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.

This is an edited version of the standards-related resources overview. For example, the type of resource entries have been fine tuned and more informative hyperlinks are added.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Name | Type of resource | IDS Standards DB[[1]](#footnote-1) | Further information | Used in CLARIN centre | TLA Comments |
| CES (Corpus Encoding Standard) | Data format | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecCes> |  |  |  |
| Citer |  |  | Mentioned in CLARIN Standardisation Action Plan and LRT-6 (as ‘Handles’) |  | Is now known as PISA (see PISA) |
| CHAT (Codes for the Human Analysis of Transcripts) | Data format |  | <http://childes.psy.cmu.edu/manuals/chat.pdf> ; Mentioned in LRT-6 | TLA |  |
| CMDI (Component Metadata Infrastructure) | Metadata format  Infrastructure | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecCmdi> |  | TLA, CLARIN-DK, CLARIN-PL |  |
| Controlled Vocabulary | Guideline | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecCV> |  | TLA, CLARIN-PL | Too generic, specific controlled vocabulary standards, e.g., CLAVAS, should be mentioned |
| Country Codes |  |  | Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and LRT-6 | TLA (in CMDI, not in IMDI), CLARIN-DK, CLARIN-PL | Too generic, specific country code sets, e.g., ISO 3166, should be mentioned |
| CLARIN MDI |  |  | Mentioned in LRT-6 |  | Is now known as CMDI (see CMDI) |
| CQLF (Corpus Query Lingua Franca) | Guideline  Query language | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecCQLF> |  |  | Under development |
| CQP (Corpus Query Processor) | Tool  Query language |  | <http://cwb.sourceforge.net/files/CQP_Tutorial/> |  |  |
| CSV (Comma Separated Values) | Data format |  | <http://en.wikipedia.org/wiki/Comma-separated_values> ; Mentioned in CLARIN EU Standard FAQ and LRT-6 | TLA (<10 resources), CLARIN-DK, CLARIN-PL | There is no CSV standard. Implementations differ in escaping, field names in the first row, single or double quotes, … |
| DataCITE | Metadata schema |  | <http://schema.datacite.org/> |  |  |
| DCAM (Dublin Core Abstract Model) | Metadata schema | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecDcAM> |  |  |  |
| DCMES (Dublin Core Metadata Element Set) | Metadata schema | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecDcMes> | Mentioned in LRT-6 | TLA (IMDIfied DC) |  |
| DCR (Data Category Registry) | Knowledge resource | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecDcr> | Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and LRT-6 and D5C-3 | TLA, CLARIN-PL | Two successors of ISOcat are (most likely) upcoming: a new ISO DCR (using isocat.org) and a CLARIN Concept Registry |
| DiAML (Dialogue Act Markup Language) | Data format | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecDiAML> | Mentioned in D5C-3 |  |  |
| DictionaryEntry-RePresentation | Data format | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecDict> |  |  | What is the relationship to LMF? |
| DITA (Darwin Information Typing Architecture) | Data format | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecDita> |  |  |  |
| DOL (Distributed Ontology Language) | Knowledge representation | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecDOL> |  |  | See OntoIOp |
| DSSSL (Document Style Semantics and Specification Language) | Transformation language | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecDSSSL> |  |  |  |
| EAD (Encoded Archival Description) | Metadata schema |  | <http://www.loc.gov/ead/> ; Mentioned in LRT-6 |  |  |
| EAF (ELAN (or EUDICO) Annotation Format) | Data format |  | <http://www.mpi.nl/tools/elan/EAF_Annotation_Format.pdf> ; Mentioned in LRT-6 | TLA |  |
| Feature structures | Data format | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecFS> | Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and D5C-3 | CLARIN-PL | Should be more specific as there are several FSRs |
| GOLD (General Ontology for Linguistic Description) | Knowledge resource | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecGOLD> | Mentioned in LRT-6 |  |  |
| H264 | Data format or encoding |  | <http://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-H.264-200305-S!!PDF-E&type=items> ; Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and LRT-6 | TLA, CLARIN-PL | H264 is an encoding, should the recommendation also include a container, e.g., H.264/MPEG-4 AVC? |
| Handle | Infrastructure |  |  | TLA, CLARIN-PL |  |
| HTML (Hypertext Markup Language) | Data format | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecHTML> | Mentioned in LRT-6 | TLA, CLARIN-PL | Be more specific in HTML versions |
| HTTP (Hypertext Transfer Protocol) | Infrastructure |  |  | TLA, CLARIN-PL |  |
| HyTime | Data format | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecHyTime> |  |  |  |
| IMDI (ISLE Metadata Initiative) | Metadata schema | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecImdiSd> | Mentioned in LRT-6 | TLA, CLARIN-DK, CLARIN-PL |  |
| IPA (International Phonetic Alphabet ) | Data format? |  | <https://www.langsci.ucl.ac.uk/ipa/ipachart.html> ; Mentioned in CLARIN Standardisation Action Plan | TLA (in EAF) | Is there a standard (Unicode) representation of IPA? |
| ISBD (International Standard Bibliographic Description) | Metadata schema  Guideline | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecISBD> |  |  | Gives a structure but not really a representation (?) |
| ISO-Thesauri | Knowledge representation | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecThesauri> |  |  |  |
| ITS (Internationalization Tag Set) | Data format | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecITS> |  |  |  |
| JATS (Journal Article Tag Suite) | Data format | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecNLMJATS> |  |  |  |
| JPEG (Joint Photographic Experts Group) | Data format |  | <http://en.wikipedia.org/wiki/JPEG> ; Mentioned in LRT-6 | TLA, CLARIN-DK, CLARIN-PL |  |
| JSON (JavaScript Object Notation) | Data format |  | <http://json.org/> (see also <https://www.tbray.org/ongoing/When/201x/2014/03/05/RFC7159-JSON>) | CLARIN-PL | Could be considered too general, i.e., just like plain XML |
| LAF (Linguistic Annotation Framework) | Data format | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecLaf> | Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and D5C-3 | CLARIN-PL |  |
| KAF (Kyoto (or Knowledge) Annotation Framework (or Format)) | Data format |  | <http://weblab.iit.cnr.it/kyoto/xmlgroup.iit.cnr.it/kyoto/indexdd46.html> ; Mentioned in D5C-3 |  |  |
| Language Codes |  |  | Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and LRT-6 and D5C-3 | TLA, CLARIN-DK | Too generic, specific language code sets should be mentioned, e.g., ISO 639-3 |
| LMF (lexical Markup Framework) | Data format | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecLMF> | Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and LRT-6 and D5C-3 | TLA, CLARIN-DK, CLARIN-PL |  |
| MAF (Morpho-syntactic Annotation Framework) | Data format | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecMaf> | Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and D5C-3 |  |  |
| MARC (Machine-readable Cataloging) | Metadata schema |  | <http://www.loc.gov/marc/> ; Mentioned in LRT-6 |  |  |
| MARTIF (Machine-readable Terminology Interchange Format) | Data format |  | <http://www.ttt.org/clsframe/negotiated.html> ; Mentioned in D5C-3 |  |  |
| METS (Metadata Encoding and Transmission Standard) | Metadata schema | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecMETS> | <http://www.loc.gov/standards/mets/> ; Mentioned in LRT-6 |  |  |
| MJPEG 2000 (Motion JPEG 2000) | Data format |  | <http://www.jpeg.org/jpeg2000/j2kpart3.html> ; Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and LRT-6 | TLA, CLARIN-PL |  |
| MLIF (Multilingual Information Framework) | Data format | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecMLIF> | Mentioned in D5C-3 | CLARIN-PL |  |
| MPEG21 DID (Digital Item Declaration) | Data format |  | <http://mpeg.chiariglione.org/standards/mpeg-21/digital-item-declaration> ; Mentioned in LRT-6 |  |  |
| MPEG7 | Data format |  | <http://mpeg.chiariglione.org/standards/mpeg-7/mpeg-7.htm> ; Mentioned in LRT-6 |  |  |
| MPEG-2 | Data format |  | <http://mpeg.chiariglione.org/standards/mpeg-2> ; Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and LRT-6 | TLA, CLARIN-PL |  |
| MPEG-1 Layer 3/MP3 | Data format |  | <http://mpeg.chiariglione.org/standards/mpeg-1/audio> ; Mentioned in LRT-6 | TLA, CLARIN-DK |  |
| Multilingual Thesaurus | Guideline | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecThesauri-IFLA> |  |  |  |
| NISO MIX (Metadata for Images in XML Schema) | Metadata schema | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecMIX> | <http://www.loc.gov/standards/mix/> |  |  |
| nRQL (new Racer Query Language) | Tool  Query language |  | <http://racer.sts.tuhh.de/Racer-2.0/users-guide-2-0.pdf#page=85> ; Mentioned in D5C-3 |  | Propriety? Recommend SPARQL |
| OAI-PMH (Open Archives Initiative - Protocol for Metadata Harvesting) | Infrastructure |  | <http://www.openarchives.org/pmh/> ; Mentioned in LRT-6 | TLA, CLARIN-PL |  |
| OLAC Metadata | Metadata schema | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecOlac> | <http://www.language-archives.org/> ; Mentioned in LRT-6 | TLA, CLARIN-DK, CLARIN-PL |  |
| OLiA | Knowledge resource | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecOLiA> |  |  |  |
| OntoIOp (Ontology Integration and Interoperability) | Knowledge representation | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecOntoIOp> | <http://ontoiop.org/> |  |  |
| Ontology Mapping Language | Knowledge representation |  | <https://www.deri.ie/fileadmin/documents/DERI-TR-2004-06-30.pdf> ; Mentioned in D5C-3 | CLARIN-PL |  |
| OAI-ORE (Open Archives Initiative - Object Reuse and Exchange) | Metadata schema  Infrastructure |  | <http://www.openarchives.org/ore/> ; Mentioned in LRT-6 |  |  |
| OWL (Web Ontology Language) | Knowledge representation | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecOWL> | Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and LRT-6 | CLARIN-DK |  |
| PCM (Pulse-code modulation) | Data format |  | <http://en.wikipedia.org/wiki/Pulse-code_modulation> ; Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and LRT-6 | TLA (in WAV) | PCM is an encoding used by various formats, e.g., WAV, shouldn’t these formats be recommended? |
| PDF/A (Portable Documentation Format) | Data format |  | <http://www.pdfa.org/> ; Mentioned in LRT-6 | TLA (PDF <= 1.6), CLARIN-PL |  |
| PISA (Persistent identification and sustainable access) | Infrastructure  Guideline | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecPisa> |  | TLA |  |
| PNG (Portable Network Graphics) | Data format |  | <http://www.iso.org/iso/catalogue_detail.htm?csnumber=29581> ; Mentioned in LRT-6 | TLA, CLARIN-PL |  |
| PML (Prague Markup Language) | Data format |  | <http://ufal.mff.cuni.cz/jazz/PML/doc/> ; Mentioned in D5C-3 |  |  |
| RDF (Resource Description Framework) | Knowledge representation  Data format | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecRDF> | Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and LRT-6 | CLARIN-DK, CLARIN-PL | See RDF/XML, RDFS |
| RDF/XML | Data format | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecRDFXML> | Mentioned in CLARIN EU Standard FAQ | CLARIN-PL | Should recommend one common RDF serialization |
| RDFS (RDF Schema) | Schema language | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecRDFS> | Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and LRT-6 |  |  |
| RELAX NG (REgular LAnguage for XML Next Generation) | Schema language | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecRNG> | Mentioned in LRT-6 | TLA, CLARIN-DK |  |
| Representation of names and scripts |  |  | Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and LRT-6 |  | Too generic, specific standards, e.g., ISO 15924, should be recommended. |
| Representation of entries in dictionaries |  |  | Mentioned in LRT-6 and D5C-3 |  | Too generic, specific standards should be recommended. |
| RDQL (RDF Data Query Language) | Query language |  | <http://www.w3.org/Submission/RDQL/> ; Mentioned in D5C-3 |  | Recommend SPARQL |
| REST (Representational state transfer) | Infrastructure |  | <http://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm> ; Mentioned in LRT-6 | TLA (in software), CLARIN-DK, CLARIN-PL |  |
| RQL (RDF Query Language) | Query language |  | <http://www.w3.org/TandS/QL/QL98/pp/rdfquery.html> ; Mentioned in D5C-3 |  | Recommend SPARQL |
| RTF (Rich Text Format) | Data format |  | <http://www.microsoft.com/en-us/download/details.aspx?id=10725> ; Mentioned in LRT-6 | CLARIN-PL |  |
| SeRQL (Sesame RDF Query Language) | Query language  Tool |  | <http://openrdf.callimachus.net/sesame/2.7/docs/users.docbook?view#chapter-serql> ; Mentioned in D5C-3 |  | Recommend SPARQL |
| SGML (Standard Generalized Markup Language) | Data format | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecSGML> |  |  |  |
| SKOS (Simple Knowledge Organization System) | Knowledge representation | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecSKOS> | Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and LRT-6 and D5C-3 |  |  |
| SemAF (Semantic annotation framework) | Data format | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecSemAF> | Mentioned in D5C-3 |  |  |
| SemRoleML (Semantic role markup language) | Data format | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecSemRoleML> |  |  |  |
| SimpL-1 (Simplified natural language) | Guideline? | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecSimpL1> |  |  |  |
| SOAP (Simple Object Access Protocol) | Infrastructure |  | <http://www.w3.org/TR/soap/> ; Mentioned in LRT-6 | CLARIN-PL |  |
| SPARQL (Simple Protocol and RDF Query Language) | Query language | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecSPARQL> | Mentioned in D5C-3 |  |  |
| SRX (Segmentation Rules eXchange) | Data format or  Knowledge representation? | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecSRX> |  | CLARIN-PL |  |
| SRU/CQL | Infrastructure |  | <http://www.loc.gov/standards/sru/> | TLA |  |
| Structured vocabularies for information retrieval | Data format | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecSV> |  | CLARIN-PL |  |
| SynAF (Syntactic annotation framework) | Data format | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecSynAF> | Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and D5C-3 |  |  |
| Systems to manage terminology, knowledge and content |  |  | Mentioned in D5C-3 | CLARIN-PL | Too generic |
| TermLR | Knowledge resource |  | <http://semanticweb.kaist.ac.kr/research/tc37sc4/new_doc/ISO_TC_37_SC4_N312_List_of_terms.pdf> ; Mentioned in D5C-3 |  | Can the TC 37 i-Term terminology resource be opened up for the public? |
| TBX (TermBase eXchange) | Data format | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecTBX> | Mentioned in LRT-6 and D5C-3 |  |  |
| TEI (Text Encoding and Interchange) Guidelines | Guideline  Data format  Metadata schema | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecTei> | Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and LRT-6 and D5C-3 | TLA? (Could be in some XML), CLARIN-DK, CLARIN-PL |  |
| TextMD (Technical Metadata for Text) | Metadata schema | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecTextMD> | <http://www.loc.gov/standards/textMD/> |  |  |
| TIFF (Tagged Image File Format) | Data format |  | <http://partners.adobe.com/public/developer/en/tiff/TIFF6.pdf> ; Mentioned in LRT-6 | TLA, CLARIN-DK |  |
| TimeML (Markup Language for Temporal and Event Expressions) | Data format | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecTimeML> | Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and D5C-3 | CLARIN-PL |  |
| Tipster | Data format |  | <http://www-nlpir.nist.gov/related_projects/tipster/> ; Mentioned in LRT-6 |  |  |
| TMF (Terminological markup framework) | Data format? | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecTMF> | Mentioned in LRT-6 and D5C-3 |  |  |
| TMS (terminology management systems) | Guideline | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecTMS> |  |  |  |
| TMX (Translation Memory eXchange) | Data format | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecTMX> | Mentioned in LRT-6 and D5C-3 |  |  |
| Topic Maps | Knowledge representation | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecTopicMaps> |  |  |  |
| TRIPLE | Query language |  | <http://triple.semanticweb.org/> ; Mentioned in D5C-3 |  | Recommend SPARQL |
| Turtle | Data format | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecTurtle> |  |  | See RDF, RDF/XML |
| Unicode | Data format |  | <http://www.unicode.org/> ; Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and D5C-3 | TLA, CLARIN-PL |  |
| URI (Uniform Resource Identifier) / URN (uniform resource name) | Infrastructure |  | <http://www.w3.org/TR/uri-clarification/> ; Mentioned in LRT-6 |  |  |
| WordNet | Knowledge resource  Knowledge representation |  | <http://wordnet.princeton.edu/> ; Mentioned in D5C-3 | CLARIN-DK, CLARIN-PL | There are many WordNets. This can refer to the original, i.e., or the approach. The original can be used as a pivot. The approach would need recommended data formats. |
| WordSeg | Data format | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecWordSeg> | Mentioned in D5C-3 |  |  |
| WSDL (Web Service Description Language) | Infrastructure |  | <http://www.w3.org/TR/wsdl> ; Mentioned in LRT-6 | CLARIN-PL |  |
| XCES (Corpus Encoding Standard in XML) | Data format | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecXces> | Mentioned in LRT-6 and D5C-3 | CLARIN-PL | See CES |
| XHTML (Extensible HyperText Markup Language) | Data format | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecXHTML> |  | CLARIN-PL |  |
| XML (Extensible Markup Language) | Data format | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecXML> | Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and LRT-6 and D5C-3 | TLA, CLARIN-DK, CLARIN-PL | Could be considered too general. Better to recommend specific XML schemas/vocabularies. |
| XMLNS (XML Namespace) | Data format | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecXMLNS> |  | TLA |  |
| XPath | Query language | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecXPath> |  | TLA (in software), CLARIN-DK, CLARIN-PL |  |
| XQuery | Query language | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecXQuery> |  | TLA (in software), CLARIN-DK |  |
| XSD (XML Schema Definition) | Schema language | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecXSD> | Mentioned in LRT-6 | TLA |  |
| XSL-FO (XSL Formatting Objects) | Data format | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecXSLFO> |  | TLA (in software) |  |
| XSLT (Extensible Stylesheet Language Transformations) | Transformation language | <http://clarin.ids-mannheim.de/standards/views/view-spec.xq?id=SpecXSLT> |  | TLA (in software), CLARIN-DK, CLARIN-PL |  |

# 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 preferrably a line of ISOcat URIs for each column)
* feature structure representation: ISO 24610-1:2006
* representation of primary sources: TEI (Text Encoding Initiative)
* knowledge engeneering: 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 lossles 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]](#footnote-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]](#footnote-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]](#footnote-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]](#footnote-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]](#footnote-6) contains short information items about a large number of the specifications.

Regarding LMF:

STO-LMF Syntax

http://cst.ku.dk/english/sto\_ordbase/STO-LMF-syntax-documentation-v1.pdf

STO-LMF Morphology

http://cst.ku.dk/english/sto\_ordbase/STO-LMF-morphology-documentation-v1.pdf

Regarding CMDI/TEI-P5

Text metadata (DK-CLARIN Technical Report, June 2011)

https://www.clarin.dk/documentation/DKCLARIN\_Metadata\_TEIP5DKCLARIN\_text-header\_jun2011.pdf

and https://www.clarin.dk/documentation/DKCLARIN\_textCorpusProfile-documentation-appendices.xlsx)

Text formatting (DK-CLARIN Technical Report, June 2011)

https://www.clarin.dk/documentation/DKCLARIN\_TEIP5DKCLARIN\_WP21\_text-format\_jun2011.pdf

Using TEI, CMDI and ISOcat in CLARIN-DK, in: LREC paper 2014:

http://www.lrec-conf.org/proceedings/lrec2014/pdf/325\_Paper.pdf

textCorpusProfile – metadata for text corpora in CLARIN-DK (Draft, July 2014)

https://www.clarin.dk/documentation/DKCLARIN\_textCorpusProfile-documentation-text.docx

# 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?

1. For those items a technical description is available at the URL <http://clarin.ids-mannheim.de/standards/>. [↑](#footnote-ref-1)
2. <http://www.clarin.eu/file/1511> [↑](#footnote-ref-2)
3. <http://www.clarin.eu/sites/default/files/Standards%20for%20LRT-v6.pdf> [↑](#footnote-ref-3)
4. <http://www.clarin.eu/sites/default/files/standards-CLARIN-ShortGuide.pdf> [↑](#footnote-ref-4)
5. <http://hdl.handle.net/1839/00-DOCS.CLARIN.EU-51> [↑](#footnote-ref-5)
6. <http://www.flarenet.eu/sites/default/files/FLaReNet_Standards_Landscape.pdf> [↑](#footnote-ref-6)