

Overview of standards-related resources in CLARIN centres

The following table gives an overview of standards-related information that is available in CLARIN centres. We have used the standards database located at <http://clarin.ids-mannheim.de/standards/> as basis for assembling this tabular overview.

Name	Type of resource	URL for direct information ¹	Further information
CES	technical	http://clarin.ids-mannheim.de/standards/	
Citer			Mentioned in CLARIN Standardisation Action Plan and LRT-6 (as 'Handles')
CHAT			Mentioned in LRT-6
CMDI	technical	http://clarin.ids-mannheim.de/standards/	
Controlled Vocabulary	technical	http://clarin.ids-mannheim.de/standards/	
Country Codes			Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and LRT-6
CLARIN MDI			Mentioned in LRT-6
CQLF	technical	http://clarin.ids-mannheim.de/standards/	
CSV			Mentioned in CLARIN EU Standard FAQ and LRT-6
DCAM	technical	http://clarin.ids-mannheim.de/standards/	
DCMES	technical	http://clarin.ids-mannheim.de/standards/	Mentioned in LRT-6
DCR	technical	http://clarin.ids-mannheim.de/standards/	Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and LRT-6 and D5C-3
DiAML	technical	http://clarin.ids-mannheim.de/standards/	Mentioned in D5C-3
DictionaryEntry-RePresentation	technical	http://clarin.ids-mannheim.de/standards/	
DITA	technical	http://clarin.ids-mannheim.de/standards/	
DOL	technical	http://clarin.ids-mannheim.de/standards/	
DSSSL	technical	http://clarin.ids-mannheim.de/standards/	
EAD			Mentioned in LRT-6

¹ For items with the URL <http://clarin.ids-mannheim.de/standards/> a technical description is available at the IDS standards database.

EAF			Mentioned in LRT-6
Feature structures	technical	http://clarin.ids-mannheim.de/standards/	Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and D5C-3
GOLD	technical	http://clarin.ids-mannheim.de/standards/	Mentioned in LRT-6
H264			Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and LRT-6
HTML	technical	http://clarin.ids-mannheim.de/standards/	Mentioned in LRT-6
HyTime	technical	http://clarin.ids-mannheim.de/standards/	
IMDI	technical	http://clarin.ids-mannheim.de/standards/	Mentioned in LRT-6
IPA			Mentioned in CLARIN Standardisation Action Plan
ISBD	technical	http://clarin.ids-mannheim.de/standards/	
ISO-Thesauri	technical	http://clarin.ids-mannheim.de/standards/	
ITS	technical	http://clarin.ids-mannheim.de/standards/	
JATS	technical	http://clarin.ids-mannheim.de/standards/	
JPEG			Mentioned in LRT-6
LAF	technical	http://clarin.ids-mannheim.de/standards/	Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and D5C-3
KAF			Mentioned in D5C-3
Language Codes			Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and LRT-6 and D5C-3
LMF	technical	http://clarin.ids-mannheim.de/standards/	Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and LRT-6 and D5C-3
MAF	technical	http://clarin.ids-mannheim.de/standards/	Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and D5C-3
MARC			Mentioned in LRT-6
MARTIF			Mentioned in D5C-3
METS	technical	http://clarin.ids-mannheim.de/standards/	Mentioned in LRT-6
MJPEG2000			Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and LRT-6
MLIF	technical	http://clarin.ids-mannheim.de/standards/	Mentioned in D5C-3
MPEG21 DID			Mentioned in LRT-6

MPEG7			Mentioned in LRT-6
MPEG-2			Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and LRT-6
MPEG-1 Layer 3/MP3			Mentioned in LRT-6
Multilingual Thesaurus	technical	http://clarin.ids-mannheim.de/standards/	
NISO MIX	technical	http://clarin.ids-mannheim.de/standards/	
nRQL			Mentioned in D5C-3
OAI PMH			Mentioned in LRT-6
OLAC Metadata	technical	http://clarin.ids-mannheim.de/standards/	Mentioned in LRT-6
OLiA	technical	http://clarin.ids-mannheim.de/standards/	
OntoIop	technical	http://clarin.ids-mannheim.de/standards/	
Ontology Mapping Lang			Mentioned in D5C-3
ORE			Mentioned in LRT-6
OWL	technical	http://clarin.ids-mannheim.de/standards/	Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and LRT-6
PCM			Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and LRT-6
PDF/A			Mentioned in LRT-6
PISA	technical	http://clarin.ids-mannheim.de/standards/	
PNG			Mentioned in LRT-6
PML			Mentioned in D5C-3
RDF	technical	http://clarin.ids-mannheim.de/standards/	Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and LRT-6
RDF/XML	technical	http://clarin.ids-mannheim.de/standards/	Mentioned in CLARIN EU Standard FAQ
RDFS	technical	http://clarin.ids-mannheim.de/standards/	Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and LRT-6
RELAX NG	technical	http://clarin.ids-mannheim.de/standards/	Mentioned in LRT-6
Representation of names and scripts			Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and LRT-6
Representation of entries in dictionaries			Mentioned in LRT-6 and D5C-3
RDQL			Mentioned in D5C-3
REST			Mentioned in LRT-6
RQL			Mentioned in D5C-3

RTF			Mentioned in LRT-6
SeRQL			Mentioned in D5C-3
SGML	technical	http://clarin.ids-mannheim.de/standards/	
SKOS	technical	http://clarin.ids-mannheim.de/standards/	Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and LRT-6 and D5C-3
SemAF	technical	http://clarin.ids-mannheim.de/standards/	Mentioned in D5C-3
SemRoleML	technical	http://clarin.ids-mannheim.de/standards/	
SimpL-1	technical	http://clarin.ids-mannheim.de/standards/	
SOAP			Mentioned in LRT-6
SPARQL	technical	http://clarin.ids-mannheim.de/standards/	Mentioned in D5C-3
SRX	technical	http://clarin.ids-mannheim.de/standards/	
Structured Vocabulary	technical	http://clarin.ids-mannheim.de/standards/	
SynAF	technical	http://clarin.ids-mannheim.de/standards/	Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and D5C-3
Systems to manage terminology, knowledge and content			Mentioned in D5C-3
TermLR			Mentioned in D5C-3
TBX	technical	http://clarin.ids-mannheim.de/standards/	Mentioned in LRT-6 and D5C-3
TEI Guidelines	technical	http://clarin.ids-mannheim.de/standards/	Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and LRT-6 and D5C-3
TextMD	technical	http://clarin.ids-mannheim.de/standards/	
TIFF			Mentioned in LRT-6
TimeML	technical	http://clarin.ids-mannheim.de/standards/	Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and D5C-3
Tipster			Mentioned in LRT-6
TMF	technical	http://clarin.ids-mannheim.de/standards/	Mentioned in LRT-6 and D5C-3
TMS	technical	http://clarin.ids-mannheim.de/standards/	
TMX	technical	http://clarin.ids-mannheim.de/standards/	Mentioned in LRT-6 and D5C-3
Topic Maps	technical	http://clarin.ids-mannheim.de/standards/	
TRIPLE			Mentioned in D5C-3
Turtle	technical	http://clarin.ids-mannheim.de/standards/	

Unicode			Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and D5C-3
URI/URN			Mentioned in LRT-6
WordNet			Mentioned in D5C-3
WordSeg	technical	http://clarin.ids-mannheim.de/standards/	Mentioned in D5C-3
WSDL			Mentioned in LRT-6
XCES	technical	http://clarin.ids-mannheim.de/standards/	Mentioned in LRT-6 and D5C-3
XHTML	technical	http://clarin.ids-mannheim.de/standards/	
XML	technical	http://clarin.ids-mannheim.de/standards/	Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and LRT-6 and D5C-3
XMLNS	technical	http://clarin.ids-mannheim.de/standards/	
XPath	technical	http://clarin.ids-mannheim.de/standards/	
XQuery	technical	http://clarin.ids-mannheim.de/standards/	
XSD	technical	http://clarin.ids-mannheim.de/standards/	Mentioned in LRT-6
XSL-FO	technical	http://clarin.ids-mannheim.de/standards/	
XSLT	technical	http://clarin.ids-mannheim.de/standards/	

Further information:

The “CLARIN EU FAQ on Standards” located at <http://www.clarin.eu/faq/what-standards-are-recommended-clarin> contains an open list naming the following standards (already included in the list above):

- character encoding: ISO -10646 UNICODE, UTF-8
- country codes: ISO 3166
- language codes: ISO 639-1 and 639-3
- codes for the representation of names of scripts: ISO 15924
- text format: XML
- text format: CSV (comma separated with "-quotes, with a header line and preferably a line of ISOcat URIs for each column)
- feature structure representation: ISO 24610-1:2006
- representation of primary sources: TEI (Text Encoding Initiative)
- knowledge engineering: RDF, RDF-S, SKOS, OWL

- audio/speech: PCM (Pulse Code Modulation) for digitizing sound waves, the Alphabet of the International Phonetic Association for phonetic transcriptions;
- video/multimodality: MJPEG2000 lossless as backend format, MPEG2 or H.264 for handling and processing
- data categories: ISO DCR and ISOcat
- annotation of temporal entities: TimeML (part of TC 37/SC 4)
- morpho-syntactic annotation: MAF (Morpho-syntactic Annotation Framework), ISO/DIS 24611
- syntactic annotation: SynAF (Syntactic Annotation Framework), ISO/CD 24615
- lexical annotation: LMF (Lexical Markup Framework), ISO 24613:2008
- linguistic annotation: LAF (Linguistic Annotation Framework), ISO/DIS 24612

Basic information about most of these standards can be found in version 8 of the “*CLARIN Standardisation Action Plan*”². In addition it contains a reference to ISO/CD 24619 Persistent identification and access in language technology application (Citer)

The document “*Standards for LRT*” (version 6)³ contains a number of standards, formats and protocols (indicated in the table above) including information about the status. Even specifications that have a negative comment in this document are included in the table above for discussion purposes.

The short guide “*Standards and Best Practices*”⁴ contains the section „How does it work?“ with a list of standards (all standards mentioned here are already included in the table above) but does not contain any further information.

The deliverable D5C-3 “*Interoperability and Standards*”⁵ contains an inventory of specifications. Standards mentioned in here are thoroughly described, however, a large number of specs are not CLARIN-related.

The FLareNet document “*The Standards’ Landscape Towards an Interoperability Framework*”⁶ contains short information items about a large number of the specifications.

² <http://www.clarin.eu/file/1511>

³ <http://www.clarin.eu/sites/default/files/Standards%20for%20LRT-v6.pdf>

⁴ <http://www.clarin.eu/sites/default/files/standards-CLARIN-ShortGuide.pdf>

⁵ <http://hdl.handle.net/1839/00-DOCS.CLARIN.EU-51>

⁶ http://www.flarenet.eu/sites/default/files/FLareNet_Standards_Landscape.pdf

Open questions and next steps

1. Which of these standards are worthwhile to be described further in the future version of the standards database (that is, which of these is used in CLARIN? Which can be discarded? Do we only concentrate on standards or do we include formats and protocols as well?
2. Who has further information about the specifications featured on the final list of standards? Who is responsible for gathering information and creating the standards list entry? Some information is already present in the IDS standards database, sometimes additional information is available in the PDF documents mentioned above.
3. When the final list is assembled, do we recommend every standard used in CLARIN or only those that have a given level of support?