>Status: optional</i> <i>Repeatability: unbounded</i>
--	--

9.2.5.2 languageDescriptionVideoInfo

definition	Groups together all information relevant to the video parts of a language description (e.g. format, languages, size etc.), if there are any (e.g. for sign language grammars)
type	component
elements	mediaType <i>Status: mandatory, with value "video"</i> <i>Repeatability: 1</i> lingualityInfo <i>Status: optional</i> <i>Repeatability: 1</i> languageInfo <i>Status: optional</i> <i>Repeatability: unbounded</i> modalityInfo <i>Status: recommended</i> <i>Repeatability: 1</i> sizeInfo <i>Status: recommended</i> <i>Repeatability: 1</i> videoContentInfo <i>Status: recommended</i> <i>Repeatability: 1</i> videoFormatInfo <i>Status: recommended</i> <i>Repeatability: unbounded</i> domainInfo <i>Status: recommended</i>

	<p><i>Repeatability: unbounded</i></p> <p>timeCoverageInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>geographicCoverageInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>creationInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p> <p>linkToOtherMediaInfo</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p>
--	---

9.2.5.3 languageDescriptionImageInfo

definition	Groups together all information relevant to the image module of a language description (e.g. format, languages, size etc.), if there are any (e.g. for sign language grammars)
type	component
elements	<p>mediaType</p> <p><i>Status: mandatory with value “image”</i></p> <p><i>Repeatability: 1</i></p> <p>lingualityInfo</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p>languageInfo</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>modalityInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p> <p>sizeInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p> <p>imageContentInfo</p> <p><i>Status: recommended</i></p>

	<p><i>Repeatability: 1</i></p> <p><u>imageFormatInfo</u></p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p><u>domainInfo</u></p> <p><i>Status: Optional</i></p> <p><i>Repeatability: unbounded</i></p> <p><u>timeCoverageInfo</u></p> <p><i>Status: Optional</i></p> <p><i>Repeatability: unbounded</i></p> <p><u>geographicCoverageInfo</u></p> <p><i>Status: Optional</i></p> <p><i>Repeatability: unbounded</i></p> <p><u>creationInfo</u></p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p> <p><u>linkToOtherMediaInfo</u></p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p>
--	---

10 Lexical/conceptual resources

10.1 resourceInfo

definition	Extension of resourceInfo for lexical/conceptual resources
elements	<p>resourceInfo – common components</p> <p><u>identificationInfo</u></p> <p><u>distributionInfo</u></p> <p><u>contactPerson</u></p> <p><u>metadataInfo</u></p> <p><u>versionInfo</u></p> <p><u>validationInfo</u></p> <p><u>usageInfo</u></p> <p><u>resourceDocumentationInfo</u></p> <p><u>resourceCreationInfo</u></p> <p><u>relationInfo</u></p> <p><u>resourceComponentType</u></p>

	lexicalConceptualResourceInfo
--	---

10.2 lexicalConceptualResourceInfo

definition	Groups together information specific to lexical/conceptual resources
type	component
elements	<p>resourceType</p> <p><i>Status: mandatory with value "lexicalConceptualResource"</i></p> <p><i>Repeatability: 1</i></p> <p>lexicalConceptualResourceType</p> <p><i>Status: mandatory</i></p> <p><i>Repeatability: 1</i></p> <p>lexicalConceptualResourceEncodingInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p> <p>creationInfo</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p>lexicalConceptualResourceMediaType</p> <p><i>Status: mandatory</i></p> <p><i>Repeatability: unbounded</i></p>

10.2.1 lexicalConceptualResourceType

definition	Specifies the subtype of lexicalConceptualResource																				
type	open controlled vocabulary																				
value space	MS-lexicalConceptualResourceType																				
values	<table> <tr> <td>value</td><td>definition</td></tr> <tr> <td>wordList</td><td></td></tr> <tr> <td>computationalLexicon</td><td></td></tr> <tr> <td>ontology</td><td></td></tr> <tr> <td>wordnet</td><td></td></tr> <tr> <td>thesaurus</td><td></td></tr> <tr> <td>framenet</td><td></td></tr> <tr> <td>terminologicalResource</td><td></td></tr> <tr> <td>machineReadableDictionary</td><td></td></tr> <tr> <td>lexicon</td><td></td></tr> </table>	value	definition	wordList		computationalLexicon		ontology		wordnet		thesaurus		framenet		terminologicalResource		machineReadableDictionary		lexicon	
value	definition																				
wordList																					
computationalLexicon																					
ontology																					
wordnet																					
thesaurus																					
framenet																					
terminologicalResource																					
machineReadableDictionary																					
lexicon																					

	other
examples	
DCLINK	
ISocatLINK	http://www.isocat.org/datcat/DC-2487

10.2.2 lexicalConceptualResourceEncodingInfo

definition	Groups all information regarding the contents of lexical/conceptual resources
type	component
elements	<p>encodingLevel</p> <p><i>Status: Mandatory</i></p> <p><i>Repeatability: unbounded</i></p> <p>linguisticInformation</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>conformanceToStandardsBestPractices</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>theoreticModel</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>externalRef</p> <p><i>Status: Optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>extratextualInformation</p> <p><i>Status: Optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>extratextualInformationUnit</p> <p><i>Status: Optional</i></p> <p><i>Repeatability: unbounded</i></p>

10.2.2.1 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 morphology syntax pragmatics other
examples	
DCLINK	
ISocatLINK	

10.2.2.2 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 semantics-QualiaStructure syntacticoSemanticLinks other
examples	
DCLINK	
ISocatLINK	

10.2.2.3 externalRef

definition	Another resource to which the lexicalConceptualResource is linked (e.g. link
------------	--

	to a wordnet or ontology)
type	xs:string
value space	
values	
examples	
DCLINK	
ISOcatLINK	

10.2.2.4 extratextualInformation

definition	An indication of the extratextual information contained in the lexicalConceptualResource; 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	<div>value</div> <div>definition</div> <div>images</div> <div>videos</div> <div>soundRecordings</div> <div>other</div>
examples	
DCLINK	
ISOcatLINK	

10.2.2.5 extratextualInformationUnit

definition	The unit of the extratextual information contained in the lexical conceptual resource
type	open controlled vocabulary
value space	MS-extratextualInformation
values	<div>value</div> <div>definition</div> <div>word</div> <div>lemma</div> <div>semantics</div> <div>example</div> <div>syntax</div>

	lexicalUnit other
examples	
DCLINK	
ISocatLINK	

10.2.3 lexicalConceptualResourceMediaType

definition	Restriction of mediaType for lexicalConceptualResources
type	component
elements	lexicalConceptualResourceTextInfo <i>Status: recommended</i> <i>Repeatability: 1</i> lexicalConceptualResourceAudioInfo <i>Status: optional</i> <i>Repeatability: 1</i> lexicalConceptualResourceVideoInfo <i>Status: optional</i> <i>Repeatability: 1</i> lexicalConceptualResourceImageInfo <i>Status: optional</i> <i>Repeatability: 1</i>

10.2.3.1 lexicalConceptualResourceTextInfo

definition	Groups information on the textual part of the lexical/conceptual resource
type	component
elements	mediaType <i>Status: Mandatory with value "text"</i> <i>Repeatability: 1</i> lingualityInfo <i>Status: Mandatory</i> <i>Repeatability: 1</i> languageInfo <i>Status: Mandatory</i> <i>Repeatability: unbounded</i> modalityInfo

	<p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p> <p>sizeInfo</p> <p><i>Status: Mandatory</i></p> <p><i>Repeatability: unbounded</i></p> <p>textFormatInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>characterEncodingInfo</p> <p><i>Status: Optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>domainInfo</p> <p><i>Status: Optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>timeCoverageInfo</p> <p><i>Status: Optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>geographicCoverageInfo</p> <p><i>Status: Optional</i></p> <p><i>Repeatability: unbounded</i></p>
--	---

10.2.3.2 lexicalConceptualResourceAudioInfo

definition	Groups information on the audio part of the lexical/conceptual resource
type	component
elements	<p>mediaType</p> <p><i>Status: Mandatory with value "audio"</i></p> <p><i>Repeatability: 1</i></p> <p>lingualityInfo</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p>languageInfo</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>modalityInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p>

	sizeInfo <i>Status: recommended</i> <i>Repeatability: unbounded</i> audioContentInfo <i>Status: recommended</i> <i>Repeatability: 1</i> audioFormatInfo <i>Status: recommended</i> <i>Repeatability: unbounded</i> domainInfo <i>Status: Optional</i> <i>Repeatability: unbounded</i> timeCoverageInfo <i>Status: Optional</i> <i>Repeatability: unbounded</i> geographicCoverageInfo <i>Status: Optional</i> <i>Repeatability: unbounded</i>
--	--

10.2.3.3 lexicalConceptualResourceVideoInfo

definition	Groups information on the video part of the lexical conceptual resource
type	component
elements	mediaType <i>Status: Mandatory with value "video"</i> <i>Repeatability: 1</i> lingualityInfo <i>Status: optional</i> <i>Repeatability: 1</i> languageInfo <i>Status: optional</i> <i>Repeatability: unbounded</i> modalityInfo <i>Status: recommended</i> <i>Repeatability: 1</i> sizeInfo <i>Status: recommended</i>

	<p><i>Repeatability: unbounded</i></p> <p>videoContentInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p> <p>videoFormatInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>domainInfo</p> <p><i>Status: Optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>timeCoverageInfo</p> <p><i>Status: Optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>geographicCoverageInfo</p> <p><i>Status: Optional</i></p> <p><i>Repeatability: unbounded</i></p>
--	---

10.2.3.4 lexicalConceptualResourceImageInfo

definition	Groups information on the image part of the lexical/conceptual resource
type	component
elements	<p>mediaType</p> <p><i>Status: Mandatory with value "image"</i></p> <p><i>Repeatability: 1</i></p> <p>lingualityInfo</p> <p><i>Status: optional</i></p> <p><i>Repeatability: 1</i></p> <p>languageInfo</p> <p><i>Status: optional</i></p> <p><i>Repeatability: unbounded</i></p> <p>modalityInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: 1</i></p> <p>sizeInfo</p> <p><i>Status: recommended</i></p> <p><i>Repeatability: unbounded</i></p> <p>imageContentInfo</p>

	<i>Status: recommended</i> <i>Repeatability: 1</i> <u>imageFormatInfo</u> <i>Status: recommended</i> <i>Repeatability: unbounded</i> <u>domainInfo</u> <i>Status: optional</i> <i>Repeatability: unbounded</i> <u>timeCoverageInfo</u> <i>Status: optional</i> <i>Repeatability: unbounded</i> <u>geographicCoverageInfo</u> <i>Status: optional</i> <i>Repeatability: unbounded</i>
--	---

AppendixI- 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 Resources and Technology”, *Proceedings of the 6th International Conference of Language Resources and Evaluation*, 2008.
- Federmann, C., B. Georgantopoulos, R. delGratta, 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.
- Labropoulou, P, M. Gavrilidou, E. Desipri, S. Piperidis, M. Speranza, M. Monachini, V. Arranz, G. Francopoulo, *META-NET Deliverable D7.2.2– Specification of Metadata-Based Descriptions for Language Resources and Technologies*, 2012.