# Metadata Curation CMDI taskforces

CLARIN Centre & Developers meeting
21-22 May 2019
Utrecht, The Netherlands

# Agenda

https://tinyurl.com/TFs-utrecht-2019

- Summary of work since CAC
- Best practices and standardisation
- Concepts/CCR
- Vocabularies/CLAVAS
- Curation

- Common use cases questionnaire

# Summary of work: best practices and standardisation

- Best practices guide
- Common use cases
- Recommended components
  - Work merged (for now) with ISO 24622-3

# Summary of work: CMDI best practices guide

- Numbered best practices for modelling, authoring
- Common approaches and problems (work in progress)
- ‘Living document’
  - Current version: 1.2.0
- Open source: GitHub + Overleaf
  - All contributions are welcome!

https://www.clarin.eu/content/cmdi-best-practice-guide

# Summary of work: CMDI best practices guide

- Improvements in recent versions
  - First release based on Overleaf/LaTeX (after move from GitBook)
  - More ‘proper’ document with bibliography, acknowledgements and no severe layout issues
  - Improved labels to tag priority levels, verifiability, B-centre requirement
  - Editorial work (e.g. spelling)
  - Meta-documentation
- Work in progress
  - Document alignment with FAIR metrics
  - New best practices,
  - Expanded and/or refined existing best practices
  - Work on new common approaches and problem sections

# Summary of work: common use cases

- Work started in the **context of the CMDI Best Practices Guide**
  - meant to give advise on which CMDI metadata to include, which components to use, etc.
  - The idea arose from dealing with user feedback and support necessities of CLARIN-D's Discipline Specific Working Groups (F-AGs).
- Helpful to CLARIN newcomers
  - Learn about the usage of CMDI with hands-on information
- Separated from the CMDI Best Practices Guide (remains activity of TF)
- Aim: insight into metadata recognition for special linguistic resources
- Approach: questionnaire to be sent to CLARIN members asking experts for their input
- Currently going, also to be pursued at CLARIN Annual Conference 2019
- After that: produce Common Use Cases document to be published for CMDI users
- Common use cases VS Recommended Components:
  - **Specific information** VS **information generally important across resources**.

# Summary of work: CMDI best practices

## Common use cases

- Aim: gather descriptions and requirements from different use cases
- By resource type: *collection, corpus, text, lexicalResource, grammar, structuredDataset, annotation, image, audio, video, session, toolService, physicalObject*
- By genre: *bibliographic, newspaper, parliamentary, conversation, event, cultural heritage*
- Approach: questionnaire (under development)

# Summary of work: standardisation

"ISO CMDI pt3"

- ISO 24622:
  Language resource management -- Component metadata infrastructure (CMDI)
  - ISO 24622-1 (published)
    - Part 1: The Component Metadata Model
  - ISO 24622-2 ("under development")
    - Part 2: Component metadata specification language
  - ISO 24622-3 (to be proposed)
    - Part 3: Recommended components
- Bottom-up approach: based on current practices and knowledge gained in previous projects

# Summary of work: curation

- VLO value harmonisation (automated post-hoc curation)
  - ‘Modality’ facet: draft for vocabulary and mapping

- Curation module and link checking
  - Curation module 3.0 in development
  - Link checker developments + VLO integration

# Summary of work: harvest/import workflow

- OAI-PMH harvester
  - Ongoing development
    - Introduced use of SaxonUtils
  - Externalised harvest configuration files
    - https://github.com/clarin-eric/oai-harvest-config
    - Pull requests are welcome!
- Harvest viewer (https://vlo.clarin.eu/oai-harvest-viewer)
  - Performance and stability improvements

# Agenda

## https://tinyurl.com/TFs-utrecht-2019

- ~~Summary of work since CAC~~
- Best practices and standardisation
- Concepts/CCR
- Vocabularies/CLAVAS
- Curation

- Common use cases questionnaire

# Common use cases questionnaire

https://tinyurl.com/cmdi-common-use-cases

- Please fill in the form in and
  - (a) send in your information as it applies to you
  - (b) send conceptual feedback regarding the form if applicable