

# Media Types and Annotation Formats

Hanna Hedeland, Archiv für Gesprochenes Deutsch  
Leibniz-Institut für Deutsche Sprache, Mannheim

# What's the problem?

## Several media types in use for single formats

- application/wav (7675), application/x-wav (124), audio/vnd.wave (2216), audio/wav (5068), audio/x-wav (72390), audio/x-wave (1)

## Use of the (deprecated, RFC 6648) x- prefix and non-IANA-registered media types

- text/x-eaf+xml, text/x-chat, application/x-cmdi+xml

## But the existing IANA registered or compatible solutions are also unsatisfactory:

- Generic media types are not very useful, e.g. application/xml, text/csv
- Even specific registered media types are still not specific enough, e.g. application/tei+xml

## *What happened to the format-variant parameter suggestion?*

# Media Types and Annotation Formats in CLARIN

2014: CE-2014-0421 Relevant Formats + Google Spreadsheet

||

2021: CLARIN Standards Information System (SIS) + XML data on GitHub

- The inventory of relevant formats is still work in progress!

The guidance for the CTS R.8 “Appraisal” includes a published list of preferred formats, but this information has to be accurate, up-to-date, specific and - FAIR!

- Centres should be able to provide technical information on file formats, the SIS can provide the infrastructure
- The CCR includes the approved concepts *mime type* and *annotation format*

# Resources for media types and (annotation) file formats

- [The CLARIN Standards Information System \(SIS\)](#)
- Values of VLO's Format facet ([ConceptLink mime type](#), [dc:format](#))
- Values of [ConceptLink annotation format](#)
- [The SSHOC project's conversion hub inventory](#)
- [The \(very old\) Linguistic Annotation Wiki](#)
  
- [IANA's list of registered media types](#)
- [The Library of Congress Format Descriptions \(FDD\)](#)
- [PRONOM \(National Archives\)](#)

# Some Questions

Is the use of specific/specified *media types* feasible or do we need additional *annotation format* information?

The LoC already provides information on [mappings to PRONOM and Wikidata](#)

➤ Is Wikidata a solution for CLARIN or do we use CCR and SKOS only?

How can the information in the SIS improve interoperability or at least the knowledge about the current situation and limitations?