

# Unified list of standards: next steps forward

Piotr Bański, Hanna Hedeland, Dieter Van Uytvanck, **CLARIN Standards Committee**

Previously: questionnaires on accepted/used data formats (H. Hedeland)

– limited feedback, limited verifiability

At present: synergize the BoD and CSC efforts via the **Key Performance Indicator**:

**percentage of B-centres that publish explicit information on what formats they accept**

**Survey A.** A 2018 survey across 21 centres: from websites and DSA/CTS assessments (H. Hedeland)

**Survey B.** 7 centres in a recent BoD **KPI-related survey** (D. Van Uytvanck): see below

	Mimetype	Ext	MPI-PL	Cocoon	UDS	HZSK	AGD@IDS	BAS	ARCHE	Your Centre	Support
Wave	audio/x-wav	.wav	1	1	1	1	1		1		6
PDF	application/pdf	.pdf	1	1	1	1			1		5
Praat Textgrid	text/praat-textgrid	.TextGrid	1	1		1	1	1			5
ELAN	text/x-eaf+xml	.eaf	1	1		1	1	1			5
MPEG 4 video	video/mp4	.mp4	1	1		1	1	1			5
Plain text	text/plain	.txt	1	1	1				1		4
FLAC	audio/flac	.flac		1			1		1		3
SVG	image/svg+xml	.svg	1		1				1		3
TIFF	image/tiff	.tiff	1		1				1		3
HTML	text/html	.html	1		1				1		3
CHAT	text/x-chat	.cha	1	1		1					3
XML	text/xml	.xml	1				1		1		3
MPEG 1 video	video/mpeg	.mpg	1			1	1				3
TEI	application/tei+xml	.tei		1		1			1		3
ANVIL	??	.anvil				1	1				2
AIFF	audio/x-aiff	.aiff					1	1			2
GIF	image/gif	.gif	1		1						2
JPEG	image/jpeg	.jpg	1		1						2
PNG	image/png	.png	1		1						2
CSV	text/csv	.csv				1			1		2
Transcriber	text/x-trs	.trs		1			1				2
XSD	text/xml	.xsd	1						1		2
MPEG 2 video	video/mpeg	.mpeg	1				1				2
Matroska	video/x-matroska	.mkv		1					1		2
Folker	??	.fln				1					1
PhonDat 1	??	??						1			1
PhonDat 2	??	??						1			1
Emu Database	??	??						1			1
SAM Format	?? (text/plain)	??						1			1
AGS Format	?? (text/xml)	??						1			1

(22 B-centres in the Centre Registry as of Oct 1 2019: the **KPI is at 31.8 %**)

Provisos:

- multiple MIME types or extensions for a single format
- some format references are far too broad (“XML”...)
- status of the format in the field (international standard, local standard, tool format, ...)
- “age” of the format/standard (newly introduced, established, retiring, ...)
- preference by the centre: (preferred, unproblematic, acceptable)
- directionality (ingest, availability/export, internal)
- data types (e.g. documentation vs. transcription @HZSK need different formats)

→ ignore these for the time being, we’ll get to that in March

Simple (simplistic) threshold-based measure of standard popularity: the number of centres that publish information about supporting this standard (see above).

From there, the CSC will un-apply the provisos, in order to further refine the unified list of CLARIN standards...

**Please make sure that your centre joins those that count for the KPI!**