

CLARIN-NL Annual Report 2014

1 Introduction

This document provides a description of the activities carried out in CLARIN-NL for 2014.

Explanations and expansions of acronyms in this document can be found on the CLARIN-NL website¹ and in Appendix D.

2 Executive Summary

The major achievement for 2014 is the securing of funding for the CLARIAH-CORE project (12 million euro), which, on the one hand, is a successor to the CLARIN-NL project, and, on the other, an extension of it to other disciplines and data types.

Special attention has been paid to educational activities: several educational modules have been developed, various lectures and tutorials have been given, including a course at the 2014 LOT Summer school, and more are planned for early 2015: a course at the LOT 2015 Winter school, and plenary sessions at TIN and CLIN.

The CLARIN-NL Portal has been set up, which includes (inter alia) faceted search to data and services yielded by CLARIN-NL, so that researchers can easily find and select what is on offer.

All projects from the 4 calls have been finished.

3 Planned Activities

1.1 EU-Line (WBS 1.9)

As planned, the budget for this item was spent on the ERIC membership fee:

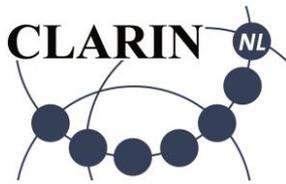
Alice Dijkstra is the representative for the Netherlands in the CLARIN ERIC General Assembly, with Jan Odijk as expert.

CLARIN-NL is very active with regard to the European enterprise to create a CLARIN infrastructure under the guidance of CLARIN ERIC, and in other CLARIN or infrastructure-related projects and initiatives. This can be illustrated with many examples:

- CLARIN ERIC is hosted by the Netherlands, more specifically Utrecht University
- The CLARIN ERIC director is Steven Krauwer.
- NL is represented in the National Coordinator's Forum, attends the monthly (virtual) meetings, and is active in it. (Jan Odijk/Daan Broeder)
- Active participation in CLARIN ERIC committees, e.g.²
 - Marc Kemps-Snijders, Member Standing Committee for CLARIN Technical Centres³

¹ <http://www.clarin.nl/node/8>

² This is the same as in 2013.



- Daan Broeder member Centre Assessment Committee
- Daan Broeder, member CLARIN ERIC standards committee
- Menzo Windhouwer, member CLARIN ERIC standards committee
- Eelco Ferwerda, member CLARIN ERIC Legal Issues Committee
- Arjan van Hessen, Web Editorial Board
- Arjan van Hessen, Arwin van der Zwan, Hetty Winkel in the CLARIN ERIC Office
- CLARIN-NL employees have organized LREC tutorials and workshops on CLARIN-related topics (metadata, semantic interoperability)
- Other tutorials and workshops that attract international researchers have been organized
- A large delegation from the Netherlands attended the second [CLARIN Annual Conference](#) in Soesterberg, the Netherlands, and many of them had contributions in the form of presentations and/or showing demonstrators (about 10 presentations, demos or posters with NL contributors).
- Many CLARIN-NL people and/or organizations are involved in important European projects and initiatives that are directly relevant to infrastructures in general and the CLARIN infrastructure in particular, e.g.
 - [DASISH](#)⁴, which brings together [all 5 ESFRI research infrastructure initiatives](#)⁵ in the social sciences and humanities
 - [EUDAT](#)⁶, which brings together data management service providers and research communities who are directly involved with the design of data services. The consortium includes key representatives from research community infrastructures from diverse disciplines.
 - [RDA](#)⁷ (Research Data Alliance), which is an organisation that aims to accelerate and facilitate standardizing and harmonizing research data sharing and exchange.
 - [CLARA](#) is training a new generation of researchers who will be able to cooperate across national boundaries on the establishment of a common language resources infrastructure and its exploitation for the construction of the next generation of language models with wide theoretical and applied significance

1.2 NL-Line

1.2.1 Technical (WBS 1.1)

The project to create essential infrastructural functionality (IIP project) and the project to create search facilities in metadata and actual data in the infrastructure (S&D project) have run well and have been finished. The CLAVAS project, in which a SKOS-based knowledge system to provide integrated access to a variety of managed vocabularies (copied from other knowledge sources) was created, so that CLARIN users can transparently use concepts and data categories from each of them, also has been fully finished.

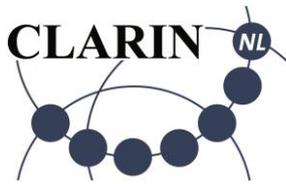
³ Dieter Van Uytvanck, is Chair Standing Committee for CLARIN Technical Centres (on behalf of Germany)

⁴ <http://dasish.eu/>

⁵ <http://dasish.eu/links/>

⁶ <http://www.eudat.eu/>

⁷ <http://rd-alliance.org/>



In 2014 a CLARIN Centre Meeting as set up. In the course of the year it was extended with IISH in light of the CLARIAH project. The centre meeting was held 4 times in 2014, and was focused on the certification and the role of centres in the new CLARIAH project.

The procedure to be recognized as a CLARIN Type B Centre was started up by INL, Meertens, and Huygens ING, and was successfully concluded by all (Huygens ING in early 2015). DANS has not participated yet but indicated an intention to participate in a later stage. MPI was already certified as a CLARIN Type B centre in an earlier round.

In the course of 2014, the MPI direction indicated that it wanted to reduce its activities in the context of the TLA. This implies that it would, after about 2 years, not function any more as a CLARIN type A-centre, and though the data it hosts would be secure, it is not clear how easily they can be accessed and whether they will be made available in a CLARIN-compatible manner in the future.

Furthermore, the ISOcat system for semantic interoperability was evaluated in CLARIN, and it was decided to phase-out its use and to turn to a different system. Support for ISOcat by the ISO registration authority and maintainer (MPI-PL) would finish by December 2014, which made it very urgent to create an alternative in 2014. For this reason, a project (called 14-014 OpenSKOS) was started up in the Netherlands to set up such a new system, called the CLARIN Concept Registry (CCR), which was developed and is hosted at Meertens Institute and is based on OpenSKOS, which Meertens already has a lot of experience with (cf. e.g. CLAVAS). The CCR has been implemented, the CLARIN relevant data categories from ISOcat have all been transferred to the CCR, and a clean-up action has started there. References to data categories in ISOcat remain resolvable.

It was observed many times already that the quality of the metadata provided by the CLARIN centers created seen in the light of one of the major functions of metadata (discoverability of resources) was quite low. A project was started up in 2013 and finished in 2014 to define a procedure to improve metadata quality (project 13-011 MCC). The resulting report is available and will be used in CLARIAH. In addition, test in 2014 (reported on in [\[Odiijk 2014\]](#)) revealed that resources are very difficult to find in the Virtual Language Observatory (VLO), and an analysis of these problems suggests that improvements are needed for metadata and the VLO. These issues have been raised at the European level (inter alia in the metadata curation taskforce), and have been taken up there, and will also be taken up nationally in CLARIAH.

1.2.2 CLARIN Data Providers

The Netherlands has a range of centres that already or since recently have making data available in their core mission and that are of great importance to humanities researchers. These include the *Koninklijke Bibliotheek* (KB, National Library), *het Nederlands Instituut voor Beeld & Geluid* (NIBG, Netherlands Institute for Sound and Vision), *het Nationaal Archief* (NA, the National Archive), the *Digitale Bibliotheek van de Nederlandse Literatuur* (DBNL, Digital Library of Dutch Literature), the various university libraries, and also organizations such as Department of information Services of the *Tweede Kamer* (Parliament). CLARIN-NL wants to make sure that data from such organizations is available to humanities researchers in the CLARIN infrastructure in a CLARIN-compliant manner. For

this reason, a special type of centre has been defined, a *CLARIN Data Provider* (also sometimes called a CLARIN D-Centre), and several of the organizations mentioned have been invited to make a proposal to become such a *CLARIN Data Provider*:

- KB has submitted a plan, which has been executed. The project finished mid-2014. The amount of metadata that the KB provides is so large that currently they are not included in the VLO (because by their sheer quantity of 1M records they would overwhelm the other metadata). This aspect is taken into account in plans for new versions of the VLO, and, of course, in Nederlab.
- NIBG executed its plan successfully already in 2013. The so-called Academia-collection is now accessible in a CLARIN-compatible manner, e.g. via the VLO. Access to the actual data does require a subscription to the Academia-service, which most (but not all) Dutch universities have. Most foreign universities and research institutes do not have such a subscription.
- DBNL (now part of the KB) finished its work by the end of January 2014.
- UBU finished its work early in 2014.

1.2.3 CMDI Metadata scheme for tools

Though full-scale testing of the CMDI *metadata scheme for tools* developed earlier was planned, this has not been carried out due to lack of available resources. However, part of it has been done in the context of the CLARIN portal, and the plan is to export the information included there into CMDI metadata descriptions and extend these in CLARIAH.

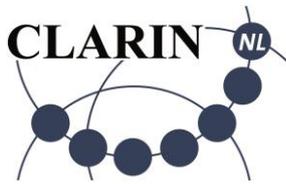
1.2.4 CMDI Metadata as Linked Open Data

Since the Linked Data (LD) paradigm is rapidly gaining popularity, it was decided to start up a project to automatically convert CMDI metadata as Linked Data. This project (14-003 CMDI2RDF) was carried out and CMDI data can now be published as Linked open data via a Virtuoso Triple store. This work will be continued in CLARIAH, where LD will play an important role. In CLARIAH it will also be investigated and attempted to convert metadata represented as LD to CMDI.

1.2.5 Integration

An important activity started up in 2013 was an *integration activity* (the *CLAPOP project*) in which the various components created separately in a variety of projects (infrastructure functionality, search functionality, curated data, curated tools, demonstrators, web applications, etc. etc.) come together and are made accessible via a single portal to the CLARIN infrastructure. The [CLARIN-NL portal](#) has been created, and it contains

- Faceted search for services (tools, web applications, web services), which include some properties which were proposed in the *Metadata for Tools project* as facets
- Faceted Search for data
- Overview of the Dutch CLARIN Centres
- Links to metadata search applications and to federated content search
- Overview of educational packages developed in CLARIN-NL and other educational material



- An initial version of registry and faceted browsing for standards and best practices supported by CLARIN, this work was meant as a show-model to discuss a standardisation registry with CLARIN EU and other CLARIN projects.

The portal uses CLARIN federated identity management setup to avoid multiple authentications when already logged-in with a CLARIN service.

The idea is that the portal will continue to exist also long after the CLARIN-NL project has finished. Its maintenance will be dealt with in CLARIAH after the end of the CLARIN-NL project.

Though it currently is a CLARIN-NL portal, it should turn into a portal containing all information that is relevant to CLARIN in the Netherlands, and perhaps even, in the context of CLARIAH, to a Digital Humanities portal. The services registry and faceted search were quite successful, and has been adopted by CLARIN Poland, and discussions are on-going with CLARIAH Austria and LINDAT (Czech Republic) on how to raise it to the EU-level.

1.2.6 Federated Login

Certain tools have ad-hoc login procedures instead of federated login as required by CLARIN. We planned to have these tools adapted so that they would support federated login. This is on-going for PILNAR, was difficult to realize for Gabmap (no knowledgeable software developer available for this), and was delayed by other problems (hosting) for BILAND/WAHP.

1.2.7 Interoperability and User-friendliness

It was observed that the interoperability of the developed functionality could be significantly improved by developing converters and wrappers. A project to develop converters (14-011 OpenConvert) has been started up in 2014 and has concluded in February 2015. It has created an initial set of converters, more specifically to convert from plain text, HTML, ePub, Word 1997-2003, Word 2010, and ALTO into TEI or FoLiA.

The development of wrappers has not been taken on systematically, though some individual projects provide the relevant functionality (e.g. the Frog tool for Part of Speech tagging allows not only plain text as input but also FoLiA and yields FoLiA as output. This will be taken up in CLARIAH again.

It was planned to make the TTNWW application more user-friendly. However, due to lack of resources, and in part by a dependency on the results of the OpenConvert project, no work on this has been done. It will be taken up again in CLARIAH.

For the same reason, a project to have TTNWW systematically also output CMDI metadata with provenance information was not started up.

1.2.8 Integration of Enrichment and Search

It was made possible to enrich ones data using TTNWW, e.g. with pos-tags, or full syntactic trees, but the output of these enrichments is a set of XML files that is not of much further use for the typical humanities researcher. It was therefore planned to enable uploading such enriched corpora in search engines. This has been done

- In the PaQU (14-010 Parse and Query) project. In the PaQu application, one can search in a Treebank for a pair of words and their grammatical properties and relation. One can upload a parsed corpus, or a plain text corpus, which will be parsed by Alpino, after which it is available for search. This application has been successfully tested in a small research project reported on at the 2015 LOT winter school CLARIN course, at CLIN, and at TIN. The application is fully functional, but it is the intention to add federated login to have it hosted at INL in 2015 .
- In the AutoSearch project a search application with the functionality of the Contemporary Dutch Corpus search application allows the upload of a text corpus enriched with pos-codes by Frog, e.g. in the TTNWW application. It was delivered early 2015.
- For GrETEL we did not manage to start up a project to provide similar functionality, mainly because of lack of resources.
- It was also planned to investigate linking OpenSONAR search with Cornetto in order to enable semantic filtering of search results, but this has been postponed to CLARIAH, since OpenSONAR only became available by the end of 2014.

The three applications mentioned will be further extended in the context of CLARIAH.

1.2.9 Data, Tools and Services (WBS 1.2/1.3/1.4)

Calls

All projects from all Calls (Calls 1 through 4) have been finished. Most of them have also been formally concluded, but for some the formal conclusion (submission and approval of final reports, and/or final payment) is still ongoing.

The results of Call 4 projects (and of other projects started up in 2014) have been demonstrated on March 13, 2015 (CLARIN-NL final Event/ CLARIAH Kick-off).

The Nederlab demonstrator project has finished, and an initial version of the Nederlab application has been made available since March 13, 2015.

All calls have been calls for data curation and demonstrator projects. In the approved projects data and tools developed and in use in the Netherlands will be curated and a series of demonstrators will be developed on the basis of existing technologies and applications to attempt to make them CLARIN-compatible. A major aim with these projects is to test the standards proposed within CLARIN against these data and tools, and to be able to provide evidence-based feedback on the proposed standards and thus to influence their determination in the CLARIN context.

In each project metadata in accordance with CMDI will have to be created and tested, making use of the experience gained and the tools developed in the Metadata project described above. Each demonstrator project will create a demonstrator web application, in which a clear separation between user interface modules and core modules is made. An API has to be defined and documented for the core module. The user interface will make calls to the core module via this API. This will make it possible later to easily turn the core module into a web service.

All projects have to deliver a document describing the requirements the infrastructure should meet and or desirable features it should offer, so that these can form input for the requirements and specifications of the CLARIN infrastructure.



In the first call, 17 project proposals were submitted. Of these, 11 were awarded funding by the Board on the basis of recommendations from the National Advisory Panel (NAP) and the Executive Board (EB). In addition, the CKCC project was awarded funding.

In the second call, 17 project proposals were submitted. Of these, 9 were awarded funding by the Board on the basis of recommendations from the National Advisory Panel (NAP) and the Executive Board (EB).

In the third call, Closed Call, five researchers were invited to submit a proposal. The disciplines covered were Philology/Classical Studies, Studies on selective language impairment, Religion Studies, the study of discourse, and descriptive linguistics/language documentation. Four proposals were submitted, and have started early 2012.

No proposal was submitted for the discipline descriptive linguistics/language documentation, but the researcher in question was requested to still submit a proposal, and this became the LAISEANG project led by Marianne Klamer.

In the third call, Open Call, 22 project proposals were submitted. 2 proposals were rejected because of IPR-issues. Of the remaining 20 project proposals, 9 were awarded funding by the Board on the basis of recommendations from the International Advisory Panel (IAP), the National Advisory Panel (NAP) and the Executive Board (EB).

In the fourth call, Closed Call, 2 project proposals were submitted and awarded funding. In the Call 4 Open Call 18 project proposals were submitted. One of these was rejected by the EB because of non-compliance, and of the remaining 17, 9 could be awarded funding.

An overview of the projects is provided here. For each project, it consists of the project acronym and title, the project coordinator, a very short description of the project goal, a very short description of the scientific impact of the project, and its current status. The status describes the situation as of 31 December 2014, though where needed also its status as of March 2015 is specified. More detailed descriptions of the projects can be found on the CLARIN-NL website.⁸

1.2.10 Call 1 Projects

ADEPT	Assaying Differences via Edit-Distance of Pronunciation Transcriptions
Project Coordinator	Dr. C. Gooskens (Groningen University)
Goal	The goal of the project is to provide a web application capable of measuring the differences in sets of phonetic (or phonemic) transcriptions via edit distance.
Budget	€ 57K
Centre	Meertens Instituut
Scientific impact	A user-friendly web application version of this existing tool will make it usable for a wide range of humanities researchers such as, phonologists, phoneticians, dialectologists and second-language learning specialists.

⁸ <http://www.clarin.nl/node/76>

Status	The project has effectively finished and has submitted its final reports, which have been approved. [web service]
---------------	---

AAM-LR	Automatic Annotation of Multi-modal Language Resources
Project Coordinator	Prof.dr. L. Boves (Radboud University Nijmegen)
Goal	The AAM-LR project aims at building a demonstrator of a web service that will help field researchers to annotate audio- and video-recordings
Budget	€ 56K
Centre	MPI
Scientific impact	It will facilitate all field research in which digital audio or video recordings are used.
Status	The project has effectively finished and has submitted its final reports, which have been approved. [web application ⁹ show case ¹⁰]

Adelheid	A Distributed Lemmatizer for Historical Dutch
Project Coordinator	Dr. H. van Halteren (Radboud University Nijmegen)
Goal	This project aims at providing a web-application with which an end user can have historical Dutch text tokenized, lemmatized and part-of-speech tagged.
Budget	€ 55K
Centre	MPI
Scientific impact	It will facilitate research on historical texts in fields such as historical linguistics, literature science and history.
Status	The project has finished and has submitted its final reports, which have been approved. The Second Call project INPOLDER makes use of the results of Adelheid. [web service]

DUELME-LMF	Converting DUELME into LMF format
Project Coordinator	Prof.dr. J.E.J.M. Odijk (Utrecht University)
Goal	The goal of the project is develop converters between DUELME and LMF format, and to create a curated DUELME resource fully compliant with standards supported by CLARIN.
Budget	€ 60K

⁹ <http://lux17.mpi.nl/aamlr>. This is a temporary URL, which may disappear in the future

¹⁰ <http://www.clarin.nl/showcase/aam-lr/268>

Centre	INL
Scientific impact	The DUELME database makes it possible to address a variety of research questions related to Dutch multi-word expressions in research areas such as computational linguistics and natural language processing, theoretical linguistics and psycholinguistics.
Status	The project has effectively finished, and has submitted its final reports, which have been approved. [metadata]

INTER-VIEWS	Curation of Interview Data
Project Coordinator	Dr. H. van den Heuvel (Radboud University Nijmegen)
Goal	The INTER-VIEWS project will make a corpus of interview data available to the community of researchers in the humanities.
Budget	€ 58K
Centre	DANS
Scientific impact	This project will contribute to facilitating all research that makes use of interviews, and more specifically research into the Second World War.
Status	The project has effectively finished, and has submitted its final reports, which have been approved. [web service]

MIMORE	Microcomparative Morphosyntax Research Tool
Project Coordinator	Prof.dr. S. Barbiers (Meertens Institute and Utrecht University)
Goal	The demonstrator tool MIMORE will create a common search engine for the databases DynaSAND , DiDDD and MAND
Budget	€ 54K
Centre	Meertens Instituut
Scientific impact	The tool is especially relevant to research in variation linguistics. It will allow investigating theoretical questions concerning the language system and language variation, and the geographic distribution of (morpho-) syntactic variables.
Status	The project has effectively finished, and has submitted its final reports, which have been approved. [web service]

SignLinC	Linking lexical databases and annotated corpora of signed languages
Project Coordinator	Dr. O.A. Crasborn (Radboud University Nijmegen)
Goal	This project aims to link two independently evolved data sets for a signed language: the Corpus NGT and the lexical database of the Dutch Sign Centre.
Budget	€ 56K
Centre	MPI

Scientific impact	The linked databases and corpora will facilitate research in the area of sign language linguistics
Status	The project has effectively finished, and has submitted its final reports, which have been approved. [web service]

TICCLops	Text-Induced Corpus Clean-up online processing system
Project Coordinator	Dr. M. Reynaert (Tilburg University)
Goal	This demonstrator project will develop a tool that allows CLARIN users to submit their corpora for fully automatic spelling correction and normalization
Budget	€ 54K
Centre	INL
Scientific impact	The tool will facilitate all research that makes use of text corpora.
Status	The project has effectively finished, and has submitted its final reports, which have been approved. [web service]

TDS Curator	A web-services architecture to curate the Typological Database System
Project Coordinator	Dr. A. Dimitriadis (Utrecht University)
Goal	TDS Curator will make the Typological Database System (TDS) into a sustainable service that conforms to CLARIN infrastructural requirements.
Budget	€ 57K
Centre	DANS
Scientific impact	The (TDS) facilitates typological linguistic research by providing integrated access to multiple independently developed typological databases through a common web interface
Status	The project has finished and has submitted its final reports, which have been approved [web service]

TQE	Transcription Quality Evaluation
Project Coordinator	Dr. H. Strik (Radboud University Nijmegen)
Goal	The TQE project aims to create a completely automatic Transcription Quality Evaluation (TQE) tool.
Budget	€ 57K
Centre	MPI
Scientific impact	The tool will be useful for validating, obtaining, and selecting phone transcriptions, for detecting phone strings (e.g. words) with deviating pronunciation, and, in general, it can be usefully applied in all research - in various (sub-) fields of

	humanities and language and speech technology in which audio and phonetic transcriptions are involved.
Status	The project has effectively finished, and has submitted its final reports, which have been approved. [web service]

WFT-GTB	Integrating the Wurdboek fan 'e Fryske Taal into the Geïntegreerde Taalbank
Project Coordinator	Drs. H. Sijens (Fryske Akademy)
Goal	The project aims to carry out data curation of the <i>Wurdboek fan de Fryske Taal (WFT)</i> database and integrate the data in the <i>Geïntegreerde Taalbank (GTB)</i> and demonstrate it via the GTB dictionary web application.
Budget	€ 56K
Centre	INL
Scientific impact	Integrating the WFT dictionary in the GTB will facilitate studying formal, semantic and idiomatic aspects of Modern West Frisian by allowing extensive forms of free and structured (in particular thematic) search queries, and it will enable comparative studies with Dutch materials.
Status	The project has effectively finished and demonstrated its results at Euralex 2010 held in Leeuwarden. It has submitted its final reports, which have been approved. [web application]

CKCC project	CKCC project
Project Coordinator	Drs. G. Gerritsen (Huygens ING)
Goal	With respect to CLARIN-NL, the project increases its use of language technology and produces the results in CLARIN-compatible manner. The project runs until Nov 1, 2012.
Budget	€ 59K
Centre	DANS/Huygens ING
Scientific impact	The CKCC project (<i>Circulation of Knowledge and Learned Practices in the 17th-century Dutch Republic, 'Geleerdenbrievenproject'</i>) is an independently financed (NWO) project that is carried out by a consortium of Dutch institutes (Descartes, Historical Literature Amsterdam, Huygens, KB (National Library), DANS and VKS) and coordinated by Wijnand Mijnhardt (Utrecht University). It investigates, on the basis of a corpus of 20,000 letters of scientists from the 17th century in the Dutch Republic and using language technology, the research question of how knowledge circulated in the 17 th century. It was selected in the CLARIN-EU call for humanities and social sciences projects as the project proposal that “[would] best demonstrate the use of LRT and would show the potential of a research infrastructure in the humanities” (CLARIN Newsletter 6, p. 3). ¹¹

¹¹ http://www.clarin.eu/files/cnl06_web_0.pdf

Status	The project has finished and is currently being evaluated. [web application]
---------------	--

1.2.11 Call 2 Projects

Arthurian-Fiction	Arthurian Fiction in Medieval Europe
Project Coordinator	Dr. B. Besamusca (Utrecht University)
Goal	ArthurianFiction is a mixed demonstrator and data curation project that aims to carry out data curation for two databases with data for literary research in the area of European Arthurian fiction and to create a demonstrator that will allow users access to the data for searching and editing.
Budget	€ 47K
Centre	Huygens ING
Scientific impact	It will enable access to the Arthurian Fiction database to a wide range of humanities researchers such as literary historians in all European languages, cultural historians, medieval book specialists and art historians.
Status	The project is finished and has submitted its final reports, which have been approved. [web application]

C-DSD	Curating the Dutch Song Database
Project Coordinator	Dr. E. Stronks (Utrecht University)
Goal	C-DSD is a data curation project that aims to curate the Dutch Song Database (DSD), a database in the field of Literary Studies containing (meta-) data on 140.000 songs and their 15.000 sources from the Middle Ages to the present day.
Budget	€ 29K
Centre	Meertens Institute
Scientific impact	This project will create new opportunities for (international) collaboration with other projects with similar source materials.
Status	The project submitted its final reports, which have been approved. [metadata in the VLO]

COAVA	Cognition, Acquisition and Variation Tool
Project Coordinator	Dr. L. Cornips (Meertens Institute Amsterdam)
Goal	COAVA is a mixed demonstrator and data curation project proposal in which a tool is developed for easily exploring the linguistic characteristics of objects from two databases of language variation (historical dialectology) on the one hand and three

	first language acquisition databases on the other hand.
Budget	€ 66K
Centre	Meertens Institute
Scientific impact	It will enable interdisciplinary research into the relation between language acquisition and language variation.
Status	The project has finished and submitted its final reports, which have been approved. [web application]

INPOLDER	Integrated Parser and Lemmatizer Dutch in Retrospect
Project Coordinator	Prof.dr. A. van Kemenade (Radboud University Nijmegen)
Goal	INPOLDER is a mixed demonstrator and data curation project that aims to provide syntactically analysed corpus material for historical Dutch. It uses the Adelheid tagger for PoS-tagging and a variant of the Penn-Helsinki parser for historical texts for enriching the texts with parse trees.
Budget	€ 69K
Centre	Meertens Institute
Scientific impact	It will fill the gap in availability of syntactically analysed corpus material for historical Dutch, enabling historical syntactic, morpho-syntactic and lexical research as well as benefiting non-linguistic historical research.
Status	The project has finished and the final reports have been submitted and approved. [web application]

IPROSLA	Integrating and publishing resources on sign language acquisition
Project Coordinator	Prof.dr. P. Fikkert (Radboud University Nijmegen)
Goal	IPROSLA is a resource curation project that aims to integrate two different data sets on sign language acquisition by documenting the two data sets with CMDI, and archiving them at the MPI language archive. The data consist of a set of longitudinal data of deaf children from deaf and hearing parents collected at the UvA, and a new collection of longitudinal data collected at the RU from hearing children of deaf parents.
Budget	€ 67K
Centre	MPI
Scientific impact	By integrating two datasets in integrated archive and making them visible and accessible, it will enable better and more research on Sign Language.
Status	The project has finished and submitted its final reports, which have been approved. [metadata via the VLO]

NEHOL	Negerhollands Database
Project Coordinator	Prof.dr. P.C. Muysken (Radboud University Nijmegen)
Goal	NEHOL is a data curation project that aims to make available to the CLARIN community the data from the Dutch-lexifier Creole language Negerhollands, the now extinct Creole language of the Virgins Islands.
Budget	€ 43K
Centre	MPI
Scientific impact	It will enable research into Negerhollands and thereby into Creole languages and language creolization processes in general.
Status	The project has finished and submitted its final reports, which have been approved. [metadata via the VLO]

VU-DNC	VU Diachronic Newspaper Corpus
Project Coordinator	Prof.dr. W. Spooren (Free University Amsterdam)
Goal	VU-DNC is a data curation project that aims to make a unique diachronic corpus of Dutch newspaper articles from five major Dutch newspapers from 1950/1951 and 2002 (2 million words) available to humanities researchers, to extend the discourse annotation with additional lexico-grammatical features and to create a gold standard benchmark for OCR-post-correction tools, all of this in a fully CLARIN-compliant manner.
Budget	€ 59K
Centre	INL
Scientific impact	It will enable research into the recent history and development of the Dutch written language, as well as research into subjectivity and quotations and the lexico-grammatical features encoding them.
Status	The project has finished and submitted its final reports, which have been approved. [Data and its metadata]

WAHSP	Web-application for historical sentiment mining in public media
Project Coordinator	Prof.dr. T. Pieters (Utrecht University)
Goal	WAHSP is a demonstrator project that aims to provide advanced forms of text mining (more specifically, sentiment mining) in large historical datasets of newspapers and journals in the form of a CLARIN compliant web-application, addressing research questions of historians and policy researchers.
Budget	€ 120K
Centre	Huygens ING
Scientific	It will enable historical research by providing facilities for mining large historical

impact	data sets, where it focuses on the opinions/perceptions regarding the use and abuse of drugs between 1900 and 1945.
Status	The project has finished and submitted its final reports, which have been approved. [web application] . There was a problem with login, but it was agreed to solve this in combination with BILAND (see there).

WIP	War in Parliament
Project Coordinator	Dr. H. Piersma (NIOD)
Goal	WIP is a mixed demonstrator and data curation project that aims to make the Dutch Hansard database compliant with the CLARIN infrastructure and to provide an advanced search engine for it in order to make it possible to carry out historical and social science research, in particular with respect to the Second World War.
Budget	€ 118K
Centre	DANS
Scientific impact	It will enable historical research by providing tools for text mining large text corpora, applied in this project in particular for references to the Second World War in parliamentary debates.
Status	The project has finished and submitted its final reports, which have been approved. [web application]

1.2.12 Call 3 Projects (Closed Call)

FESLI	Functional Elements in Specific Language Impairment
Project Coordinator	Prof.dr. F. Weerman (University of Amsterdam)
Goal	FESLI (CLARIN-NL-11-006) is a mixed demonstrator and resource curation project in the Closed Call that aims to curate Dutch corpora from monolingual and bilingual children with Specific Language Impairment (SLI) and provide tools based on the CLARIN-NL project COAVA that allow quantitative and qualitative comparisons to be made esp. with regard to functional elements in language between the SLI data and CHILDES monolingual data.
Budget	€ 79K
Centre	Meertens Instituut
Scientific impact	The project will curate important sets of monolingual and multilingual SLI data, improve search facilities in the data and will thus greatly facilitate research into SLI.
Status	The project has finished and submitted its final reports, which have been approved. [web application]

PILNAR	Pilgrimage Narratives: creating a germ corpus for studying the profile of the
---------------	--

	modern pilgrim
Project Coordinator	Prof.dr. P. Post (Tilburg University)
Goal	PILNAR (CLARIN-NL-11-007) is a mixed demonstrator and resource curation project in the Closed Call that aims to curate a corpus of pilgrimage narratives and use the heuristic tool "Fields of the Sacred" as a search tool for disclosing the material.
Budget	€ 79K
Centre	Meertens Instituut
Scientific impact	It will make searching in pilgrimage narratives much easier and more powerful, this boosting research in the religion sciences.
Status	The project has finished and submitted its final reports, which have been approved. It has been decided to implement the lacking federated login separately [web application]

	Grieks-Nederlands woordenboek (Greek-Dutch dictionary)
GrNe Project Coordinator	Prof.dr. I. Sluiter (Leiden University)
Goal	GrNe (CLARIN-NL-11-008) is a mixed demonstrator and resource curation project in the Closed Call that aims curate a part of the dictionary Ancient Greek-Dutch currently under construction and to create a tool for search functionality including an interface in which (inter alia) inflected forms can be input and the entries of their lemmata will be returned using an existing lemmatizer (called 'parser' in the proposal, which appears to be a misnomer).
Budget	€ 55K
Centre	University Library of UvA
Scientific impact	It will make a part of the Classical Greek dictionary currently being produced available via a web interface, with increased search functionality offering search options both for lemmas and for inflected word forms, thus greatly facilitating research and education in classical philology.
Status	The project has finished and submitted its final reports, which have been approved. [web application]

	Towards a Discourse Annotation system for Dutch language corpora
DiscAn Project Coordinator	Prof.dr. T.J.M. Sanders (Utrecht University)
Goal	DiscAn (CLARIN-NL-11-031) is a resource curation project in the Closed Call that aims curate a set of existing corpora with discourse annotations, to develop the foundations of a discourse annotation system, and to develop an interdisciplinary Dutch-Flemish discourse community of linguists from various subdisciplines.
Budget	€ 80K

Centre	MPI
Scientific impact	The project will bring together and harmonize a wide range of discourse annotated corpora that are currently located in various places, in multiple different format sand generally not easily accessible, thus greatly facilitating research into discourse properties of natural language and organizing a research community working on this topic.
Status	The project has finished and submitted its final reports, which have been approved. The project has been followed up by the Mindisc project (see there). Metadata in the VLO .

1.2.13 Call 3 Projects (Open Call)

BILAND	Towards a flexible and stable CLARIN-supported web-application for bilingual historical analysis of discourses in news media
Project Coordinator	Prof.dr. T. Pieters (Utrecht University)
Goal	BILAND (CLARIN-NL-11-017) is a demonstrator project in the Open Call that aims to enable advanced forms of discourse analysis in large historical datasets in the CLARIN infrastructure.
Budget	€ 120K
Centre	Huygens ING
Scientific impact	The project will greatly facilitate research in history and of advanced discourse analysis in historical datasets.
Status	The project has finished and submitted its final reports, which have been approved. It was decided to set up a separate project to solve the remaining problems, viz. (1) Federated login was still not taken care of, despite the agreements made for WAHSP; (2) it turned out that a large part of the application runs on UvA servers, not on Huygens servers. However, so far not much progress has been made here [web application]

MultiCon	Multilayer Concordance Functions in ELAN and ANNEX
Project Coordinator	Dr. O.A. Crasborn (Radboud University Nijmegen)
Goal	MultiCon (CLARIN-NL-11-018) is a demonstrator project in the Open Call that aims to create collocation functionality for multimodal data in the tools ELAN and ANNEX by improving their query and concordance functionality and by providing visualization options for multilayer collocations.
Budget	€ 55K
Centre	MPI
Scientific impact	The project will offer new research opportunities in multimodality studies, both in linguistics and beyond.

Status	The project has finished and submitted its final reports, which have been approved. Link to ELAN
---------------	--

D-LUCEA	Database of the Longitudinal Utrecht Collection of English Accents
Project Coordinator	Dr. H. Quené (Utrecht University)
Goal	D-LUCEA (CLARIN-NL-11-011) is a resource curation project in the Open Call that aims to curate a database of existing speech recordings of L1 and L2 speakers of English.
Budget	€ 67K
Centre	MPI
Scientific impact	This project will greatly facilitate linguistic and phonetic research into L1 and L2 acquisition.
Status	The project has finished and submitted its final reports, which have been approved.

EMIT-X	Early-Modern Image and Text eXchange
Project Coordinator	Dr. E. Stronks (Utrecht University)
Goal	EMIT-X (CLARIN-NL-11-009) is a resource curation project in the Open Call that aims to curate Emblems Project Utrecht (EPU) resources by converting them into a commonly agreed upon format (no name given) based on work by Rawles and Stäcker and making them available for OAI-PMH harvesting.
Budget	€ 20K
Centre	Huygens ING
Scientific impact	The project will greatly facilitate research in literary studies, esp. the study of emblems.
Status	The project has finished and submitted its final reports, which have been approved. Metadata in the VLO.

NameScope	Mapping the Landscape of Names in Modern Dutch Literature
Project Coordinator	Prof.dr. K. van Dalen-Oskam (Huygens ING)
Goal	NameScope (CLARIN-NL-11-029) is a demonstrator project in the Open Call that aims to annotate a substantial amount of literary works with a rich tag set for named entities using a named entity recognizer and make them accessible and visualize them via a demonstrator web application.
Budget	€ 120K
Centre	INL

Scientific impact	The project will greatly enhance the research opportunities in literary studies esp. the study of named entities and their roles in literary works.
Status	The project has finished and submitted its final reports, which have been approved. [web application]

VK	Verrijkt Koninkrijk (Enriched Kingdom)
Project Coordinator	Dr. K. Ribbens (NIOD)
Goal	VK (CLARIN-NL-11-014) is a mixed demonstrator and resource curation project in the Open Call that aims to make a demonstrator to gather data on pillarization in Dr. Lou de Jong's VK and to enrich the resulting data and make them available to the CLARIN community.
Budget	€ 111K
Centre	DANS
Scientific impact	The project will greatly facilitate historical research, esp. on the Second World War, and on Lou de Jong's work on this topic, and more specifically on the representation of pillarization in this work.
Status	The project has finished and submitted its final reports, which have been approved. [web application]

Cornetto-LMF-RDF	Curated Cornetto database in LMF and RDF
Project Coordinator	Prof.dr. P. Vossen (Free University Amsterdam)
Goal	Cornetto-LMF-RDF (CLARIN-NL-11-020) is a mixed demonstrator and resource curation project in the Open Call that aims to convert the Dutch semantic lexical database Cornetto to LMF (lexicon part) and RDF/SKOS (WordNet part).
Budget	€ 117K
Centre	INL
Scientific impact	This project will greatly enhance the research opportunities in study of semantics in particular and linguistics more generally.
Status	The project has finished and submitted its final reports, which have been approved. Since the project spent much less than planned, it was decided to spend the remaining funds to further improvements of the interface and the efficiency of the underlying engine, in the 14-013 Cornetto II project. See below. [web application]

MIGMAP	Interactive migration maps for the 20th century
Project Coordinator	Dr. G. Bloothoof (Utrecht University and Meertens Institute Amsterdam)

Goal	MIGMAP (CLARIN-NL-11-012) is a demonstrator project in the Open Call that aims to develop a web application which visualizes migration by generation and gender based on the Meertens “Kaat” software module.
Budget	€ 54K
Centre	Meertens Institute
Scientific impact	The project will facilitate research into migration and linguistics, esp. language change.
Status	The project has finished and submitted its final reports, which have been approved. [web application]

Polimedia	Interlinking multimedia for the analysis of media coverage of political debates
Project Coordinator	Prof.dr. H. Beunders (Erasmus University Rotterdam)
Goal	PoliMedia (CLARIN-NL-11-022) is a mixed demonstrator and resource curation project in the Open Call that aims to showcase the potential of cross-media analysis for research in the humanities by curating annotated data sets of different media types and developing a demonstrator that enables researchers to deploy such an interlinked collection for analysis of media coverage.
Budget	€ 111K
Centre	NIBG
Scientific impact	The project will enhance research opportunities for the cross-media analysis in historical studies.
Status	The project has finished and submitted its final reports, which have been approved. [web application]

LAISEANG	Language Archive of Insular South East Asia and West New Guinea
Project Coordinator	Prof.dr. Marianne Klamer (Leiden University)
Goal	This project will ensure the preservation of unique records of languages and the cultures encapsulated by them in the region. The language resources have been gathered by twenty linguists at, or in collaboration with Dutch universities over the last 40 years, and will be compiled and archived in collaboration with The Language Archive (TLA) in Nijmegen. The resulting archive will constitute an unrivalled collection of multimedia materials and written documents from over 50 languages in Insular South East Asia and West New Guinea.
Budget	73k
Centre	MPI
Scientific impact	It will preserve unique records of language and culture documentation and make these available for research by the whole research community
Status	The project has finished and submitted its final reports, which have been approved. [Metadata in the VLO] .

1.2.14 Call 4 Projects

QuaMerdes	Quantitative Content Analysis of Media Researchers' Data
Project Coordinator	Dr. Jasmijn van Gorp (Utrecht University)
Goal	QuaMerdes is a demonstrator project from the closed call that will enable quantitative content analysis of television and printed media. It will expand the existing tool MeRDES to support principles of quantitative content analysis and explore new data visualisation paradigms. QuaMeRDES will be evaluated in a specific case study that enables media studies scholars to come to new insights about how representations of migrants on Dutch television are related to social, cultural and political values in Dutch society.
Budget	76k
Centre	NISV/KB
Scientific impact	It will enable quantitative content analysis of television and printed media.
Status	The project has finished and submitted its final reports, which have been approved. The application has ad-hoc login, and it was requested to add federated login. Login was necessary because of IPR-restrictions on certain data. However, it is likely that the login can be removed in full, so a decision on this has been postponed. If needed, federated login will be implemented in CLARIAH. [web application]

@PhilosTEI	<i>TICCLing Philosophy: a TEI corpus-building workflow towards a new computational methodology for philosophy</i>
Project Coordinator	Dr. Arianna Betti (Free university Amsterdam)
Goal	@PhilosTEI is a demonstrator project from the closed call in which an open source, web-based, user-friendly workflow is developed from textual digital images to TEI, based on an OCRopus / Tesseract web service and a multilingual version of the OCR-post correction web service TICCLops. The tool will be demonstrated on a multilingual, multi-script corpus of important 18th-20th-century European philosophical texts.
Budget	70k
Centre	HI
Scientific impact	It will contribute to increasing the digitally available empirical material, esp. for philosophy, but also for other disciplines
Status	The project has finished but still has to submit its final reports. [web application]

eBNM+	Linked Data on Middle Dutch Sources Kept Worldwide
--------------	---

Project Coordinator	Dr. André Bouwman (Leiden University Libraries)
Goal	<i>e-BNM+</i> is a mixed demonstrator and resource curation project from the open call with two goals: (1) to convert e-BNM collection of textual, codicological and historical information about thousands of Middle Dutch manuscripts kept worldwide into a flexible data structure that will turn e-BNM into a key open access resource to which many other resources can be easily linked; and (2) to create a web application for consultation, using faceted search, and collaborative editing.
Budget	68k
Centre	HI
Scientific impact	It will stimulate research that makes uses of Middle Dutch manuscripts.
Status	The project has finished and submitted its final reports, which have been approved. [web application]

VALID	<i>Vulnerability in Acquisition: Language Impairments in Dutch. Curating five valuable data sets</i>
Project Coordinator	Dr. Jetske Klatter (Radboud University Nijmegen)
Goal	VALID is a resource curation project from the open call that aims to curate five existing, digital data sets of language pathology data collected in the Netherlands, primarily on Dutch, in order to make them available for scientific research in CLARIN-compatible format, as a first, major step in the development of a VALID data archive.
Budget	64k
Centre	MPI
Scientific impact	It will enable better research on language pathology data, primarily for the Dutch language.
Status	The project has finished and submitted its final reports, which have been approved. [Metadata in the VLO]

COBWWWEB	Connections Between Women and Writings Within European Borders
Project Coordinator	Dr. Suzan van Dijk (Huygens ING)
Goal	COBWWWEB is a mixed demonstrator and resource curation project from the open call that aims to (1) connect the WomenWriters database to other national collections in women's literature; (2) build a research application for scholars; and (3) create a set of standards to exchange data based on CLARIN guidelines for shared metadata and service-based infrastructures, which will be implemented for a variety of databases by a range of international partners.

Budget	95k
Centre	HI
Scientific impact	It will enable better research on female writers and the reception of their works
Status	The project has finished and submitted its final reports, which have been approved. [web application]

OpenSoNaR	Online Personal Exploration and Navigation of SoNaR
Project Coordinator	Dr. Martin Reynaert (Tilburg University)
Goal	OpenSoNaR is a demonstrator project from the open call that aims to develop a tool for exploring the SoNaR-500 reference corpus by layman and specialist researcher alike, using an online frontend (WhiteLab) providing a range of interfaces that provide user-driven functionality, combined with the existing INL-developed back-end system (BlackLab).
Budget	120k
Centre	INL
Scientific impact	It will provide better empirical support for linguistic and literary research
Status	The project has finished and submitted its final reports, which have been approved. [web application]

SHEBANQ	System for HEBrew Text: ANnotation for Queries and Markup
Project Coordinator	Prof.dr. Wido van Peursen (Free University Amsterdam)
Goal	SHEBANQ is a mixed demonstrator and resource curation project from the open call that aims to curate the WIVU Hebrew Text Database and to develop a web application demonstrator that enables researchers to perform linguistic queries on the web resource and preserve significant results as annotations to this resource.
Budget	101k
Centre	DANS
Scientific impact	It will enable better research in the Hebrew bible texts, and in theology
Status	The project has finished and submitted its final reports, which have been approved. [web application]

DSS	Dutch Ships and Seamen
Project Coordinator	Prof.dr. Lex Heerma van Voss (Huygens ING)

Goal	DSS is a mixed demonstrator and resource curation project from the open call that aims to bring together the rich maritime historical data currently preserved in many different databases. To that end it will provide a tool chain and methodology for converting legacy datasets, and applies it to an initial set of selected datasets.
Budget	98k
Centre	HI
Scientific impact	It will enable better research into the maritime history of the Netherlands and its predecessors
Status	The project has finished and submitted its final reports, which have been approved. [web application]

EXILSEA	Exploiting ISOcat's Language Sections in ELAN and ANNEX
Project Coordinator	Dr. Onno Crasborn (Radboud University Nijmegen)
Goal	EXILSEA is a mixed demonstrator and resource curation project from the open call that aims to make annotated audio-visual resources more accessible to users of different languages by better exploiting the multilingual features of ISOcat in the ELAN and ANNEX tools. In addition it will enhance the Corpus NGT, the world's first open access sign language corpus, by updating the existing IMDI metadata to CLARIN-standard CMDI descriptions using bilingual ISOcat categories.
Budget	65k
Centre	MPI
Scientific impact	It will enable multilingual support for research into Sign Language
Status	The project has finished and submitted its final reports, which have been approved. [ELAN]

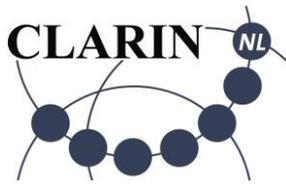
ColTime	Collaboration on time-based resources
Project Coordinator	Dr. Onno Crasborn (Radboud University Nijmegen)
Goal	ColTime is a demonstrator project from the open call that aims to extend ELAN and ANNEX for the annotation and display of time-based resources such as audio and video with a referencing and note exchanging system.
Budget	68k
Centre	MPI
Scientific impact	It will enable better research on time-based resources
Status	The project has finished but still has to submit one of its final reports. [ELAN]

RemBench	A Digital Workbench for Rembrandt Research
Project Coordinator	Prof.dr. Volker Manuth (Radboud University Nijmegen)
Goal	RemBench is a demonstrator project from the open call that aims to demonstrate how (art) historians can benefit from linking eLaborate to resources created by museums, archives and libraries. The demonstrator will use eLaborate to link the Rembrandt Documents (RemDoc) database with related resources created by the Rijksbureau voor Kunsthistorische Documentatie (RKD), as well as with a standard university library catalogue.
Budget	121k
Centre	HI
Scientific impact	It will enable better research on the Rembrandt Documents
Status	The project has finished and submitted its final reports, which have been approved. [web application]

1.2.15 Other Projects

Cooperation between the Netherlands and Flanders CLARIN-NL cooperates with CLARIN Flanders in a cooperation project. The overall project consists of two parts:

- **Part I (1356k Euro, NL: 904k Euro, VL: 452k Euro)**
 - In this part a specific cooperation plan has been formulated in which the Netherlands and Flanders aim to adapt various existing components developed in CGN and especially STEVIN and incorporate them into a work flow system for web services that is currently being developed in CLARIN. The whole system will run on servers of recognized CLARIN-centres and aims to offer facilities to researchers from the humanities with little or no technical expertise. These facilities must 1) allow them to address their research questions in a better and easier manner, and 2) create opportunities for formulating completely new types of research questions
 - In the Netherlands this part consists on the one hand of the infrastructure implementation plan described above (section 3) in which the basis for the technical infrastructure will be designed and implemented (with a budget of 504k euro). On the other hand 400kEuro will be invested for the further realization of this part (in the TTNWW project). The formal decision on the exact usage of this amount is planned to be made by the CLARIN-NL board in early 2010. Flanders contributes in this part 452k Euro in accordance with the usual NL: 2/3-VL: 1/3 proportion (all in the TTNWW project).
- **Part II (920k Euro; NL: 680k Euro; VL: 340k Euro)**
 - The contribution from the Netherlands consists of 680k Euro spent on the first open call for demonstrator en resource curation projects.



- In Flanders the research community represented in CLIF has formulated project proposals for a total amount of 340k Euro.

The project has finished and submitted its final reports, which have been approved. [Link to workflow system](#)¹².

ISOcat Coordinator

The work by the ISOcat Coordinator (Ineke Schuurman) started up in 2011 has continued in 2014.

CLAM-support

In the context of the TTNWW project, Maarten van Gompel (initially Tilburg University, currently Radboud University Nijmegen) has developed CLAM, software to facilitate turning existing tools into RESTful web services. CLAM is successful and is in use by most participants in the TTNWW project. For this reason, systematic support, maintenance and extension of the CLAM-functionality are required, which has been implemented in a project financed by CLARIN-NL to maintain and further extend CLAM, and to provide support to its users. (carried out by Maarten van Gompel since 2012, 1/10 appointment). This has been continued in 2014, extending the support also to FoLiA. A separate report on the activities of 2014 is available.

Centres of Expertise (WBS 1.6)

The CLARIN-NL Centre of Expertise on the curation of data, led by the Radboud University of Nijmegen but with close involvement of experts from the CLARIN Centres, has discontinued the CLARIN-NL Data Curation Service in 2014. It finished the last resources worked on early in 2014.

An actual overview of the data curation results and plans can be found on the [DCS project area](#)¹³ on the CLARIN-NL website. Here's a snapshot:

Completed:

- LESLLA (delivered in August 2013)
- IPNV Interviews with veterans (delivered in May 2013)
- Woordenboek Gelderse Dialecten, Rivierengebied (delivered in June 2013)
- Woordenboek Gelderse Dialecten, Veluwe part) (delivered in June 2013)
- Curation organisation names for CLAVAS / OpenSkos (delivered in June 2013; finalised in August 2013)
- Six dialect dictionaries from Brabant (delivered in October 2013)
- WLD and WBD part III (delivered in December 2013)
- DBD/TCULT (delivered in January 2014)
- Roots of Etnolect (collection Linda van Meel) (delivered in January 2014)

¹² <http://yago.meertens.knaw.nl/apache/TTNWW/>

¹³ <http://www.clarin.nl/node/147>

- Traces of Contact (collection Kofi Yakbo)

These resources have been finished but no further DCS activities were planned for 2014. A small amount of money was reserved to ensure proper documentation and preservation of knowledge for a possible restart of the DCS in the CLARIAH project, but this money was not necessary for this purpose.

As to the second Centre of Expertise, [Nederlab](#)¹⁴, on data and tools for the study of the Dutch language and culture across time, the Nederlab demonstrator project has finished. An initial version of the Nederlab application has been made available at the CLARIN-NL Final Event on March 13, 2015.

As a successor to the Cornetto-LMF-RDF project, a new project, 14-013 Cornetto II has been started up to create an improved interface to the Cornetto database and to resolve some performance issues. This project has finished in February 2015 but still has to submit its final reports.

The CLARIN-NL 13-012 LASSY project to improve and maintain the Groningen Word relations interface has finished but still has to submit its final reports.

1.2.16 Education and Awareness (WBS 1.7)

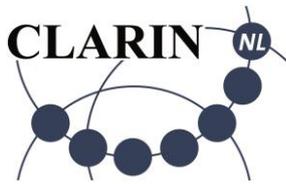
An overview is given of the various activities for stimulating education, training, and awareness carried out in 2014.

Conferences and workshops visited¹⁵

From	To	Event	City	Ctry	Who
14-jan-14	15-jan-14	THATCamp	Den Haag	financial	AvH
17-jan-14	17-jan-14	CLIN	Leiden	NL	many
10-apr-14	11-apr-14	'Mining Digital Repositories'	Den Haag	NL	JO
25-mei-14	1-jun-14	LREC 2014	Reykjavik	IS	many
6-jun-14	6-jun-14	Bessensap	Utrecht	NL	HB, AvH
23-Jun-14	26-Jun-14	CLARIN std. cie / ISO-TC37/SC4	Berlin	DE	DB, MW

¹⁴ <http://www.nederlab.nl/>

¹⁵ JO=Jan Odijk, DB=Daan Broeder, MW=Menzo Windhouwer, AvH=Arjan van Hessen, AD = Alice Dijkstra, HB=Hans Bennis



23-okt-14	25-okt-14	CLARIN Conference	Soesterberg	NL	many
26-nov-14	26-nov-14	CLARIN ERIC General Assembly	Tübingen	DE	JO, AD
1-dec-14	2-dec-14	Clarino Meeting ¹⁶	Tromsø	NO	JO
8-dec-14	9-dec-14	Exploring Historical Sources with Language Technology: Results and Perspectives	Den Haag	NL	JO
15-dec-14	16-dec-14	NeDiMAH	Den Haag	NL	DB

At CLIN, there were lectures and posters on SHEBANQ, OPENSonar (2x), GrETEL (2x), ISOcat, TiCCLOps/Nederlab, and NameScape.

At LREC, there were about 8 presentations on CLARIN-NL

At the CLARIN Annual Meeting, which was hosted by the Netherlands in 2014, there were about 10 presentation, posters and/or demo's from the Netherlands.

Events organized by CLARIN-NL

From	To	Description	City	Ctry	Who
4-mrt-14	4-mrt-14	CMDI Tutorial	Utrecht	NL	DB, MW

The *CLARIN-NL Final Event* is planned for March 13, 2015, in combination with the CLARIAH Kick-off.

Publications

A *Lingua* special issue is being prepared, with Jan Odijk as guest editor. Seven papers have been submitted. We provide some details about these contributions:

Topics covered: synchronic syntax (in 4 contributions), language variation (in 3 contributions), language acquisition, language contact, language typology, phonology/phonetics, and diachronic syntax (each in 1 contribution).¹⁷

#Authors involved: 20

Distribution of the primary authors by Country: Netherlands 4; Belgium, Czech Republic and Germany each 1.

¹⁶ Jan Odijk is member of the CLARINO Advisory Board.

¹⁷ Many contributions cover multiple topics

Distribution of all authors by Country: Netherlands: 14, Czech Republic: 4; Belgium and Germany each 1.

Enhanced publications: Though Lingua explicitly invited authors to submit enhanced publications, only one contribution comes with a data set in a separate file as support for the article. We will discuss with Lingua how this will be dealt with.

The review process is on-going, but it has some delay because many candidate reviewers refuse to review. It is expected that the special issue will be published by the end of 2015.

Here's a (not fully complete) overview of other publications (partly found using Google Scholar):¹⁸

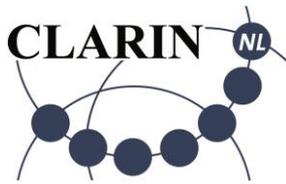
- Amigó, E., Carrillo-de-Albornoz, J., Chugur, I., Corujo, A., Gonzalo, J., Meij, E., ... & Spina, D. (2014). Overview of RepLab 2014: author profiling and reputation dimensions for Online Reputation Management. In *Information Access Evaluation. Multilinguality, Multimodality, and Interaction* (pp. 307-322). Springer International Publishing.
<http://nlp.uned.es/~damiano/pdf/replab2014overview.pdf>
- Baltussen, L. B., Blom, J., Medjkoune, L., Pop, R., Van Gorp, J., Huurdeman, H., & Haaijer, L. (2014). Hard Content, Fab Front-End: Archiving Websites of Dutch Public Broadcasters. *Alexandria*, 25(1), 69-91.
- Boer, Victor de, Jur Leinenga, Matthias van Rossum and Rik Hoekstra. 'Dutch Ships and Sailors Linked Data Cloud', Accepted for publication in *Proceedings of the International Semantic Web Conference (ISWC 2014)*, 19-23 October, Riva del Garda, Italy, 2014. [\[Draft\]](#)
- Bravo Balado, A. *Information extraction on newspaper archives for historical research. A Dutch maritime history case study*. M.Sc. thesis VU University Amsterdam (forthcoming), 2014.
- Broeder, D. et al. Experiences with the ISOcat Data Category Registry. *Proceedings of LREC 2014*, Reykjavik, Iceland. [\[URL\]](#)
- Broeder, D., & Lannom, L. W. (2014). Data Type Registries: A Research Data Alliance Working Group. *D-Lib Magazine*, 20(1), 2.
<http://dialnet.unirioja.es/servlet/articulo?codigo=4572541&orden=1&info=link>
- Buccio, E. D., Nunzio, G. M. D., & Silvello, G. (2014). A linked open data approach for geolinguistics applications. *International Journal of Metadata, Semantics and Ontologies*, 9
- Cai, F., Liang, S., & de Rijke, M. (2014, July). Personalized document re-ranking based on bayesian probabilistic matrix factorization. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval* (pp. 835-838). ACM.
<http://ilps.science.uva.nl/sites/ilps.science.uva.nl/files/sigir2014-personalisationFei.pdf>
- Chuklin, A., & de Rijke, M. (2015). The Anatomy of Relevance: Topical, Snippet and Perceived Relevance in Search Result Evaluation. arXiv preprint arXiv:1501.06412.
<http://arxiv.org/pdf/1501.06412>
- Chuklin, A., Zhou, K., Schuth, A., Sietsma, F., & de Rijke, M. (2014, July). Evaluating intuitiveness of vertical-aware click models. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval* (pp. 1075-1078). ACM.
<http://staff.science.uva.nl/~mdr/Publications/sigir2014-sp-clickmodels.pdf>
- Chuklin, A., Schuth, A., Zhou, K., & Rijke, M. D. (2015). A Comparative Analysis of Interleaving Methods for Aggregated Search. *ACM Transactions on Information Systems (TOIS)*, 33(2), 5.
<http://ilps.science.uva.nl/sites/ilps.science.uva.nl/files/tois2015-interleaving-for-aggregated-search.pdf>

¹⁸ For a more complete and up-to-date overview, see <http://www.clarin.nl/node/1959>

- Crasborn, O. & H. Sloetjes, Improving the Exploitation of Linguistic Annotations in ELAN. Proceedings of LREC 2014, Reykjavik, Iceland. [URL]
- Crasborn, O., Hulsbosch, M., Lampen, L., & Sloetjes, H. (2014). New multilayer concordance functions in ELAN and TROVA. In Tilburg Gesture Research Meeting [TiGeR 2013]. <http://pubman.mpg.de/pubman/item/escidoc:1760119/component/escidoc:1760120/crasborn.pdf>
- Dorofeeva, A. (2014). Towards Digital Humanities Tool Criticism. https://openaccess.leidenuniv.nl/bitstream/handle/1887/29597/MA_thesis_Dorofeeva_TowardsDigitalHumanitiesToolCriticism.pdf?sequence=1
- Durco and Windhouwer, The CMD Cloud. Proceedings of LREC 2014, Reykjavik, Iceland. [URL]
- M. van Gompel - CLAM: Computational Linguistics Application Mediator – Documentation – LST Technical Report Series LST-14-02 Nijmegen, 2014 [PDF] [URL]
- van Gompel, M. (2014). FoLiA: Format for Linguistic Annotation. <http://antalvandenbosch.ruhosting.nl/fofia.pdf>
- M. van Gompel and M. Reynaert - CLAM: Quickly deploy NLP command-line tools on the web. Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics, 2014 [PDF] [URL]
- Van Gorp, J. (2014). modes of self-representation: Visualized identities of former yugoslav migrant women in the netherlands. *Crossings: Journal of Migration & Culture*, 5(1), 153-171.
- Graus, D., van Dijk, D., Tsagkias, M., Weerkamp, W., & de Rijke, M. (2014, July). Recipient recommendation in enterprises using communication graphs and email content. In Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval (pp. 1079-1082). ACM. <http://graus.nu/paper/sigir2014-recipientrecommendation.pdf>
- Graus, D., Odijk, D., Tsagkias, M., Weerkamp, W., & de Rijke, M. (2014, July). Semanticizing search engine queries: the University of Amsterdam at the ERD 2014 challenge. In Proceedings of the first international workshop on Entity recognition & disambiguation (pp. 69-74). ACM. http://web-gram.research.microsoft.com/ERD2014/Docs/submissions/erd14_submission_8.pdf
- Graus, D., Tsagkias, M., Buitinck, L., & de Rijke, M. (2014). Generating pseudo-ground truth for predicting new concepts in social streams. In *Advances in Information Retrieval* (pp. 286-298). Springer International Publishing. <https://staff.science.uva.nl/m.derijke/Publications/ecir2014-fp-ookb.pdf>
- Heuvel, H. van den, Sanders, E., Klatter, J., Hout, R. van, Fikkert, P., Baker, A., Jong, J. de, Wijnen, F. & Trilsbeek, P. (2014). Data curation for a VALID Archive of Dutch Language Impairment Data. In: *Dutch Journal of Applied Linguistics*, 3(2), pp. 127-135. DOI:10.1075/dujal.3.2.02heu
- Hofmann, K., Whiteson, S., Schuth, A., & de Rijke, M. (2014). Learning to rank for information retrieval from user interactions by K. Hofmann, S. Whiteson, A. Schuth, and M. de Rijke with Martin Vesely as coordinator. *ACM SIGWEB Newsletter*, (Spring), 5. <https://staff.science.uva.nl/s.a.whiteson/pubs/hofmannsigweb14.pdf>
- Hofmann, K., Schuth, A., Bellogin, A., & de Rijke, M. (2014). Effects of Position Bias on Click-Based Recommender Evaluation. In *Advances in Information Retrieval* (pp. 624-630). Springer International Publishing. <http://www.anneschuth.nl/wp-content/uploads/2014/01/ecir2014-sp-interleaving.pdf>
- Huijnen, P., Laan, F., de Rijke, M., & Pieters, T. (2014). A Digital Humanities Approach to the History of Science. In *Social Informatics* (pp. 71-85). Springer Berlin Heidelberg. <http://ilps.science.uva.nl/sites/ilps.science.uva.nl/files/histoinformatics2013-wahsp.pdf>

- Huijnen, P., Laan, F., de Rijke, M., & Pieters, T. (2014). A Digital Humanities Approach to the History of Science Eugenics Revisited in Hidden Debates by Means of Semantic Text Mining. *SOCIAL INFORMATICS, SOCINFO 2013*, 8359, 70-84.
http://dspace.library.uu.nl/bitstream/handle/1874/303426/histoinformatics2013_wahsp.pdf?sequence=1
- Klatter, J., Hout, R. van, Heuvel, H. van den, Fikkert, P., Baker, A., Jong, J. de, Wijnen, F., Sanders, E. & Trilsbeek, P. (2014). Vulnerability in Acquisition, Language Impairments in Dutch: Creating a VALID Data Archive. *Language Resources and Evaluation Conference Proceedings (LREC) May 26-31, 2014, Reykjavik*. pp. 357-364.
- Kleppe, M., Hollink, L., Kemman, M., Juric, D., Beunders, H., Blom, J., ... & Houben, G. J. (2014). PoliMedia. Analysing Media Coverage of political debates by automatically generated links to Radio & Newspaper Items. http://repub.eur.nl/pub/51285/linkedup_veni2013_04.pdf
- Lefortier, D., Serdyukov, P., Romanenko, F., & de Rijke, M. (2014). Blending Vertical and Web Results. In *Advances in Information Retrieval* (pp. 184-196). Springer International Publishing.
<https://staff.science.uva.nl/m.derijke/Publications/ecir2014-fp-verticals.pdf>
- Liang, S., Ren, Z., & de Rijke, M. (2014). The impact of semantic document expansion on cluster-based fusion for microblog search. In *Advances in Information Retrieval* (pp. 493-499). Springer International Publishing.
<https://staff.fnwi.uva.nl/s.liang/uploads/Main/HomePage/ecir2014.pdf>
- Liang, S., Ren, Z., & de Rijke, M. (2014, July). Fusion helps diversification. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval* (pp. 303-312). ACM. <https://staff.science.uva.nl/m.derijke/Publications/sigir2014-fp-fusion.pdf>
- Liang, S., Ren, Z., & de Rijke, M. (2014, August). Personalized search result diversification via structured learning. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 751-760). ACM.
- Odijk, J. 'CLARIN-NL: Major Results', in *Proceedings LREC 2014, Reykjavik, May 2014*, pp 2187-2193 [pdf] [URL]
- Odijk, J. 'The CLARIN infrastructure in the Netherlands: What is it and how can you use it?', unpublished paper, Utrecht University [pdf] [URL]
- Odijk, J. 'The CLARIN infrastructure in the Netherlands: Design and Construction', unpublished paper, Utrecht University [pdf] [URL]
- Odijk, J. 'Discovering Resources in CLARIN: Problems and Suggestions for Solutions', internal CLARIN Report, Utrecht University [pdf] [URL]
- Oostdijk, N. & H. van den Heuvel, *The Evolving Infrastructure for Language Resources and the Role for Data Scientists. Proceedings of LREC 2014, Reykjavik, Iceland*. [URL]
- Pander Maat, H., Kraf, R., Van den Bosch, A., Dekker, N., Van Gompel, M., Kleijn, S., Sanders, T., and Van der Sloot, K. (2014). T-Scan: A new tool for analysing Dutch Text. *Computational Linguistics in the Netherland Journal*, 4, pp. 53-74. [PDF] [URL]
- Peetz, M. H., Meij, E., & de Rijke, M. (2014). Using temporal bursts for query modeling. *Information retrieval*, 17(1), 74-108. <http://edgar.meij.pro/wp-content/papercite-data/pdf/irj-2013-peetz.pdf>
- Ponstein, R. *Reconciling Dutch ships and sailors*. M.Sc. thesis VU University Amsterdam (forthcoming), 2014
- Postma, M., & Vossen, P. (2014). Open Source Dutch WordNet.
http://kyoto.let.vu.nl/~postma/odwn/Documentation/odwn_documentation.pdf
- Rehm, G., Uszkoreit, H., Odijk, J. E. J. M., & Zabarskaite, J. (2014). The Strategic Impact of META-NET on the Regional, National and International Level. *Proceedings of LREC 2014*
http://dspace.library.uu.nl/bitstream/handle/1874/295272/405_Paper.pdf?sequence=1

- Ren, Z., Peetz, M. H., Liang, S., van Dolen, W., & de Rijke, M. (2014, July). Hierarchical multi-label classification of social text streams. In Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval (pp. 213-222). ACM. <http://staff.science.uva.nl/~sliang/uploads/Main/Publications/sigir2014-classification.pdf>
- Reynaert, M. Synergy of Nederlab and @PhilosTEI: Diachronic and Multilingual Text-induced Corpus Clean-up. Proceedings of LREC 2014, Reykjavik, Iceland. [URL]
- Reynaert, M. (2014). TICCLops: Text-Induced Corpus Clean-up as online processing system. COLING 2014, 52. <http://www.aclweb.org/anthology/C/C14/C14-2.pdf#page=64>
- Reynaert, M. (2014, May). On OCR ground truths and OCR post-correction gold standards, tools and formats. In Proceedings of the First International Conference on Digital Access to Textual Cultural Heritage (pp. 159-166). ACM <http://www.nederlab.nl/wp/wp-content/uploads/2014/03/Datech2014-poster.pdf>
- Reynaert, M., van de Camp, M., Taalmonsters, D., & van Zaanen, M. (2014). OpenSoNaR: user-driven development of the SoNaR corpus interfaces. COLING 2014, 124. <http://www.aclweb.org/anthology/C/C14/C14-2.pdf#page=136>
- Roorda, D. (2015). The Hebrew Bible as Data: Laboratory-Sharing-Experiences. arXiv preprint arXiv:1501.01866. <http://arxiv.org/pdf/1501.01866>
- Roorda, D., Kalkman, G., Naaijer, M., & van Cranenburgh, A. (2014). LAF-Fabric: a data analysis tool for Linguistic Annotation Framework with an application to the Hebrew Bible. arXiv preprint arXiv:1410.0286. <http://arxiv.org/pdf/1410.0286>
- Sanders, E. et al. The Dutch LESLLA Corpus, Proceedings of LREC 2014, Reykjavik, Iceland. [URL]
- Schuth, A., Sietsma, F., Whiteson, S., & de Rijke, M. (2014). Optimizing Base Rankers Using Clicks. In Advances in Information Retrieval (pp. 75-87). Springer International Publishing. <http://staff.science.uva.nl/~mdr/Publications/ecir2014-fp-bm25.pdf>
- Severyn, A., Moschitti, A., Tsagkias, M., Berendsen, R., & de Rijke, M. (2014, July). A syntax-aware re-ranker for microblog retrieval. In Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval (pp. 1067-1070). ACM. <http://ilps.science.uva.nl/sites/ilps.science.uva.nl/files/sigir2014-sp-microblog.pdf>
- Sijs, V. D. N. (2014). De voortzetting van de historische taalkunde met andere middelen. http://depot.knaw.nl/14952/1/2014_Sijs_Voortzetting_historische_taalkunde_oratie.pdf
- van Sluijs, R. (2014). What's Past Is Past: Variation in the Expression of Past Time Reference in Negerhollands Narratives. *Journal of Germanic Linguistics*, 26(03), 272-321.
- Stronks, E. (2015). 'Dees kennisse zuldy te kope vinnen'. Liedcultuur en de waarde van 'know how' in de vroegmoderne Republiek. *De Zeventiende Eeuw. Cultuur in de Nederlanden in interdisciplinair perspectief*, 30(2), 147-167. <http://www.de-zeventiende-eeuw.nl/index.php/dze/article/view/9932/10409>
- Voskarides, N., Odijk, D., Tsagkias, M., Weerkamp, W., & de Rijke, M. (2014). Query-dependent contextualization of streaming data. In Advances in Information Retrieval (pp. 706-712). Springer International Publishing. <http://wouter.weerkamp.com/downloads/ecir2014-query-dependent-contextualization.pdf>
- Wright, S. E., Windhouwer, M., Schuurman, I., & Broeder, D. (2014). Segueing from a Data Category Registry to a Data Concept Registry. In Terminology and Knowledge Engineering 2014 (pp. 11-p). <https://hal.archives-ouvertes.fr/hal-01005840/document>
- Zeldenrust, D. A. (2014). Access to data, The Soundbites collection of the Meertens Institute and a flexible approach to the curation and dissemination of humanities digital resources. In Digital Humanities Benelux Conference.
- Zeldenrust, D. A. (2014). The Exploration and Visualisation of Big Data in the Humanities, Comparing Data Representation of a Large-scale E-infrastructure (CLARIN) and a Dedicated Virtual



Research Environment (Soundbites). In European Social Science History Conference 2014 (ESSHC).

Zoghi, M., Whiteson, S. A., de Rijke, M., & Munos, R. (2014, February). Relative confidence sampling for efficient on-line ranker evaluation. In Proceedings of the 7th ACM international conference on Web search and data mining (pp. 73-82). ACM.
<https://staff.fnwi.uva.nl/s.a.whiteson/pubs/zoghiwdsdm14.pdf>

Presentations

Here's a (not fully complete) overview:

Augustinus, L., 'NEDERBOOMS: Treebank Mining for Data-based Linguistics', Lecture at the LOT Summerschool Course "CLARIN for Linguists", Nijmegen, Jun 24, 2014 [[pdf](#)]

Augustinus, L. & F. van Eynde: *Dat we ons daar nog mee kunnen bezig houden. Looking for cluster creepers in Dutch treebanks*. Presentation at CLIN 2014, Leiden, Jan 17, 2014.

Augustinus, L., 'Hands-on session GrETEL', Lecture at the LOT Summerschool Course "CLARIN for Linguists", Nijmegen, Jun 24, 2014 [[pdf](#)]

Bennis, H. 'Digitale geesteswetenschappen zijn niet te stoppen', presentation at [NWO Bessensap](#), Jun 6, 2014 [[abstract](#)] [[pptx](#)]

Bloem, J. *et al.* Using automatically annotated corpora in language variation research. Presentation at the CLARIN Annual Conference 2014, Soesterberg, Netherlands, 23-25 Oct.

Bloem, J. *et al.* Using automatically annotated corpora in language variation research. Demo at the CLARIN Annual Conference 2014, Soesterberg, Netherlands, 23-25 Oct.

Blumtritt, J. *et al.* User Delegation in the CLARIN Infrastructure. Poster at the CLARIN Annual Conference 2014, Soesterberg, Netherlands, 23-25 Oct.

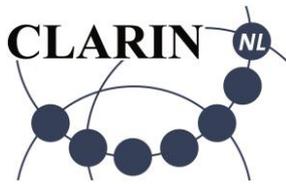
Broeder, D & J. Odijk 'CLARIN Centers for a Sustainable Infrastructure'. Presentation at the CLARIN Annual Conference 2014, Soesterberg, Netherlands, 23-25 Oct. ([URL](#))

Does, J. de *et al.* *Named Entity Recognition and Resolution for Literary Studies*. Presentation at CLIN 2014, Leiden, Jan 17, 2014.

Goosen, T. *et al.* *CMDI 1.2: Improvements in the CLARIN Component Metadata Infrastructure*. Poster at the CLARIN Annual Conference 2014, Soesterberg, Netherlands, 23-25 Oct.

Goosen T. & T. Eckart. *Virtual Language Observatory 3.0: What'sNew?*. Poster at the CLARIN Annual Conference 2014, Soesterberg, Netherlands, 23-25 Oct.

Heuvel, H. van *den et al.* Data curations by the Dutch Data Curation Service. Poster at the CLARIN Annual Conference 2014, Soesterberg, Netherlands, 23-25 Oct.



Hoogervorst, Tom G. 'Heritage through digitisation', Talk presented at the *ASEASUK 14th update*, 12-14 September 2014, Brighton, United Kingdom [[pdf](#)]

Kemman, M. & M. Kleppe. Too Many Varied User Requirements for Digital Humanities Projects. Poster at the CLARIN Annual Conference 2014, Soesterberg, Netherlands, 23-25 Oct.

Kemps-Snijders, **Marc**, 'TTNWW: TST tools voor het Nederlands als Webservices in een Workflow', Lecture at the LOT Summerschool Course "CLARIN for Linguists", Nijmegen, Jun 25, 2014 [[pdf](#)]

Klatter, J. (2014). **Presentation** on VALID at Conference *E-Health voor mensen met communicatieve beperkingen: wetenschap en praktijk*, Nijmegen, October 3, 2013.

Klatter, J. (2014). **Vulnerability** in Acquisition, Language Impairments in Dutch: Creating a VALID Data Archive, presentation at *Language Resources and Evaluation Conference (LREC)*, Reykjavik, May 26-31, 2014.

Maks, I., Cornetto and WordNet: network of relations based on the words of a language , Lecture at the LOT Summerschool Course "CLARIN for Linguists", Nijmegen, Jun 26, 2014 [[pdf](#)]

Niestadt, J.& J. De Does Shooting ducks in a pond: hunting for patterns with BlackLab. Presentation at CLIN 2014, Leiden, Jan 17, 2014.

Odijk, J. 'The CLARIN Infrastructure / GrETEL', MA-Rotation lecture, Utrecht, Feb 13, 2014. [[pptx](#)]

Odijk, J. 'CLARIN: What's in it for Linguists?', Uilendag lecture, Utrecht, Mar 27, 2014. [[pptx](#)]

Odijk, J. 'Searching for Constructions with GrETEL, Syntax Interface Meeting, Mar 31, 2014 [[pptx](#)]

Odijk, J. 'CLARIN-NL: Major Results', in Poster presentation at LREC 2014, Reykjavik, May 29, 2014, [[pdf](#)]

Odijk, J. 'CLARIN for Linguists: Introduction', lecture at the LOT Summerschool Course "CLARIN for Linguists", Nijmegen, Jun 23, 2014 [[pdf](#)]

Odijk, J., 'CLARIN for Linguists: Portal and Searching for Data', lecture at the LOT Summerschool Course "CLARIN for Linguists", Nijmegen, Jun 23, 2014 [[pdf](#)]

Odijk, J., 'CLARIN for Linguists: Search Illustration 1', lecture at the LOT Summerschool Course "CLARIN for Linguists", Nijmegen, Jun 23, 2014 [[pdf](#)]

Odijk, J. 'CLARIN for Linguists: Storing resources in CLARIN', lecture at the LOT Summerschool Course "CLARIN for Linguists", Nijmegen, Jun 27, 2014 [[pdf](#)]

Odijk, J. 'CLARIN for Linguists: Concluding Overview', lecture at the LOT Summerschool Course "CLARIN for Linguists", Nijmegen, Jun 27, 2014 [[pdf](#)]



Pieters, T. *et al.* Eugenics revisited in hidden debates by means of multilingual semantic text-mining with the Biland demonstrator. Poster at the CLARIN Annual Conference 2014 Soesterberg, Netherlands, 23-25 Oct.

Reynaert, M. Diachronic TICCL or Text-Induced Corpus Clean-up in Nederlab. Presentation at CLIN 2014, Leiden, Jan 17, 2014.

Roorda, R. et al *Workbench for Data analysis in the Hebrew Bible*. Presentation at CLIN 2014, Leiden, Jan 17, 2014.

Scholman. M. & T. Sanders. Annotating coherence relations in corpora of language use. Poster at the CLARIN Annual Conference 2014 Soesterberg, Netherlands, 23-25 Oct.

Schuurman, I. et al. *The linguist is always right! Oops ... Our experiences with User Interface Testing*. Presentation at CLIN 2014, Leiden, Jan 17, 2014.

Van Eynde, F., 'Number agreement in copular constructions. A treebank-grounded investigation', Lecture at the LOT Summerschool Course "CLARIN for Linguists", Nijmegen, Jun 24, 2014 [\[pdf\]](#)

Windhouwer, M. & Schuurman: *ISOcat and RELcat, two cooperating semantic registries*. Presentation at CLIN 2014, Leiden, Jan 17, 2014.

Zaanen, M. van et al.: *Online Personal Exploration and Navigation of SoNaR*. Presentation at CLIN 2014, Leiden, Jan 17, 2014.

Support for other events

From	To	Event	City	Form of Support
14-jan-14	15-jan-14	THATCamp	Den Haag	financial
17-jan-14	17-jan-14	CLIN	Leiden	financial

Support for travel of Dutch researchers to CLARIN-related events

Date	Description	City	Country	Who	Organization
26-may-14	LREC	Reykjavik	IS	Han Sloetjes	MPI
20-jul-14	XXIVth Congress of The International Arthurian Society	Bukarest	RO	Bart Besamusca	UU



11-aug-14	Methods in Dialectology XV	Groningen	NL	Terese Leinonen	
-----------	--	-----------	----	-----------------	--

CLARIN- NL Newsflashes

CLARIN- NL Newsflashes have been sent out to all registered persons on the following dates:

- [NewsFlash 30 \(23-05-2014\)](#)
- [NewsFlash 29 \(28-01-2014\)](#)

All Newsflashes until 31-dec-2014 can be accessed via <http://www.clarin.nl/node/82>.

CLARIN-EU Newsflashes

In 2014, no CLARIN (ERIC) newsflashes have appeared.

CLARIN-EU Newsletter

In 2014 no CLARIN (ERIC) Newsletters have appeared.

eData & Research

CLARIN-NL passed on its participation in the eData & Research Magazine in 2014 to the CLARIAH-SEED project. Several articles on CLARIN-NL or related topics have appeared in [eData&Research](#). On behalf of CLARIAH-SEED, Erica Renckens participates in the editorial team and Arjan van Hessen in the board.

The articles that have been published in eData & Research in 2014 can be accessed via <http://www.clarin.nl/node/188>.

Movies and Screen Captures

Several projects have made movies or screen captures to illustrate their work or their demonstrators. Rembench also added such a short movie on its help page:

Project	Link to Movie / Screen capture
GabMap	http://www.youtube.com/watch?v=anElatt5fFM&feature=plcp
Interviews / Veteran Tapes	http://www.youtube.com/watch?v=VpCXh_NVVso
WFT-GTB	http://www.youtube.com/watch?v=fevAPknCaDo&feature=player_embedded
RemBench	http://rembench.huygens.knaw.nl/help

[OpenSONAR](#) and [@PhilosTEI](#) each has a guided tour through the application via their Help button



The EB commissioned the production of 9 short movies on CLARIN-NL projects (to Dyzlo). Below one can find the CLARIN-NL projects that have been selected for these movies. All movies are available.

All movies and screen captures can be reached via <http://www.clarin.nl/node/403>

Project	Link to Movie
MIMORE	http://vimeo.com/35305355 http://www.youtube.com/watch?v=TBDGrqKYbIA
SignLinc	http://vimeo.com/35305490 http://www.youtube.com/watch?v=je3ZKPw6w5s
Arthurian Fiction	http://vimeo.com/35305561 http://www.youtube.com/watch?v=t0y-ecrVD5w
WFT-GTB	http://vimeo.com/39044483 http://www.youtube.com/watch?v=aSWKJucbWY8
WIP	http://vimeo.com/48195599 http://www.youtube.com/watch?v=tEqAIH2o2rc&feature=plcp
Search & Develop	http://vimeo.com/40289489 http://www.youtube.com/watch?v=6CkCfaj7rao&feature=plcp
CKCC	http://vimeo.com/68203559 http://www.youtube.com/watch?v=SI0lg7SCOcc
PILNAR	http://vimeo.com/100788014 https://www.youtube.com/watch?v=iMWhO3D_n_Q
Polimedia	http://vimeo.com/100787499 https://www.youtube.com/watch?v=u24oRCj7xrQ

The movies have been made available in streaming mode via Vimeo and YouTube, have subtitles (at least in English) and the originals have been made available on solid state memory.

Website

The website is maintained by an external company (*Civinex*).

The CLARIN-NL website is a project-internal website. For aspects that go beyond the project, we will use the CLARIN-NL portal, a preliminary version of which has gone live in June 2014 and which will be officially launched at the CLARIN-NL final event.

Education

Six educational modules were developed and are available via the [CLARIN-NL Portal](#).

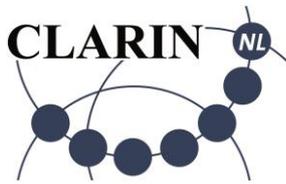
Several other pieces of educational material are available there, including a recorded lecture (by JO) on searching for grammatical constructions with GrEtel

Various educational activities have been organized in 2014. The following table lists the most important ones:



From	To	Event	City	Ctry	Who
13-feb-14	13-feb-14	'The CLARIN Infrastructure / GrE TEL', MA Rotation Lecture	Utrecht	NL	JO
4-mrt-14	4-mrt-14	CMDI Tutorial	Utrecht	NL	DB
31-mrt-14	31-mrt-14	Searching for Constructions with GrE TEL, Syntax Interface Meeting	Utrecht	NL	JO
23-apr-14	23-apr-14	Your Examples in GrE TEL I	Utrecht	NL	JO
20-may-14	20-may-14	Your Examples in GrE TEL II	Utrecht	NL	JO
23-jun-14	27-jun-14	LOT Summerschool CLARIN Course	Nijmegen	NL	JO + guest teachers
12-aug-14	12-aug-14	Tutorial Gabmap	Groningen	NL	Terese Leinonen & Wilbert Heeringa
14-oct-14	14-oct-14	CLARIN and GrE TEL: Lecture for the UU Digital Humanities Batchelor Course	Utrecht	NL	JO
17-dec-14	20-jan-15	Corpus Analysis Course	Utrecht	NL	JO

A winter school course is planned for January 2015 (LOT winter school in Amsterdam), as well as plenary sessions at TIN 2015 and CLIN 2015. Some funds were reserved for these seasonal schools but not spent (some relevant but very small travel costs were booked on the general travel budget). Through these activities, many of the relevant researchers, especially linguistics researchers, have been reached.



Helpdesk

The CLARIN-NL Helpdesk¹⁹ has run all year. The amount of requests is rather low, and therefore the Helpdesk was easily managed in 2014 by Jan Odijk and Daan Broeder. There are more frequent informal requests directly to these persons from researchers who inquire about CLARIN and its use.

It was suggested to develop a CLARIN-NL App for mobile devices, and some money was reserved for it. A CLARIN-NL App might have good PR value but whether it would have any actual usage value is doubtful. Because of lack of resources and lack of ideas for such an App, no work was done on this (and no money spent on it).

1.2.17 National Coordination Point

The national coordination point has functioned well in 2013. Prof.dr. Jan Odijk (1FTE) is the Programme Director. Drs. Arwin van der Zwan was the project secretary (partially .3 FTE, partially .4 FTE).

The governance structure consists of an Executive Board (EB), Board, National Advisory Panel (NAP) and International Advisory Panel (IAP). Their tasks and responsibilities have been laid down in the consortium agreement.

Executive Board The Executive Board consists of 4 members:

- Prof.dr. Jan Odijk, programme director
- Prof.dr. Hans Bennis, chair and representative of the humanities researchers
- Dr. Arjan van Hessen, responsible for complementary activities such as PR, education, training and awareness creation.
- Ir. Daan Broeder, technical director

Apart for the programme director (1FTE), the members of the EB have a 1/5FTE appointment to carry out their tasks.

The executive board has held regular meetings (once every two weeks) to formulate, prepare and implement policy procedures and concrete activities.

Board The Board consists of 8 members. A full list of its members can be found in Appendix A.

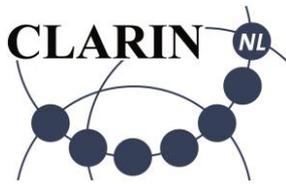
The current composition of the Board can also be found on the CLARIN-NL website (<http://www.clarin.nl/category/board/board>)

The Board has held 1 meeting in 2014 (April 15, 2014).

NAP The National Advisory Panel consists of 16 members, providing a good representation of the fields of the humanities, language and speech processing and infrastructure design and implementation. A full list of the members can be found in Appendix B

The current composition of the CLARIN-NL NAP can also be found on the CLARIN-NL website (<http://www.clarin.nl/category/board/national-advisory-panel>).

¹⁹ <http://trac.clarin.nl/trac>



The NAP has held no meetings in 2014, since it was not possible to get enough members together. A concluding meeting will be held early 2015.

IAP The International Advisory Panel consists of 11 members. A full list of the members can be found in Appendix C and on the CLARIN-NL website (<http://www.clarin.nl/category/board/international-advisory-panel>).

The IAP has held no meetings in 2014.

Consortium Agreement The consortium agreement, which was approved by the Board on Nov 2, 2009 and which was initially signed by 21 partners, has remained unchanged. The consortium is open to new partners, and several partners have joined the consortium in 2012, bringing the total number of partners to 33 (see <http://www.clarin.nl/node/7>). The consortium agreement can be found on the CLARIN-NL website (login required): <http://www.clarin.nl/node/72>.

CLARIN-NL Long Term Working Plan The CLARIN-NL Long Term Working Plan was approved by the Board on August 19, 2009, and has remained unchanged. It can be found on the CLARIN-NL website: <http://www.clarin.nl/system/files/CLARIN-NL%20Multiyear%20Programme%20090409-2.pdf>

Working Plan 2014 The Working Plan for 2014 was approved by the Board. It can be found on the CLARIN-NL website (<http://www.clarin.nl/page/about/47>)

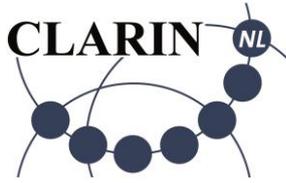
Overall, the governance structure as set up functions appropriately.

4 2014 Finances

A detailed report on the CLARIN-NL finances in 2013 can be found in the 2013 Financial Report.

5 Other Activities

As a response to the invitation of the *Commissie voor de update van de Nationale Roadmap voor Grootchalige Onderzoeksfaciliteiten* (Committee for the update of the National Roadmap for Large Scale Research Facilities), CLARIN-NL and the Dutch players in DARIAH have joined forces and formulated a new proposal for a large scale national infrastructural project called *Common Lab Research Infrastructure for the Arts and the Humanities (CLARIAH)*, uniting all humanities organizations in the Netherlands. The project has been put on the Roadmap and has received 1 million euro as 'seed capital', to maintain the dynamics of the consortium and to prepare a new, improved, proposal. The CLARIAH-SEED project has developed plans and executed all of them in 2014. On October 1, 2013, a revised proposal was submitted, which has been positively evaluated. Details about it can be found on the CLARIAH website (www.clariah.nl). The new project, called CLARIAH-CORE, will start on January 1, 2015 and will run for four years.

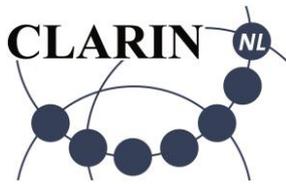


Appendix A Composition of the Board

Name	Affiliation
Prof.dr. Geert Booij (until Sep 1, 2012)	LUCL, Universiteit van Leiden
Prof.dr. Lou Boves	CLST, Radboud Universiteit Nijmegen
Dr. Peter Doorn	DANS
Prof.dr. Martin Everaert	UIL-OTS, Universiteit van Utrecht
Prof.dr. Jaap van den Herik	TiCC, Universiteit van Tilburg
Prof.dr. Aafke Hulk (until Sep 1, 2012)	ACLCL, Universiteit van Amsterdam
Prof.dr.ir John Nerbonne	CLCG, Universiteit van Groningen
Prof.dr. Els Stronks (as of Sep 1, 2012)	Universiteit van Utrecht
Ir. Peter Wittenburg (as of Sep 1, 2012)	Max Planck Instituut voor Psycholinguïstiek, Nijmegen

Appendix B Composition of the NAP

Name	Affiliation
Prof.dr. Willem Adelaar (until Sep 1, 2012)	LUCL, Universiteit van Leiden
Prof.dr. Sjef Barbiers	Meertens Instituut
Dr. Jeannine Beeken	INL
Prof.dr Antal van den Bosch	Radboud Universiteit Nijmegen
Dr. Gosse Bouma	Universiteit van Groningen
Drs. Hennie Brugman	Meertens Instituut
Prof.dr. Karina van Dalen-Oskam	Huygens Instituut voor Nederlandse Geschiedenis
Dr. Paul Doorenbosch	Koninklijke Bibliotheek
Dr. Willemijn Heeren	Universiteit van Twente
Ir. Marc Kemps-Snijders	Meertens Instituut
Prof.dr. Kees Mandemakers	Internationaal Instituut voor Sociale Geschiedenis
Dr. Nelleke Oostdijk	CLST, Radboud Universiteit Nijmegen
Prof.dr. Reinier Salverda	Fryske Akademy
Prof.dr. Ted Sanders (until Sep 1, 2012)	UiL-OTS, Universiteit Utrecht
Prof.dr. Piek Vossen	Vrije Universiteit Amsterdam
Prof.dr. Sally Wyatt (until Sep 1, 2012)	eHumanities (KNAW)
Dr. Joris van Zundert	Huygens Instituut voor Nederlandse Geschiedenis
Prof.dr. Arjen Versloot (as of Sep 1, 2012)	Universiteit van Amsterdam
Prof.dr. Toine Pieters (as of Sep 1, 2012)	Universiteit van Utrecht

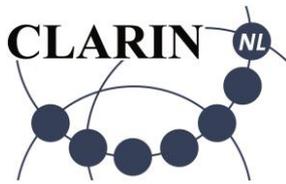


Appendix C Composition of the IAP

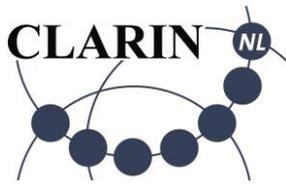
Name	Affiliation	Country
Prof.dr Scott Farrar	University of Washington, Department of Linguistics	USA
Prof.dr David Hoover	New York University	USA
Prof.dr. Clifford Lynch	Coalition for Networked Information	USA
Prof.dr. Harold Short	King's College London, Centre for Computing in the Humanities	UK
Prof.dr. Benjamin T'sou	City University of Hong Kong	China
Prof.dr Hugo Van hamme	Katholieke Universiteit Leuven	Belgium
Prof.dr. Tamás Váradi	Linguistics Institute, Hungarian Academy of Sciences	Hungary
Prof.dr. Koenraad De Smedt	University of Bergen	Norway
Prof.dr. Erhard Hinrichs	University of Tübingen	Germany
Prof.dr. Jan Hajič	Charles University Prague	Czech republic

Appendix D Acronyms

Acronym / Term (URL)	Type	Expansion	Dutch expansion
@PhilosTEI	CLARIN-NL subproject	TICCLing Philosophy: a TEI corpus-building workflow towards a new computational methodology for philosophy	
AAI	software	Authentication and Authorisation Infrastructure	
AAM-LR	CLARIN-NL subproject	Automatic Annotation of Multi-modal Language Resources	
ACL	organisation	Amsterdam Center for Language and Communication	
Adelheid	CLARIN-NL subproject	A Distributed Lemmatizer for Historical Dutch	
ADEPT	CLARIN-NL subproject	Assaying Differences via Edit-Distance of Pronunciation Transcriptions	
AGORA	project		
Alfalab	project		
ALTO	format	Analyzed Layout and Text Object	
Anéla	organisation	Dutch Association for Applied Linguistics	
ANNEX	software	Annotation Explorer	
ANP	company	National Netherlands News Agency	Algemeen Nederlands Persbureau
ANP radio bulletin collection	data	ANP radio bulletin	



		collection	
ANWB	organisation	Dutch automobile association	Algemene Nederlandse Wielrijdersbond
API	term	Application Programmer Interface	
ARBIL	software		
Arthurian Fiction	CLARIN-NL subproject	Arthurian Fiction in Medieval Europe	
Asymenc	project	Asymmetrical Encounters: Digital Humanities Approaches to Reference Cultures in Europe, 1815–1992; HERA Project	
AutoSearch	CLARIN-NL subproject		
AV	term	Audio-Visual	
AVAtch	project	Advancing Video Audio Technology in Humanities Research	
AXES	project	Access to AudioVisual Archives	
BG	organisation	Netherlands Institute for Sound and Vision	Instituut voor Beeld en Geluid
BIGGrid	project	Dutch eScience Grid	
BILAND	CLARIN-NL subproject	Towards a flexible and stable CLARIN-supported web-application for bilingual historical analysis of discourses in news media	
BioPort	data	Biographic Portal	Biografisch Portaal



Birth of TV	project	Birth Television Archive	
BlackLab	software	Backend search engine	
BLARK	term	Basic Language Resource Kit	
BRIDGE	project	Building Rich Links to Enable Television History Research	
BSON	software	Binary JSON	
CATCH	project	NWO Programme Continuous Access To Cultural Heritage	
CATCHPlus	project		
CCR	Concept registry	CLARIN Concept Registry	
CDH	organisation	Centre for Digital Humanities	
C-DSD	CLARIN-NL subproject	Curating the Dutch Song Database	
CELEX	data	Dutch Centre for Lexical Information	
Center for Language Studies	organisation	Centre for Language Studies	
CEPR	organisation	Center for Economic policy Research	
CESSDA	infrastructure	Council of European Social Science Data Archives	
CGN	data	Spoken Dutch Corpus	Corpus Gesproken Nederlands
CHAT	organisation	Centre for Humanities and Technology	
CHILDES	infrastructure	Child Language Data Exchange System	
CHIP	project	Project in the CATCH programme	



CHOICE	project	Project in the CATCH programme	
CineXPRES	project		
CIO	term	Chief Integration Officer	
CKCC	project	<i>Circulation of Knowledge and Learned Practices in the 17th-century Dutch Republic,</i>	Geleerdenbrievenproject
CLAM	software	Computational Linguistics Application Mediator	
CLAPOP	CLARIN-NL subproject	CLARIN-NL Portal Project	
CLARA	project	Common Language Resources and their Applications	
CLARIAH	infrastructure	Common Lab Infrastructure for the Humanities	
CLARIAH-AT	infrastructure project	Austrian CLARIAH project	
CLARIAH-CORE	Infrastructure project	Dutch main CLARIAH project	
CLARIAH-SEED	infrastructure project	CLARIAH Seed Capital project	
CLARIN	infrastructure	The CLARIN infrastructure	
CLARIN ERIC	organisation	CLARIN ERIC	
CLARIN-EU	infrastructure project	The European CLARIN enterprise	
CLARIN-NL	infrastructure project	The Netherlands national project for CLARIN	
CLARIN-PP	infrastructure project	The European CLARIN preparatory phase project	
CLARIN-prep	infrastructure project	The European CLARIN preparatory	



		phase project	
CLARINO	Infrastructure project	Norwegian CLARIN infrastructure project	
CLAVAS	CLARIN-NL subproject	CLARIN Vocabulary Service	
CLAVAS	software	CLARIN Vocabulary Service	
CLCG	organisation	Center for Language and Cognition Groningen	
CLICKNL	organisation	Creativity, Learning, Innovation, Co-creation, Knowledge	
CLIF	organisation	Computational Linguistics in Flanders	
CLIN	conference	Computational Linguistics in the Netherlands (local conference)	
Clio Infra	infrastructure		
CLIO-DAP	project		
CLS	organisation	Centre for Language Studies	
CLST	organisation	Centre for Language and Speech Technology	
CLUE	organisation	Research institute for the heritage and history of the Cultural Landscape and Urban Environment	
CMDI	software	Component MetaData Infrastructure	



CMDI2RDF	CLARIN-NL subproject	Project to convert CMDI to RDF linked data	
COAVA	CLARIN-NL subproject	Cognition, Acquisition and Variation Tool	
COBWWWEB	CLARIN-NL subproject	Connections Between Women and Writings Within European Borders	
COGIS	organisation	Dutch expert centre on the (psycho)social effects of war, persecution, aggression and violence	Kennisinstituut sociale en psychische gevolgen van oorlog, vervolging en geweld
ColTime	CLARIN-NL subproject	Collaboration on time-based resources	
Comerda	software		
COMMIT	project		
COREX	software	Corpus Exploitation Tool	
Cornetto	data		
CORNETTO-LMF-RDF	CLARIN-NL subproject	Curated Cornetto database in LMF and RDF	
CORNETTO II	CLARIN-NL subproject	Successor of CORNETTO-LMF-RDF	
CPG	organisation	Centre for Parliamentary History	Centrum voor Parlementaire Geschiedenis
CTO	term	Chief Technical Officer	
Culturele Dynamiek	grant programme	NWO grant programme: Cultural Dynamics	
D3	software	Data Driven Documents	
DAM-LR	project	Distributed Access Management for Language	



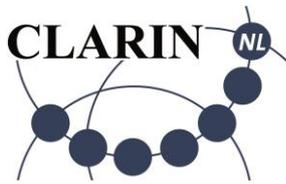
		Resources	
DANS	organisation	Data Archiving and Networking Services	
DARIAH	infrastructure	Digital Research Infrastructure for the Arts and Humanities	
DARIAH-EU	infrastructure project	The European DARIAH enterprise	
DARIAH-NL	infrastructure project	DARIAH in the Netherlands	
DARIAH-PP	infrastructure project	the DARIAH Preparatory Phase Project	
DASISH	infrastructure project	Data Service Infrastructure for the Social Sciences and the Humanities	
DBD	data	Dutch Bilingual Database	
DBNL	organisation	Digital Library for Dutch Literature	
D-COI	data		
DCR	data	Data Category Registry	
DCS	CLARIN-NL subproject	Data Curation Service	
DDS	software	Data Deposition Service	
DEN	organisation	Digital Heritage Netherlands	Kenniscentrum Digitaal Erfgoed
DH	term	Digital Humanities	
DiDDD	data	Diversity in Dutch DP Design	
DiscAn	CLARIN-NL subproject	Towards a Discourse Annotation system for Dutch language corpora	
DLU	organisation	Dutch Language Union	

D-LUCEA	CLARIN-NL subproject	Database of the Longitudinal Utrecht Collection of English Accents	
DOBES	project	Documentation of Endangered Languages	
DSD	data	Dutch Song Database	Liederbank
DOI	identifier	Digital Object Identifier	
DSS	CLARIN-NL subproject	Dutch Ships and Seamen	
DUELME	data	Dutch Electronic Lexicon of Multiword Expressions	
DUELME-LMF	CLARIN-NL subproject	Converting DUELME into LMF format	
Dynamics of Memory	grant programme	NWO grant programme: Dynamics of Memory	
DynaSAND	software	Dynamic Syntactic Atlas of the Dutch Dialects	Dynamische Syntactische Atlas van de Nederlandse Dialecten
EAGLES	project	Expert Advisory Group on Language Engineering Standards	
EB	term	Executive Board	
e-BNM	data	Bibliotheca Neerlandica Manuscripta database	
eBNM+	CLARIN-NL subproject	Linked Data on Middle Dutch Sources Kept Worldwide	
EC	organisation	European Commission	
EDBO	data	Early Dutch Books Online	

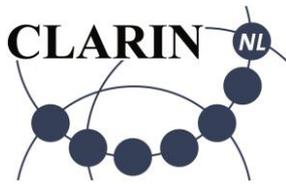
EDBO	data	Early Dutch Books Online	
eHg	organisation	eHumanities Group	
EHPS-Net	network	European Historical Population Samples Network	
EHRI	infrastructure project	European Holocaust Research Infrastructure	
eHumanities Group	organisation	KNAW eHumanities Group	
EIP	instrument	European Innovation Partnerships	
ELAN	software	EUDICO Linguistic Annotator (Annotation Tool)	
ElasticSearch	software	open source, distributed real-time search and analytics engine for the cloud	
EMIT-X	CLARIN-NL subproject	Early-Modern Image and Text eXchange	
EPIC	consortium	European Persistent Identifier Consortium	
ePistolarium	software		
ePub	Data format	Electronic Publication	
ERC	organisation	European Research Council	
ERIC	organisation	European Research Infrastructure Consortium	
ESF EURYI	grant programme	European Science Foundation European Young Investigator Awards	



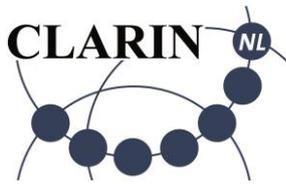
ESFRI	organisation	European Strategy Forum on Research Infrastructures	
EU	organisation	European Union	
EUDAT	project	European Data Infrastructure	
EUR	organisation	Erasmus University Rotterdam	
European Film Gateway	organisation		
Europeana	data		
EURYI	grant programme	European Science Foundation European Young Investigator Awards	
EUScreen	project		
EXILSEA	CLARIN-NL subproject	Exploiting ISOcat's Language Sections in ELAN and ANNEX	
eXist	software	Native XML database	
EYE	organisation	EYE Film Museum Amsterdam	
FA	organisation	Fryske Akademy	
FESLI	CLARIN-NL subproject	Functional Elements in Specific Language Impairment	
FIDLR	project	Framework for the Interoperability of Dutch Language Resources	
FIDLR-Start	project	Project to prepare the FIDLR project proposal	
FlareNet	project	Fostering Language Resources	



		Network, European Network project.	
FlockDB	software	Distributed Fault-Tolerant Graph Database	
FOLiA	Data format	Format for Linguistic Annotation	
FP7	programme	Framework Programme 7 (European Union)	
Frontwise	company		
FTE	term	Full-Time Equivalent	
Gabmap	software	Application resulting from the CLARIN-NL ADEPT subproject	
GIA	organisation	Groningen Institute of Archaeology	
GIS	term	Geographic Information System	
Global labour	project	Global Collaboratory on the History of Labour Relations 1500-2000	
Gridline	company		
GrNe	CLARIN-NL subproject	Greek-Dutch dictionary	Grieks-Nederlands woordenboek
GTB	software	Integrated Language Database	Geïntegreerde Taalbank
GTRP	project	Goeman, Taeldeman, van Reenen Project	
HERA	organisation	Humanities in the European Research Area	
HI	organisation	Huygens ING Institute	
HI-POD	project	Historical Patterns of	



		Development	
HLT	term	Human Language Technologies	
HLZ	project	HSN LINKS Zeeland	
HSN	infrastructure	Historical Sample of the Netherlands	
HSS	term	Humanities and Social Sciences	
HTML	Data format	HyperText Markup Language	
Huygens ING	organisation	Huygens ING Institute	
HvA	organisation	Hogeschool van Amsterdam	
IAP	term	International Advisory Panel	
IARPA SESAME Project	project	Video SEarch with Speed and Accuracy for Multimedia Events	
IBM	company	International Business Machines	
IBM NL	company	IBM Netherlands	
ICON UU	organisation	Institute for Cultural Inquiry Utrecht University	
ICORDI	project		
ICPSR	organisation	Interuniversity Consortium for Political and Social Research	
ICT	term	Information and Communication Technology	
ICT Roadmap for the Top Sectors	roadmap		
ID	term	Identifier	
IIAV	organisation	International Information Centre and Archive for the Women's	Internationaal Informatiecentrum en Archief voor de Vrouwenbeweging



		Movement	
IIP	CLARIN-NL subproject	Infrastructure Implementation Project	
IISG	organisation	International Institute for Social History	Internationaal Instituut voor Sociale Geschiedenis
IISH	organisation	International Institute for Social History	Internationaal Instituut voor Sociale Geschiedenis
Images for the Future	project		Beelden voor de Toekomst
IMDI	software	ISLE MetaData Initiative	
IMIX	project	NWO programme for Interactive Multimodal Information eXtraction	
iMmovator	network	Cross media Network	
INL	organisation	Institute for Dutch Lexicology	Instituut voor Nederlandse Lexicologie
INPOLDER	CLARIN-NL subproject	Integrated Parser and Lemmatizer Dutch in Retrospect	
Institute for Historical Literature and Cultural Studies	organisation		
INTERA	project	Integrated European language data Repository Area	
Interop	project	NSF Community-Based Data Interoperability Networks Programme	
INTER-VIEWS	CLARIN-NL subproject	Curation of Interview Data	
IOP	programme	Innovation-targeted Research Programmes	Innovatiegerichte Onderzoekprogramma's

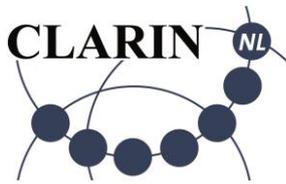
IPNV	data	Interview Project Nederlandse Veteranen	Interview Project Dutch Veterans
IPR	term	Intellectual Property Rights	
IPROSLA	CLARIN-NL subproject	Integrating and publishing resources on sign language acquisition	
ISLA	organisation	Intelligent Systems Lab Amsterdam	
ISLE	project	International Standards for Language Engineering	
ISO	organisation	International Standards Organisation	
ISO TC37/SC4	standard	ISO Subcommittee focusing on Language Resource Management	
ISOcat	data	Project to implement the ISO DCR standard	
ISOcat	software	ISO Data Category Registry	
JSON	software	JavaScript Object Notation	
KB	organisation	National Library of the Netherlands	
KB Newspaper Archive	data	National Library Newspaper Archive	
KDC	organisation	Catholic Documentation Centre	Katholiek Documentatie Centrum
KITLV	organisation	Royal Netherlands Institute of Southeast Asian and Caribbean	Koninklijk Instituut voor Taal-, Land- en Volkenkunde

		Studies	
KNAW	organisation	Royal Netherlands Academy of Arts and Sciences	
KNAW Over Grenzen	KNAW grant programme	KNAW Crossing Boundaries	KNAW Over Grenzen
Knowledge Concepts	company		
LAISEANG	CLARIN-NL subproject	Language Archive of Insular South East Asia and West New Guinea	
LD	Method for publishing data	Linked Data	
LDE CGHD	organisation	Leiden Delft Erasmus Center for Global Heritage and Development	
LESLLA	data	Low Educated Second Language and Literacy Acquisition	
Leverhulme Trust	grant programme		
LINDAT	Infrastructure project	Czech CLARIN infrastructure project	
LINKS	project		
LinkTV	project		
LIRICS	project	Linguistic Infrastructure for Interoperable Resources and Systems	
LMF	standard	Lexical Markup Framework	
LMF	standard	Lexical Markup Framework	
Lorentz	organisation	Lorentz Center	
LOT	organisation	Netherlands Graduate School of Linguistics	Landelijke Onderzoeksschool Taalkunde
LREC	conference	Language	

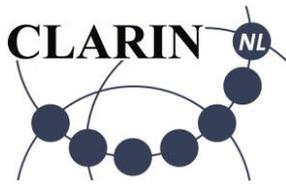


		Resources and Evaluation Conference	
LRT	term	Language Resources and Technologies	
LT	term	Language Technology	
LUCAS	organisation	Leiden University Center for the Arts in Society	
LUCL	organisation	Leiden University Centre for Linguistics	
MA	term	Master of Arts	
Maddison Project	project		
MAND	data	Morphological Atlas of the Dutch Dialects	
MCC	CLARIN-NL Subproject	Metadata Content Coordinator	
MD	CLARIN-NL subproject	Metadata Project	
MD4T	CLARIN-NL subproject	Metadata for Tools Project	
Meertens	organisation	Meertens Institute	
Meertens Institute	organisation	Meertens Institute	
Merdes	software	Media Researchers Data Exploration Suite	
META-NET	project		
META-SHARE	sharing facility		
METS	standard	Metadata Encoding and Transmission Standard	
MI	organisation	Meertens Institute	
MIGMAP	CLARIN-NL subproject	Interactive migration maps for the 20th century	
MIMORE	CLARIN-NL	Microcomparativ	

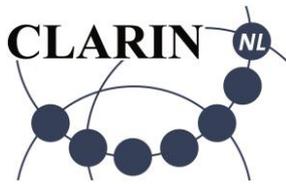
	subproject	e Morphosyntax Research Tool	
Mindisc	CLARIN-NL subproject	Minimal Set for Discourse Annotation	
MPG	organisation	Max-Planck-Gesellschaft	
MPI	organisation	Max Planck Institute for Psycholinguistics	
MT	term	Management Team	
MU	organisation	Maastricht University	
Multicon	CLARIN-NL subproject	Multilayer Concordance Functions in ELAN and ANNEX	
MuSeUM	project	Project in the CATCH programme	
NA	organisation	National Archive	Nationaal Archief
Namespace	CLARIN-NL subproject	Mapping the Landscape of Names in Modern Dutch Literature	
NAP	committee	National Advisory Panel	
Nederlab	project		
NeDiMAH	network	Network for Digital Methods in the Arts and Humanities	
NEHOL	CLARIN-NL subproject	Negerhollands Database	
NGT	data	Dutch Sign Language Corpus	Corpus van de Nederlandse Gebarentaal
NIAS	organisation	Netherlands Institute for Advanced Study in the Humanities and Social Sciences	



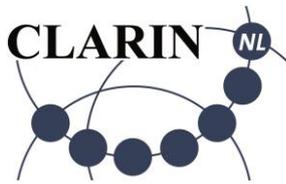
NIBG	organisation	Netherland Institute for Sound and Vision	Nederlands Instituut voor Beeld en Geluid
NIOD	organisation	Netherlands Institute for War Documentation	Nederlands Instituut voor Oorlogsdocumentatie
NISV	organisation	Netherlands Institute for Sound and Vision	Instituut voor Beeld en Geluid
NL	country	the Netherlands	
NLeSc	organisation	Netherlands eScience Centre	
NLP	term	Natural Language Processing	
Notas	organisation	Dutch Organisation for Language and Speech Technology	Nederlandse Organisatie voor Taal- en Spraaktechnologie
Noterik	company	Noterik BV	
NSF	organisation	National Science Foundation	
NWO	organisation	Netherlands Organisation for Scientific Research	
NWO Creative Industry	grant programme	NWO grant programme: Creative Industry	
NWO Open Competition	grant programme	NWO grant programme: Open Competition	NWO Open Competitie
NWO-Groot	grant programme	NWO Infrastructure Investment Fund	
OAI	organisation	Open Archives Initiative	
OAI/PMH	standard	Protocol for Metadata Harvesting	
OAuth2	software	Oauth 2.0	
OCR	term	Optical Character Recognition	
OCROPUS	software	OCROPUS	
OHT	project	Oral History	



		Today	
OLAC	standard	Open Languages Archive Community	
OpenConvert	CLARIN-NL subproject		
OpenSONAR	CLARIN-NL subproject	Online Personal Exploration and Navigation of SoNaR	
OPENSKOS	CLARIN-NL subproject	To create the CCR	
OWL	standard	OWL Web Ontology Language	
PaQu	CLARIN-NL subproject	Parse and Query	
PC	organisation	Programme Committee	
PCM	company	large publisher in the Netherlands (Perscombinatie Meulenhoff & Co)	
PDF	Document format	Portable Document Format	
Picturae	company		
PID	term	Persistent Identifier	
PILNAR	CLARIN-NL subproject	Pilgrimage Narratives: creating a germ corpus for studying the profile of the modern pilgrim	
Pionier Grant	grant programme	specific type of grant awarded by NWO in the past	
PMH	standard	Protocol for Metadata Harvesting	
Polimedia	CLARIN-NL subproject	Interlinking multimedia for the analysis of media coverage	



		of political debates	
POS	term	Part of Speech	
Posthumus Institute	organisation	National Graduate School the Posthumus Institute	Landelijke Onderzoeksschool het Posthumus Instituut
PR	term	Public Relations	
QuaMerdes	CLARIN-NL subproject	Quantitative Content Analysis of Media Researchers' Data	
R&D	term	Research and Development	
RCE	organisation	Cultural Heritage Agency	
RDA	project	Research Data Alliance	
RDF	standard	Resource Description Framework	
RDF/SKOS	standard	RDF / Simple Knowledge Organisation System	
Readspeaker	company		
Redis	software	Open Source Key-Value Store	
RELcat	software	Relation Registry	
RemBench	CLARIN-NL subproject	A Digital Workbench for Rembrandt Research	
REST	standard	Representational State Transfer	
RKD	organisation	Netherlands Institute for Art History	
RMES	organisation	Research School for Media Studies	
RU	organisation	Radboud University Nijmegen	
RUG	organisation	Groningen University	Rijksuniversiteit Groningen



RUN	organisation	Radboud University Nijmegen	
S&D	CLARIN-NL subproject	Search & Develop	
SAMPA	standard	Speech Assessment Methods Phonetic Alphabet	
SAND	data	Syntactic Atlas of the Dutch Dialects	Syntactische Atlas van de Nederlandse Dialecten
SARA	organisation	Foundation Academic Computing Centre Amsterdam	
SCHEMAcat	software	Schema Registry	
SEALINCMedia	project	Socially-enriched access to linked cultural media	
Seecr	company	Software Craftmanship	
SHEBANQ	CLARIN-NL subproject	System for HEBrew Text: ANnotation for Queries and Markup	
SignLinC	CLARIN-NL subproject	Linking lexical databases and annotated corpora of signed languages	
SKOS	standard	Simple Knowledge Organization System	
SLI	term	Specific Language Impairment	
SME	term	Small and Medium Enterprises	
SO	term	Support Office	
SOA	standard	Service Oriented Architecture	
SONAR	data		

SPARQL	software	Query Language for RDF	
SPEX	organisation	Speech Processing Expertise Centre (RU, Nijmegen)	
SPIN	grant programme	NWO Programme in the 90s	
SPINlab	organisation	Spatial Information laboratory VU	
Spinoza Grant	grant	specific type of grant awarded by NWO	
Spinque	company		
SQL	standard	Structured Query Language	
STEVIN	project	Dutch-Flemish programme for realizing the BLARK and HLT research for Dutch	
STITCH	project	Project in the CATCH programme	
SURF	organisation	SURF	
SURFnet	organisation	University Network Services Organisation	
SURFsara	organisation		
SWOT	term	Strength Weakness Opportunity Threat Analysis	
Target Holding	company		
TB	term	Technical Board	
TBDL	organisation	Tilburg Big Data Lab	Tilburg Big Data Lab
TCULT	project	Languages and Cultures in the Utrecht districts Lombok and Transvaal	Talen en Culturen in het Utrechts Lombok en Transvaal
TDS	software	Typological Database System	



TDS Curator	CLARIN-NL subproject	A web-services architecture to curate the Typological Database System	
Teezir	company		
TEI	Standard text format	Text Encoding Initiative	
Telecats	company		
TERENA	project	Trans-European Research and Education Networking Association	
Thaesis	company		
THATCamp	Event type	The Humanities and Technology Camp	
TiCC	organisation	Tilburg Centre for Cognition and Communication	
TICCLops	CLARIN-NL subproject	Text-Induced Corpus Clean-up online processing system	
Tilburg University School of Humanities	organisation		
TIN	conference	Linguistics in the Netherlands	Taalkunde In Nederland
TLA	organisation	The Language Archive	
TQE	CLARIN-NL subproject	Transcription Quality Evaluation	
TransAtlantis	project	NWO Horizon Programme about the influence of the US on public debates in the Netherlands	
TROVA	software		
TROVe	software	TransMedia Observatory	
TTNWW	CLARIN-NL subproject	LRT Tools for Dutch as web	TST Tools voor het Nederlands als Web



		services in Work flows	services in Work flows
TU	organisation	Twente University CTIT	
Tweede Kamer BIP	organisation	Dutch Parliament Office for Automation and Projects	Tweede Kamer der Staten Generaal Bureau voor Informatisering en Projecten
Tweede Kamer DIV	organisation	Dutch Parliament Information Service	Tweede Kamer der Staten Generaal Dienst InformatieVoorziening
UBU	organisation	Utrecht University Library	
UC Davis	organisation	University California Davis	
UIL-OTS	organisation	Utrecht Institute of Linguistics OTS	
UK	country	United Kingdom	
UM	organisation	University of Maastricht	
URL	term	Universal Resource Locator	
USA	country	United States of America	
UT	organisation	University of Twente	
UU	organisation	Utrecht University	
UvA	organisation	University of Amsterdam	
VALID	CLARIN-NL subproject	Vulnerability in Acquisition: Language Impairments in Dutch. Curating five valuable data sets	
Vancis	company		
VCC	organisation	Virtual Competency Centre	
VerteldVerleden	project	Open Platform for Oral History	
VICI	grant programme	specific type of grant awarded by NWO	

Video Active	project		
VIDI	grant programme	specific type of grant awarded by NWO	
VK	CLARIN-NL subproject	Enriched Kingdom	Verrijkt Koninkrijk
VKS	organisation	Virtual Knowledge Studio	
VL	region	Flanders	Vlaanderen
VLO	software	Virtual Language Observatory	
VRE	software	Virtual Research Environment	
VSNU	organisation	Association of Collaborating Dutch Universities	Vereniging van Universiteiten
VU	organisation	VU University of Amsterdam	
VU-DNC	CLARIN-NL subproject	VU Diachronic Newspaper Corpus	
VVV	organisation	Touristic Office	
WAHSP	CLARIN-NL subproject	Web-application for historical sentiment mining in public media	
WAYF	standard	Where Are You From	
WBD	data	Brabant Dialects Dictionary	Woordenboek van de Brabantse Dialecten
WBS	term	administrative code for Utrecht University	
WFT	data	Dictionary of the Frisian Language	Wurdboek fan 'e Fryske Taal
WFT-GTB	CLARIN-NL subproject	Integrating the Wurdboek fan 'e Fryske Taal into the Geïntegreerde Taalbank	
WhiteLab	software	Search Engine front end	
WIP	CLARIN-NL	War in	

	subproject	Parliament	
WIVU database	data	Hebrew Bible Database constructed by the Werkgroep Informatica (WI) of the Free University (VU)	
WLD	data	Limburg Dialects Dictionary	Woordenboek van de Limburgse Dialecten
WMTC	organisation	Netherlands Research School of Science, Technology and Modern Culture	
WP	term	Work Package	
WTMC	organisation	Netherlands Graduate Research School of Science, Technology and Modern Culture	
XML	Standard data format	eXtensible Markup Language	
Xpath	standard	XML Path Language	
Xquery	standard	W3C XML Query	
xTAS	software	Text Analysis in a Timely Manner	