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 | IDS Standards DB[[1]](#footnote-1) | Further information | Used in CLARIN centre |
| CES | technical | yes |  |  |
| Citer |  |  | Mentioned in CLARIN Standardisation Action Plan and LRT-6 (as ‘Handles’) |  |
| CHAT |  |  | Mentioned in LRT-6 |  |
| CMDI | technical | yes |  |  |
| Controlled Vocabulary | technical | yes |  |  |
| Country Codes |  |  | Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and LRT-6 |  |
| CLARIN MDI |  |  | Mentioned in LRT-6 |  |
| CQLF | technical | yes |  |  |
| CQP |  |  |  |  |
| CSV |  |  | Mentioned in CLARIN EU Standard FAQ and LRT-6 |  |
| DataCITE |  |  |  |  |
| DCAM | technical | yes |  |  |
| DCMES | technical | yes | Mentioned in LRT-6 |  |
| DCR | technical | yes | Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and LRT-6 and D5C-3 |  |
| DiAML | technical | yes | Mentioned in D5C-3 |  |
| DictionaryEntry-RePresentation | technical | yes |  |  |
| DITA | technical | yes |  |  |
| DOL | technical | yes |  |  |
| DSSSL | technical | yes |  |  |
| EAD |  |  | Mentioned in LRT-6 |  |
| EAF |  |  | Mentioned in LRT-6 |  |
| Feature structures | technical | yes | Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and D5C-3 |  |
| GOLD | technical | yes | Mentioned in LRT-6 |  |
| H264 |  |  | Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and LRT-6 |  |
| Handle |  |  |  |  |
| HTML | technical | yes | Mentioned in LRT-6 |  |
| HTTP |  |  |  |  |
| HyTime | technical | yes |  |  |
| IMDI | technical | yes | Mentioned in LRT-6 |  |
| IPA |  |  | Mentioned in CLARIN Standardisation Action Plan |  |
| ISBD | technical | yes |  |  |
| ISO-Thesauri | technical | yes |  |  |
| ITS | technical | yes |  |  |
| JATS | technical | yes |  |  |
| JPEG |  |  | Mentioned in LRT-6 |  |
| JSON |  |  |  |  |
| LAF | technical | yes | 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 | yes | Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and LRT-6 and D5C-3 |  |
| MAF | technical | yes | 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 | yes | Mentioned in LRT-6 |  |
| MJPEG2000 |  |  | Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and LRT-6 |  |
| MLIF | technical | yes | 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 | yes |  |  |
| NISO MIX | technical | yes |  |  |
| nRQL |  |  | Mentioned in D5C-3 |  |
| OAI PMH |  |  | Mentioned in LRT-6 |  |
| OLAC Metadata | technical | yes | Mentioned in LRT-6 |  |
| OLiA | technical | yes |  |  |
| OntoIOp | technical | yes |  |  |
| Ontology Mapping Lang |  |  | Mentioned in D5C-3 |  |
| ORE |  |  | Mentioned in LRT-6 |  |
| OWL | technical | yes | 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 | yes |  |  |
| PNG |  |  | Mentioned in LRT-6 |  |
| PML |  |  | Mentioned in D5C-3 |  |
| RDF | technical | yes | Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and LRT-6 |  |
| RDF/XML | technical | yes | Mentioned in CLARIN EU Standard FAQ |  |
| RDFS | technical | yes | Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and LRT-6 |  |
| RELAX NG | technical | yes | 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 | yes |  |  |
| SKOS | technical | yes | Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and LRT-6 and D5C-3 |  |
| SemAF | technical | yes | Mentioned in D5C-3 |  |
| SemRoleML | technical | yes |  |  |
| SimpL-1 | technical | yes |  |  |
| SOAP |  |  | Mentioned in LRT-6 |  |
| SPARQL | technical | yes | Mentioned in D5C-3 |  |
| SRX | technical | yes |  |  |
| SRU/CQL |  |  |  |  |
| Structured Vocabulary | technical | yes |  |  |
| SynAF | technical | yes | 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 | yes | Mentioned in LRT-6 and D5C-3 |  |
| TEI Guidelines | technical | yes | Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and LRT-6 and D5C-3 |  |
| TextMD | technical | yes |  |  |
| TIFF |  |  | Mentioned in LRT-6 |  |
| TimeML | technical | yes | Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and D5C-3 |  |
| Tipster |  |  | Mentioned in LRT-6 |  |
| TMF | technical | yes | Mentioned in LRT-6 and D5C-3 |  |
| TMS | technical | yes |  |  |
| TMX | technical | yes | Mentioned in LRT-6 and D5C-3 |  |
| Topic Maps | technical | yes |  |  |
| TRIPLE |  |  | Mentioned in D5C-3 |  |
| Turtle | technical | yes |  |  |
| 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 | yes | Mentioned in D5C-3 |  |
| WSDL |  |  | Mentioned in LRT-6 |  |
| XCES | technical | yes | Mentioned in LRT-6 and D5C-3 |  |
| XHTML | technical | yes |  |  |
| XML | technical | yes | Mentioned in CLARIN EU Standard FAQ and Standardisation Action Plan and LRT-6 and D5C-3 |  |
| XMLNS | technical | yes |  |  |
| XPath | technical | yes |  |  |
| XQuery | technical | yes |  |  |
| XSD | technical | yes | Mentioned in LRT-6 |  |
| XSL-FO | technical | yes |  |  |
| XSLT | technical | yes |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

# 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.

# 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)