

CLARIN-NL

- National project in NL
- 2009-2015
- Budget: 9.01 m euro
- Funding by NWO
- Coordinated by Utrecht University
- [>33 partners](#)

The CLARIN Infrastructure

- A technical research infrastructure for **humanities** researchers who work with **language resources**
- find all relevant data relevant
 - find all relevant tools and services
 - apply the tools and services to the data
 - store data and tools resulting from the research via one portal

Search for Data and Tools

- **Virtual Language Observatory** : Faceted browsing and geographical navigation (CLARIN-prep)
- **Meertens CLARIN LR Search**

Data and Search in Data

Lexical Data

- **COAVA** Curated Dutch Dialect Dictionaries + CHILDES browser
- **Cornetto** data in LMF and RDF + Interface
- **GrNe** search interface to Classical Greek-Dutch dictionary
- **DuELME_MWE** database in LMF + interface

Linguistically Annotated Corpora

- **Corpus Gysseling** with search interface
- **DISCAN** discourse-annotated text corpus
- **FESLI** search in curated specific language impairment data
- **GrETEL** treebank search (CLARIN- NL / Flanders cooperation)
- **IProSLA** sign language data
- **MIMORE** search in [DiDDD](#), [Dynasand](#), and [GTRP](#) micro-comparative data
- **NEHOL** Negerhollands annotated corpus
- **TDS** typological database system
- **VU-DNC** historical newspaper corpus

Historical & Contemporary Corpora

- **BILAND** multilingual search in historical text corpora
- **Polimedia** multimedia data + search interface
- **Verrijkt Koninkrijk** data + search interface
- **WAHSP** historical sentiment mining in public media
- **WIP** search in the Dutch parliamentary proceedings

Other Data

- **Arthurian Fiction** search interface
- **C-DSD** song database
- **MIGMAP** interface for migration analysis
- **NameScope** search for names in literary works
- **PILNAR** curated pilgrimage data

Annotate Data

- **AAM-LR** annotation of audio-files
- **Multicon** enhancements of ELAN/ANNEX for multimodal collocations
- **TQE** tool for transcription quality evaluation

Process Data

- **Adelheid** tokenizer, lemmatizer, PoS-tagger for Historical Dutch
- **Gabmap** web application for dialect variation analysis
- **INPOLDER** parser for Historical Dutch
- **NameScope** Named/Entity Tagger
- **TICCLops** orthographic normalizer
- **TTNWW web service** work flow system for Dutch (normalisation, PoS-tagging, Parsing, NE-recognition, speech file transcription, ...)

And much more to come in 2014...

- See full paper for a list

INVITATION

- Use this functionality in your research!
- Provide feedback so that we can improve and extend!
- Start page: www.clarin.nl/node/404

Store Data

- **Metadata**
 - **ARBIL** Metadata Editor
 - **CMDI Component registry**
- **Semantic Interoperability**
 - **ISOCAT** web interface and services
 - **RELCAT** relation registry
 - **SCHEMACAT** schema registry
 - **CLAVAS** vocabulary service
- **Ingestion**
 - **LAMUS** (TLA)
 - **EASY** (DANS)

Wish list

- Metadata for many more **data and tools** in the VLO
- Many tools must be professionalized
- Search through multiple resources must be extended
- Processing must be followed by search / analysis options
- Interoperability must be further improved
- Parameterized queries
- Chaining Search (syntax, morphology, semantics)
- Interoperability with existing analysis and visualisation tools must be improved
- Facilities for improving verifiability and replicability