



The CMDI Explorer

Denis Arnold, Ben Campbell, Thomas Eckart, Bernhard Fisseni, Thorsten Trippel, **Claus Zinn**

CLARIN 2020 Virtual Conference



Motivation

- Scientific resources often comprise many different parts
- Their metadata description (should) mirror their complexity
- Understanding complex metadata descriptions is hard
- Tool required to support reading complex metadata (and act on it)

Background

- CMDI is de-facto standard in CLARIN universe
- Virtual Language Observatory helps users access scientific resources via their metadata, which works well for simply structured resources:

Nganasan Spoken Language Corpus (NSLC)

[Record details](#) [Links \(2\)](#) [Availability](#) [All metadata](#) [Technical Details](#) [Hierarchy](#)

Use the tree below to explore the hierarchy this record is part of. Click on the name of any of the records in this hierarchy to navigate to that item.

Nganasan Spoken Language Corpus (NSLC)

- Communication "TKF_990816_ShamanChild_flkd" of "Nganasan Spoken Language Corpus (NSLC)"
- Communication "PTK_930316_Djemnime_nar" of "Nganasan Spoken Language Corpus (NSLC)"
- Communication "ChND_041212_Girl_flkd" of "Nganasan Spoken Language Corpus (NSLC)"
- Communication "MDN_97_Chiimi_song" of "Nganasan Spoken Language Corpus (NSLC)"
- Communication "KES_061020_Habtyrkangku_song" of "Nganasan Spoken Language Corpus (NSLC)"
- Communication "ChND_99_War_flkd" of "Nganasan Spoken Language Corpus (NSLC)"
- Communication "TKF_031118_Reindeers_nar" of "Nganasan Spoken Language Corpus (NSLC)"
- Communication "JDS_160318_Chiimi_flk" of "Nganasan Spoken Language Corpus (NSLC)"
- Communication "KVB_97_Animals_flkd" of "Nganasan Spoken Language Corpus (NSLC)"
- Communication "MVL_080304_NjomuKamleguNy_flks" of "Nganasan Spoken Language Corpus (NSLC)"
- Communication "JMD_080219_MyLife_nar" of "Nganasan Spoken Language Corpus (NSLC)"
- Communication "JDH_00_Njaakju_flkd" of "Nganasan Spoken Language Corpus (NSLC)"
- Communication "MDN_97_Chukchis_flkd" of "Nganasan Spoken Language Corpus (NSLC)"

Background

- CMDI is de-facto standard in CLARIN universe
- Virtual Collection Registry helps users managing virtual collections, , which works well for simply structured resources:

Resources ▼		
Type	Reference	Actions
Resource	Peer Gynt (in English) An untraced edition, apparently from 1875.	...
Resource	HDL 2027/njp.32101068574639	...
Resource	HDL 2027/wu.89035754027	...
Resource	HDL 2027/uc1.32106002253281	...
Resource	HDL 2027/mdp.39015005058386	...
Resource	HDL 11245/1.146820 Dutch PhD thesis (with English abstract) by R.G.C. van der Zalm	...
Resource	HDL 1874/250346 Dutch BA thesis by L.G. de Jong	...



The CMDI Explorer

- Complements CLARIN tool space with missing functionality
 - Will attach to VLO
 - Will attach to VCR
 - Is connected to Switchboard
- Visualizes / helps navigating hierarchical structure of complex CMDI metadata
- Attaches actions to nodes in the resource tree
 - View
 - Download
 - Process (with Switchboard)



The CMDI Explorer - Demonstration

- <https://weblicht.sfs.uni-tuebingen.de/CMDIExplorer/>
- **Input data examples:**
 - <http://hdl.handle.net/11022/0000-0007-CFD8-D> (small)
 - <http://hdl.handle.net/11858/00-1778-0000-0008-C024-C> (Thomisticus)
 - <http://hdl.handle.net/11022/0000-0000-4F7F-B> (Rudi Völler)
 - <http://hdl.handle.net/10932/00-01B8-AE41-41A4-DC01-5> (IDS)
 - <http://hdl.handle.net/10932/00-01B8-AFF5-B57C-C801-B> (IDS subset)

CMDI Explorer x +
weblicht.sfs.uni-tuebingen.de/CMDIExplorer/input
CMDI Explorer Explore Virtual Collection Inventory Help CLARIN

Enter your metadata

Upload CMDI File Submit Handle Submit CMDI Text

http://hdl.handle.net/11022/0000-0007-CFE2-1

Submit URL

Stop CMDI Extractor

CMDI Extraction successful

100%

Filter with:

Search...

Download selection

- [-] [+] Data associated with Dima (2019) dissertation: "Composition Models for the Representation and Semantic Interpretation of Nominal Compounds" [+] [-]
- [-] [+] de-nn-only-comp
 - 11022_0000-0007-CFD8-D_README.txt
 - 11022_0000-0007-CFD8-D_dev_text.txt
 - 11022_0000-0007-CFD8-D_full_text.txt
 - 11022_0000-0007-CFD8-D_test_text.txt
 - 11022_0000-0007-CFD8-D_train_text.txt
- > de-mixed-comp
- > de-ulex
- > en-vector-rep
- > de-vector-rep
- > en-comcom
- > de-nncom-sem

[Download list.html](#)

[Download list.json](#)