



# CLARIN-NL Annual Report 2012

## 1 Introduction

This document provides a description of the activities carried out in CLARIN-NL for 2012.

## 2 Executive Summary

The year 2012 has been a very successful year for CLARIN-NL. The CLARIN ERIC was established, and this fact was celebrated on several occasions (at the Ministry of Education, Culture and Sciences; at the opening of the Academic Year 2012-2013 in Utrecht, and on a special occasion set-up by the European Committee). Many people active in CLARIN-NL have become active for the CLARIN ERIC its governance organisation and committees.

Basic functionality for the CLARIN infrastructure has been further extended (AAI, federated search, vocabulary services, CMDI metadata profile for tools, etc.). The vocabulary service developed in CLAVAS is already in use also outside of the Netherlands, inter alia in Austria. A plan for integrating the various pieces of the CLARIN infrastructure has been elaborated and is ready to be executed. The various CLARIN centres have re-assessed their ambitions and all have reconfirmed their ambition to become at least a B-centre, and some even aim to become a B+ centre or an A-centre. A range of independent data providers have expressed their willingness to make their data available in a CLARIN-compatible manner and have executed (NISV) or are executing plans (KB, DBNL) to realise this. Others are preparing plans for this.

All subprojects from the Third Call have started up and are currently running. The researchers in these projects are, as usual, supported for various CLARIN-related aspects by the HelpDesk, and for metadata and semantic interoperability via tutorials, workshops and dedicated experts. The results of the Call 2 projects have become available in 2012, adding to the results of the Call 1 projects that were available already for quite some time.

As planned, a Fourth call for data curation and demonstrator projects has been launched. It included both an Open part and a Closed Part. In total, 11 subprojects could be awarded funding (out of 20 submissions). These projects will start in the first half of 2013.

The cooperation project of the Netherlands and Flanders is coming to an end. It has, inter alia, resulted in a workflow system that enables a researcher to combine a wide variety of language and speech technology web services to enrich data for his or her research. A keynote speech on this system will be held at CLIN 2013 in Enschede. Systematic support for the CLAM-system, which originated in this project and has enabled many to turn their tools into web services in an easy manner, could be provided thanks to CLARIN-NL.

The Data Curation Centre, one of the (virtual) CLARIN-NL centres of expertise, has started its activities. It curated several resources and made a data curation guideline.



The *Nederlab* project, a project that aims to provide data and tools for the longitudinal study of the Dutch language and culture, has been supported financially and has been set up as the second (virtual) CLARIN-NL centre of expertise, ready to start its activities in 2013.

A wide range of events was organized by CLARIN-NL: various tutorials (both nationally and internationally), workshops, information sessions, etc. CLARIN-NL also supported various independent events financially, and it has supported various researchers to attend workshops and conferences. Many of these workshop and conference visits have led to presentations and publications. Newsflashes have been sent out regularly, and CLARIN-NL has supported *eData & Research*, in which a wide variety of articles on or related to CLARIN have appeared. Several movies and screen captures that show the results of various subprojects have been made. The website has been redesigned and the underlying content management system is being upgraded to a state-of-the-art version. Moreover, many CLARIN-participants wrote a [blog](#) about the progress of their projects and/or about their CLARIN-project presentations in and outside the Netherlands.

The National Coordination Point has functioned well. Some members have left the CLARIN Board and the CLARIN NAP, but excellent replacements have been found for them. The IAP has been extended with a number of renowned international experts.

Finally, a project proposal for a successor project, CLARIAH had been submitted in 2011. It was put on the National Roadmap in 2012, and it received 1 million euro of 'seed money', from the deputy minister Halbe Zijlstra on a special event for this occasion in Utrecht (March 3, 2012), The 'seed money' is to be used for maintaining the consortium dynamics and for preparing a new proposal to be submitted by 2013.

### 3 Planned Activities

#### EU-Line (WBS 1.9)

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

- Membership fee: the Netherlands has to pay to CLARIN ERIC as its host, for 2012. In 2012 the first part of the fee has been paid. The amount was slightly smaller than planned, since CLARIN ERIC started only as of Feb 2012, so that only 11/12<sup>th</sup> of the annual fee had to be paid.
- Costs for transition from CLARIN-Prep to CLARIN ERIC. A small amount of the budget was used for the transition from the European CLARIN-prep project (which ended in June 2011) to the start of CLARIN ERIC (Feb 2012). This involves mainly salaries and travel expenses for the CLARIN-prep coordinator (Steven Krauwer).

In addition, costs were temporarily charged to this budget line that were made by CLARIN ERIC before it had its own bank account etc. These costs will be invoiced to CLARIN ERIC in 2013.

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



CLARIN-NL has been very active in 2012 towards the European enterprise to create a CLARIN infrastructure under the guidance of CLARIN ERIC, and it has been very active 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.
  - Marc Kemps-Snijders, Member Standing Committee for CLARIN Technical Centres<sup>1</sup>
  - 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
  - Remco van Veenendaal, 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 CLARIN-NL attended the first CLARIN Annual meeting in Sofia, Bulgaria, and many of them had contributions in the form of presentations and/or showing demonstrators.
- 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](#)<sup>2</sup>, which brings together [all 5 ESFRI research infrastructure initiatives](#)<sup>3</sup> in the social sciences and humanities
  - [EUDAT](#)<sup>4</sup>, 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](#)<sup>5</sup> (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

---

<sup>1</sup> Dieter Van Uytvanck, is Chair Standing Committee for CLARIN Technical Centres (on behalf of Germany)

<sup>2</sup> <http://dasish.eu/>

<sup>3</sup> <http://dasish.eu/links/>

<sup>4</sup> <http://www.eudat.eu/>

<sup>5</sup> <http://rd-alliance.org/>



## NL-Line

### 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 delivered intermediate status reports, and several initial versions of crucial components of the infrastructure.

### Infrastructure Implementation Plan (IIP)

Participants in the IIP project are the centres that expressed the ambition and made the commitment to become a CLARIN type A/B centre in the Netherlands, i.e. MPI (Nijmegen), Meertens Institute (Amsterdam), INL (Leiden), and DANS (The Hague).

In the course of 2012, Huygens Institute for Dutch History (Huygens ING), which also has expressed the ambition to become a CLARIN Centre, submitted a plan to for this. The plan has been accepted, and will start in 2013.

There has been discussion during this year about the status and ambitions of the various centres.

The major issues are (1) how are the costs covered after the CLARIN-NL-project; and (2) many centres are unsure whether they can host software services, and how this should be done, and are therefore reluctant to make strong commitments. For this reason, a new centre type (B+), has been defined, In addition to the known types A and B. A B+-centre is a centre that explicitly also aims to host software and data management services (in addition to hosting data ) for the whole CLARIN community.

MPI has reconfirmed its ambition to become an A/B+ centre. DANS has made clear that it aims to be a B-centre (and has made a plan towards it). INL, HI, and MI aim to be B+-centres. All centres have to work towards certification as a CLARIN-centre. A dedicated workshop on this for whole CLARIN will be held in 2013.

It became clear that the metadata of many data and tools that reside at the centres are not yet visible via the Virtual Language Observatory. An action to improve on this has been started up.

The IIP project is responsible for contributing to the CMDI metadata infrastructure registries and tools such as the VLO metadata catalogue and the ARBIL metadata editor. It also works on building and maintaining several for CLARIN important registries as ISOcat, RELcat and SCHEMAcat

### Search & Develop (S&D)

The S&D project aims to create a network of centres in the Netherlands, focusing around text content search services. The S&D projects aims to create search facilities for search in metadata stored in a centralized database (itself filled and regularly updated via metadata harvesting), as well as federated search facilities in the actual data. A first prototype S&D federated search demonstrator with limited functionality (term search only) has been created.<sup>6</sup> At the Sofia CLARIN conference (Oct

---

<sup>6</sup> <http://lux17.mpi.nl/ds/fedsearch/>



2012), it became clear that there are different views (and implementations) on federated search, an issue that will be taken up in a dedicated CLARIN workshop in 2013.

The S&D project is otherwise running according to plan.

#### **CLAVAS: A Vocabulary Service for CLARIN**

It has become clear in the IIP project that ISOCAT cannot function as the only concept registry in the CLARIN infrastructure. For certain concepts ISOCAT is not particularly suited, and for other concepts independently maintained registries exist and duplication is to be avoided. For this reason, a separate project, CLAVAS, was started up to build an interface to a variety of concept and data category registries, so that CLARIN users can transparently use concepts and data categories from each of them. This project has some delays, though the main functionality is available and already used by other partners in Europe (e.g. Austria). One part of the project, automatic normalization of organisation names, was not successful. It has been decided to have the relevant work done manually by the Data Curation Service in 2013.

#### **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, Royal Library), *het Nederlands Instituut voor Beeld & Geluid* (NIBG, 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*. 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*, 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 approved and is currently being executed. It is expected that the targeted project end date (Feb 2013) will not be feasible, and KB is working on a revised plan.
- NIBG has submitted a plan, which was approved and executed successfully. 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 has submitted a plan, which was approved. However, just after the start of the project, DBNL came into trouble, and was transferred to the KB. The EB is now investigating with the DBNL and KB what the situation is. There is a real danger that the money provided for this project as an advance (22.5k euro) is definitively lost.
- NA has submitted a plan, which was not approved. It has been discussed with NA what the plan should look like to become approved, and NA repeatedly claimed to submit a revised plan. However, due to various circumstances, no plan was submitted. The EB gave the NA

Feb 2013 as ultimate date to submit a plan. However, NA has communicated that they will not be able to submit a plan by then or even shortly after it. The EB will reallocate the reserved money in the 2013 budget.

- The Department of information Services of the *Tweede Kamer* has not been invited to submit a proposal, because the University Library Utrecht (UBU) appeared a candidate that should be given higher priority
- UBU has submitted a plan, which is being evaluated. (It has been approved as of March 19, 2013)

The EB aimed to test how visible and accessible the data at these CLARIN Data Providers actually are, after the projects have finished. For IBG, a natural test can be formed by the *Polimedia* project that they are involved in. For KB, it was considered to set up an experiment with the WIP project, which uses text corpora from parliamentary debates that are also available at the KB. However, the results of the WAHSP project instantiate the envisaged experiment, so that a new experiment is not needed. The WAHSP-project uses actual KB-data, which proves the visibility and accessibility of the KB-data. However, performance issues make the WAHSP/BILAND developers consider an approach where a copy of the KB-data is used locally. This is a more general problem (which was predicted by many people)) that should be addressed at the level of the CLARIN infrastructure.

### Data Deposition Service

The board aimed to submit a revised version for a *Data Deposition Service* (DDS). The original plan for this DDS was made in 2011, submitted to the Board, but rejected. The EB did not submit a revised plan in 2012 (mainly due to lack of time) and does not plan to do so in 2013 either (because of lack of time and money). However, the EB still believes that DDS functionality is important and an essential ingredient for reducing costs for sustainability.

CLARIN A centres can offer data deposition services as a service to the CLARIN community. It is expected that when centres such as Meertens and MPI make CMDI metadata and PIDs part of their internal workflow and architecture, they can offer already such a service without much extra effort. It remains to be seen if this will then indeed be offered and if a limited number of centres offering such a service is sufficient.

This topic surely will have to be taken up in CLARIAH

### CMDI Metadata scheme for tools

It was planned to set up a small project to create a *CMDI metadata scheme for tools*. Actual activities for making the scheme have started in August 2012 and are targeted to end by January 2013.

### Integration

The most important activity to be started up in 2012 was an *integration activity* 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 infrastructure and the portal towards it will have to grow incrementally and any new additions, modifications etc. will have to be accompanied by good PR activities. The definitive plan for these integration activities has become available only late in 2012 and has been approved by the Board, but the actual work will start only as of 2013.

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

#### Calls

All projects from Call 1 were formally closed in the first quarter of 2012, as planned. The only exception is the TDS-Curator project (which has been formally closed early 2013)

We targeted for all projects from Call 2 to finish and get formally closed in 2012. Though all projects did finish their work in 2012, most were not formally closed in 2012. Only C-DSD, WIP and VU-DNC were formally closed in 2012. The major reasons for these delays are: (1) project coordinators take a long time before submitting their final report and deliverables even though the work on the project has already finished; (2) lack of time due to many other activities (e.g. organizing Call 4, new plans for 2013) cause delay in evaluation of the submitted reports and deliverables in the EB.

We targeted for all projects from Call 3 to start up in the first quarter of 2012. This turned out to be too optimistic, but all projects have started as of May 1, 2012.

We targeted for all Call 3 projects to finish and be formally closed ultimately by mid 2013, and that is still a realistic target.

As planned, new Call, Call 4 was set up in 2012. After an investigation by the EB, two disciplines (Philosophy and Media Studies) and two researchers in these areas were invited to submit a proposal in the Closed Call. Both project were submitted and approved. In addition, 80k from the Closed Call budget was reserved for NederLab (see below, under Expertise Centres). An Open Call has been issued. 17 submissions were made. An 18<sup>th</sup> submission (CLARIN-LL-12-024, ITASAP) was exceptionally allowed to be submitted later than the deadline. However, initial inspection for compliance showed that this project proposal did not meet the compliance criteria of Call 4, and was therefore rejected by the EB and not even submitted to IAP and NAP for evaluation. Nine projects (out of 17 compliant project proposals submitted) could be awarded funding. As planned, Call 4 was the last call in the CLARIN-NL project, and the budgets available for this from 2012, 2013 and 2014 have been used for it.

In 2012, a project proposal for an earlier Closed Call (the 'Maarten Mous project') has been submitted and approved (CLARIN-NL 12-002, LAISEANG, coordinator Marian Klamer). It will start early 2013.

The EB has investigated whether the call could be issued earlier than in previous years, but this turned out to be very difficult, for various reasons, inter alia:



- More time than anticipated was needed to select the closed call candidates
- Revised schedule did not fit in very well with the IAP

Since the call could not be issued earlier, a start date for the awarded projects by Jan 1, 2013, is not feasible. A more realistic start date for the Call 4 projects is March or April 2013, and this appears to be the case for most projects.

The Call invited, as planned, for resource curation and demonstrator projects. The same maximum budget amounts for projects were used as in 2011: max. 120k for an open call project; max 80k for a closed call project; maximum duration of 1 year. The total budget available for this call was 1 million euro, as planned (actually slightly more has been used), 240k of which has been used for closed call projects, including Nederlab.

All calls, hence also the Call 4, are 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 IMDI or (if available 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.

A detailed report on the procedure followed for the Call 4, together with reports on the evaluation by the IAP and NAP, has been made and was approved by the Board on December 18th, 2012.

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.

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 2012, though where needed also its status as of March 2013 is specified. More detailed descriptions of the projects can be found on the CLARIN-NL website.<sup>7</sup>

### Call 1 Projects

<b>ADEPT</b>	<b>Assaying Differences via Edit-Distance of Pronunciation Transcriptions</b>
<b>Project Coordinator</b>	<a href="#">Dr. C. Gooskens</a> (Groningen University)
<b>Goal</b>	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.
<b>Budget</b>	€ 57K
<b>Centre</b>	Meertens Instituut
<b>Scientific impact</b>	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.
<b>Status</b>	The project has effectively finished and has submitted its final reports, which have been approved. [ <a href="#">web service</a> ]

<b>AAM-LR</b>	<b>Automatic Annotation of Multi-modal Language Resources</b>
<b>Project Coordinator</b>	<a href="#">Prof.dr. L. Boves</a> (Radboud University Nijmegen)
<b>Goal</b>	The AAM-LR project aims at building a demonstrator of a web service that will help field researchers to annotate audio- and video-recordings
<b>Budget</b>	€ 56K
<b>Centre</b>	MPI
<b>Scientific impact</b>	It will facilitate all field research in which digital audio or video recordings are used.

<sup>7</sup> <http://www.clarin.nl/node/76>

<b>Status</b>	The project has effectively finished and has submitted its final reports, which have been approved. [ <a href="#">web application</a> <sup>8</sup> <a href="#">show case</a> <sup>9</sup> ]
---------------	---

<b>Adelheid</b>	<b>A Distributed Lemmatizer for Historical Dutch</b>
<b>Project Coordinator</b>	<a href="#">Dr. H. van Halteren</a> (Radboud University Nijmegen)
<b>Goal</b>	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.
<b>Budget</b>	€ 55K
<b>Centre</b>	MPI
<b>Scientific impact</b>	It will facilitate research on historical texts in fields such as historical linguistics, literature science and history.
<b>Status</b>	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. [ <a href="#">web service</a> ]

<b>DUELME-LMF</b>	<b>Converting DUELME into LMF format</b>
<b>Project Coordinator</b>	<a href="#">Prof.dr. J.E.J.M. Odiijk</a> (Utrecht University)
<b>Goal</b>	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.
<b>Budget</b>	€ 60K
<b>Centre</b>	INL
<b>Scientific impact</b>	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.
<b>Status</b>	The project has effectively finished, and has submitted its final reports, which have been approved. [ <a href="#">metadata</a> <sup>10</sup> ]

<sup>8</sup> <http://lux17.mpi.nl/aamlr>. This is a temporary URL, which may disappear in the future

<sup>9</sup> <http://www.clarin.nl/showcase/aam-lr/268>

<sup>10</sup>

[http://oai.dev.clarin.inl.nl/oai\\_server.php?verb=GetRecord&identifier=hdl:10032/89511c6d452ec7110717d8b3fc359e4d&metadataPrefix=cmdi](http://oai.dev.clarin.inl.nl/oai_server.php?verb=GetRecord&identifier=hdl:10032/89511c6d452ec7110717d8b3fc359e4d&metadataPrefix=cmdi)

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

<b>MIMORE</b>	<b>Microcomparative Morphosyntax Research Tool</b>
<b>Project Coordinator</b>	<a href="#">Prof.dr. S. Barbiers</a> (Meertens Institute and Utrecht University)
<b>Goal</b>	The demonstrator tool MIMORE will create a common search engine for the databases <a href="#">DynaSAND</a> , <a href="#">DiDDD</a> and <a href="#">MAND</a>
<b>Budget</b>	€ 54K
<b>Centre</b>	Meertens Instituut
<b>Scientific impact</b>	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.
<b>Status</b>	The project has effectively finished, and has submitted its final reports, which have been approved. [ <a href="#">web service</a> ]

<b>SignLinC</b>	<b>Linking lexical databases and annotated corpora of signed languages</b>
<b>Project Coordinator</b>	<a href="#">Dr. O.A. Crasborn</a> (Radboud University Nijmegen)
<b>Goal</b>	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.
<b>Budget</b>	€ 56K
<b>Centre</b>	MPI
<b>Scientific impact</b>	The linked databases and corpora will facilitate research in the area of sign language linguistics
<b>Status</b>	The project has effectively finished, and has submitted its final reports, which have been approved. [ <a href="#">web service</a> ]

<b>TICCLops</b>	<b>Text-Induced Corpus Clean-up online processing system</b>
<b>Project Coordinator</b>	Dr. <a href="#">M. Reynaert</a> (Tilburg University)

<b>Goal</b>	This demonstrator project will develop a tool that allows CLARIN users to submit their corpora for fully automatic spelling correction and normalization
<b>Budget</b>	€ 54K
<b>Centre</b>	INL
<b>Scientific impact</b>	The tool will facilitate all research that makes use of text corpora.
<b>Status</b>	The project has effectively finished, and has submitted its final reports, which have been approved. [ <a href="#">web service</a> ]

<b>TDS Curator</b>	<b>A web-services architecture to curate the Typological Database System</b>
<b>Project Coordinator</b>	<a href="#">Dr. A. Dimitriadis</a> (Utrecht University)
<b>Goal</b>	TDS Curator will make the Typological Database System (TDS) into a sustainable service that conforms to CLARIN infrastructural requirements.
<b>Budget</b>	€ 57K
<b>Centre</b>	DANS
<b>Scientific impact</b>	The (TDS) facilitates typological linguistic research by providing integrated access to multiple independently developed typological databases through a common web interface
<b>Status</b>	The project is finished but the final reports are still to be submitted. [ <a href="#">web service</a> ]

<b>TQE</b>	<b>Transcription Quality Evaluation</b>
<b>Project Coordinator</b>	<a href="#">Dr. H. Strik</a> (Radboud University Nijmegen)
<b>Goal</b>	The TQE project aims to create a completely automatic Transcription Quality Evaluation (TQE) tool.
<b>Budget</b>	€ 57K
<b>Centre</b>	MPI
<b>Scientific impact</b>	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.
<b>Status</b>	The project has effectively finished, and has submitted its final reports, which have been approved. [ <a href="#">web service</a> ]

<b>WFT-GTB</b>	<b>Integrating the Wurdboek fan 'e Fryske Taal into the Geïntegreerde Taalbank</b>
<b>Project Coordinator</b>	<a href="#">Drs. H. Sijens</a> (Fryske Akademy)
<b>Goal</b>	The project aims to carry out data curation of the <i>Wurdboek fan de Fryske Taal</i> (WFT) database and integrate the data in the <i>Geïntegreerde Taalbank</i> (GTB) and

	demonstrate it via the GTB dictionary web application.
<b>Budget</b>	€ 56K
<b>Centre</b>	INL
<b>Scientific impact</b>	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.
<b>Status</b>	The project has effectively finished and demonstrated its results at <a href="#">Euralex 2010</a> held in Leeuwarden. It has submitted its final reports, which have been approved.

CKCC project	CKCC project
<b>Project Coordinator</b>	<a href="#">Drs. G. Gerritsen</a> (Huygens ING)
<b>Goal</b>	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.
<b>Budget</b>	€ 59K
<b>Centre</b>	DANS/Huygens ING
<b>Scientific impact</b>	The CKCC project ( <i>Circulation of Knowledge and Learned Practices in the 17th-century Dutch Republic</i> , 'Geleerdenbrievenproject ') 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 <sup>th</sup> 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). <sup>11</sup>
<b>Status</b>	The project is running according to schedule and expected to finish early 2013

### Call 2 Projects

Arthurian-Fiction	Arthurian Fiction in Medieval Europe
<b>Project Coordinator</b>	<a href="#">Dr. B. Besamusca</a> (Utrecht University)
<b>Goal</b>	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

<sup>11</sup> [http://www.clarin.eu/files/cnl06\\_web\\_0.pdf](http://www.clarin.eu/files/cnl06_web_0.pdf)

	area of European Arthurian fiction and to create a demonstrator that will allow users access to the data for searching and editing.
<b>Budget</b>	€ 47K
<b>Centre</b>	Huygens ING
<b>Scientific impact</b>	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.
<b>Status</b>	The project is finished and has submitted its final reports, which are being evaluated by the EB. (March 2013: reports approved)

<b>C-DSD</b>	<b>Curating the Dutch Song Database</b>
<b>Project Coordinator</b>	<a href="#">Dr. E. Stronks</a> (Utrecht University)
<b>Goal</b>	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.
<b>Budget</b>	€ 29K
<b>Centre</b>	Meertens Institute
<b>Scientific impact</b>	This project will create new opportunities for (international) collaboration with other projects with similar source materials.
<b>Status</b>	The project submitted its final reports, which have been approved

<b>COAVA</b>	<b>Cognition, Acquisition and Variation Tool</b>
<b>Project Coordinator</b>	<a href="#">Dr. L. Cornips</a> (Meertens Institute Amsterdam)
<b>Goal</b>	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.
<b>Budget</b>	€ 66K
<b>Centre</b>	Meertens Institute
<b>Scientific impact</b>	It will enable interdisciplinary research into the relation between language acquisition and language variation.
<b>Status</b>	The project has finished but not submitted its final reports yet, (march 2013: final reports submitted and are being evaluated)

<b>INPOLDER</b>	<b>Integrated Parser and Lemmatizer Dutch in Retrospect</b>
<b>Project Coordinator</b>	<a href="#">Prof.dr. A. van Kemenade</a> (Radboud University Nijmegen)

<b>Goal</b>	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.
<b>Budget</b>	€ 69K
<b>Centre</b>	Meertens Institute
<b>Scientific impact</b>	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.
<b>Status</b>	The project has finished but the final reports have not been submitted yet. (march 2013: final reports not yet received)

<b>IPROSLA</b>	<b>Integrating and publishing resources on sign language acquisition</b>
<b>Project Coordinator</b>	<a href="#">Prof.dr. P. Fikkert</a> (Radboud University Nijmegen)
<b>Goal</b>	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.
<b>Budget</b>	€ 67K
<b>Centre</b>	MPI
<b>Scientific impact</b>	By integrating two datasets in integrated archive and making them visible and accessible, it will enable better and more research on Sign Language.
<b>Status</b>	The project has finished but not submitted its final reports yet. (March 2013: final reports submitted and approved)

<b>NEHOL</b>	<b>Negerhollands Database</b>
<b>Project Coordinator</b>	<a href="#">Prof.dr. P.C. Muysken</a> (Radboud University Nijmegen)
<b>Goal</b>	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.
<b>Budget</b>	€ 43K
<b>Centre</b>	MPI
<b>Scientific impact</b>	It will enable research into Negerhollands and thereby into Creole languages and language creolization processes in general.
<b>Status</b>	The project has finished but is still to submit its finale reports, (March 2013: final reports have been submitted and are being evaluated).

<b>VU-DNC</b>	<b>VU Diachronic Newspaper Corpus</b>
<b>Project Coordinator</b>	<a href="#">Prof.dr. W. Spooren</a> (Free University Amsterdam)
<b>Goal</b>	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 MW) 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.
<b>Budget</b>	€ 59K
<b>Centre</b>	INL
<b>Scientific impact</b>	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.
<b>Status</b>	The project has finished and submitted its final reports, which have been approved.

<b>WAHSP</b>	<b>Web-application for historical sentiment mining in public media</b>
<b>Project Coordinator</b>	<a href="#">Prof.dr. T. Pieters</a> (Utrecht University)
<b>Goal</b>	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.
<b>Budget</b>	€ 120K
<b>Centre</b>	Huygens ING
<b>Scientific impact</b>	It will enable historical research by providing facilities for mining large historical data sets, where it focuses on the opinions/perceptions regarding the use and abuse of drugs between 1900 and 1945.
<b>Status</b>	The project has finished and submitted its final reports, which are being evaluated by the EB. (March 2013: final reports have been approved).

<b>WIP</b>	<b>War in Parliament</b>
<b>Project Coordinator</b>	<a href="#">Dr. H. Piersma</a> (NIOD)
<b>Goal</b>	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.
<b>Budget</b>	€ 118K
<b>Centre</b>	DANS

<b>Scientific impact</b>	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.
<b>Status</b>	The project is has finished and submitted its final reports, which have been approved.

### Call 3 Projects (Closed Call)

<b>FESLI</b>	<b>Functional Elements in Specific Language Impairment</b>
<b>Project Coordinator</b>	<a href="#">Prof.dr. F. Weerman</a> (University of Amsterdam)
<b>Goal</b>	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.
<b>Budget</b>	€ 79K
<b>Centre</b>	Meertens Instituut
<b>Scientific impact</b>	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.
<b>Status</b>	The project is running since May 1, 2012, and targeted to finish by April 30, 2013.

<b>PILNAR</b>	<b>Pilgrimage Narratives: creating a germ corpus for studying the profile of the modern pilgrim</b>
<b>Project Coordinator</b>	<a href="#">Prof.dr. P. Post</a> (Tilburg University)
<b>Goal</b>	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.
<b>Budget</b>	€ 79K
<b>Centre</b>	Meertens Instituut
<b>Scientific impact</b>	It will make searching in pilgrimage narratives much easier and more powerful, this boosting research in the religion sciences.
<b>Status</b>	The project is running since April 1, 2012 and targeted to finish by April 1, 2013

<b>GrNe</b>	<b>Grieks-Nederlands woordenboek (Greek-Dutch dictionary)</b>
<b>Project Coordinator</b>	<a href="#">Prof.dr. I. Sluiter</a> (Leiden University)
<b>Goal</b>	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).
<b>Budget</b>	€ 55K
<b>Centre</b>	University Library of UvA
<b>Scientific impact</b>	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.
<b>Status</b>	The project is running since April 1, 2012 and targeted to finish by April 1, 2013.

<b>DiscAn</b>	<b>Towards a Discourse Annotation system for Dutch language corpora</b>
<b>Project Coordinator</b>	<a href="#">Prof.dr. T.J.M. Sanders</a> (Utrecht University)
<b>Goal</b>	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.
<b>Budget</b>	€ 80K
<b>Centre</b>	MPI
<b>Scientific impact</b>	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.
<b>Status</b>	The project is running since April 1, 2012 and targeted to finish by April 1, 2013

### Call 3 Projects (Open Call)

<b>BILAND</b>	<b>Towards a flexible and stable CLARIN-supported web-application for bilingual historical analysis of discourses in news media</b>
<b>Project Coordinator</b>	<a href="#">Prof.dr. T. Pieters</a> (Utrecht University)
<b>Goal</b>	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.
<b>Budget</b>	€ 120K
<b>Centre</b>	Huygens ING
<b>Scientific impact</b>	The project will greatly facilitate research in history and of advanced discourse analysis in historical datasets.

<b>Status</b>	The project is running since Feb 1, 2012 and originally targeted to finish by Feb 1, 2012..It has requested an extension until May 31 <sup>st</sup> , 2013, which has been given.
---------------	---

<b>MultiCon</b>	<b>Multilayer Concordance Functions in ELAN and ANNEX</b>
<b>Project Coordinator</b>	<a href="#">Dr. O.A. Crasborn</a> (Radboud University Nijmegen)
<b>Goal</b>	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.
<b>Budget</b>	€ 55K
<b>Centre</b>	MPI
<b>Scientific impact</b>	The project will offer new research opportunities in multimodality studies, both in linguistics and beyond.
<b>Status</b>	The project is running since April 1, 2012 and targeted to finish by June 1, 2013.

<b>D-LUCEA</b>	<b>Database of the Longitudinal Utrecht Collection of English Accents</b>
<b>Project Coordinator</b>	<a href="#">Dr. H. Quené</a> (Utrecht University)
<b>Goal</b>	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.
<b>Budget</b>	€ 67K
<b>Centre</b>	MPI
<b>Scientific impact</b>	This project will greatly facilitate linguistic and phonetic research into L1 and L2 acquisition.
<b>Status</b>	The project is running since April 1, 2012 and targeted to finish by April 1, 2013.

<b>EMIT-X</b>	<b>Early-Modern Image and Text eXchange</b>
<b>Project Coordinator</b>	<a href="#">Dr. E. Stronks</a> (Utrecht University)
<b>Goal</b>	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.
<b>Budget</b>	€ 20K
<b>Centre</b>	Huygens ING
<b>Scientific impact</b>	The project will greatly facilitate research in literary studies, esp. the study of emblems.

<b>Status</b>	The project is running since March 1, 2012 and originally targeted to finish by March 1, 2013. It has requested an extension until June 1, 2013, which has been given.
---------------	--

<b>NameScope</b>	<b>Mapping the Landscape of Names in Modern Dutch Literature</b>
<b>Project Coordinator</b>	<a href="#">Prof.dr. K. van Dalen-Oskam</a> (Huygens ING)
<b>Goal</b>	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.
<b>Budget</b>	€ 120K
<b>Centre</b>	INL
<b>Scientific impact</b>	The project will greatly enhance the research opportunities in literary studies esp. the study of named entities and their roles in literary works.
<b>Status</b>	The project is running since March 15, 2012 and targeted to finish by April 1, 2013.

<b>VK</b>	<b>Verrijkt Koninkrijk (Enriched Kingdom)</b>
<b>Project Coordinator</b>	<a href="#">Dr. K. Ribbens</a> (NIOD)
<b>Goal</b>	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.
<b>Budget</b>	€ 111K
<b>Centre</b>	DANS
<b>Scientific impact</b>	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.
<b>Status</b>	The project is running since Feb 1, 2012 and originally targeted to finish by Jan 31, 2013. It has requested an extension until March 31, 2013, which has been given.

<b>Cornetto-LMF-RDF</b>	<b>Curated Cornetto database in LMF and RDF</b>
<b>Project Coordinator</b>	<a href="#">Prof.dr. P. Vossen</a> (Free University Amsterdam)
<b>Goal</b>	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).
<b>Budget</b>	€ 117K

<b>Centre</b>	INL
<b>Scientific impact</b>	This project will greatly enhance the research opportunities in study of semantics in particular and linguistics more generally.
<b>Status</b>	The project is running since May 1, 2012 and targeted to finish by May 1, 2013.

<b>MIGMAP</b>	<b>Interactive migration maps for the 20th century</b>
<b>Project Coordinator</b>	<a href="#">Dr. G. Bloothoof</a> (Utrecht University and Meertens Institute Amsterdam)
<b>Goal</b>	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 "Kaart" software module.
<b>Budget</b>	€ 54K
<b>Centre</b>	Meertens Institute
<b>Scientific impact</b>	The project will facilitate research into migration and linguistics, esp. language change.
<b>Status</b>	The project is running since March 1, 2012 and is targeted to finish by Dec 31, 2012. (March 2013: It has finished and its final reports and deliverables are being evaluated.)

<b>PoliMedia</b>	<b>Interlinking multimedia for the analysis of media coverage of political debates</b>
<b>Project Coordinator</b>	<a href="#">Prof.dr. H. Beunders</a> (Erasmus University Rotterdam)
<b>Goal</b>	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.
<b>Budget</b>	€ 111K
<b>Centre</b>	NIBG
<b>Scientific impact</b>	The project will enhance research opportunities for the cross-media analysis in historical studies.
<b>Status</b>	The project is running since April 1, 2012 and targeted to finish by April 1, 2013..

## Call 4 Projects

<b>QuaMerdes</b>	<b>Quantitative Content Analysis of Media Researchers' Data</b>
<b>Project Coordinator</b>	Dr. Jasmijn van Gorp (Utrecht University)
<b>Goal</b>	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.
<b>Budget</b>	76k
<b>Centre</b>	NISV/KB
<b>Scientific impact</b>	It will enable quantitative content analysis of television and printed media.
<b>Status</b>	The project is targeted to start early 2013.

<b>@PhilosTEI</b>	<b><i>TICCLing Philosophy: a TEI corpus-building workflow towards a new computational methodology for philosophy</i></b>
<b>Project Coordinator</b>	Dr. Arianna Betti (Free university Amsterdam)
<b>Goal</b>	@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-postcorrection web service TICCLops. The tool will be demonstrated on a multilingual, multi-script corpus of important 18th-20th-century European philosophical texts.
<b>Budget</b>	70k
<b>Centre</b>	HI
<b>Scientific impact</b>	It will contribute to increasing the digitally available empirical material, esp. for philosophy, but also for other disciplines
<b>Status</b>	The project is targeted to start in September 2013.

<b>eBNM+</b>	<b>Linked Data on Middle Dutch Sources Kept Worldwide</b>
<b>Project Coordinator</b>	Dr. André Bouwman (Leiden University Libraries)
<b>Goal</b>	<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 world wide 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.
<b>Budget</b>	68k
<b>Centre</b>	HI
<b>Scientific impact</b>	It will stimulate research that makes uses of Middle Dutch manuscripts.
<b>Status</b>	The project is targeted to start early 2013.

<b>VALID</b>	<b><i>Vulnerability in Acquisition: Language Impairments in Dutch. Curating five valuable data sets</i></b>
<b>Project Coordinator</b>	Dr. Jetske Klatter (Radboud University Nijmegen)
<b>Goal</b>	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.
<b>Budget</b>	64k
<b>Centre</b>	MPI
<b>Scientific impact</b>	It will enable better research on language pathology data, primarily for the Dutch language.
<b>Status</b>	The project is targeted to start early 2013.

<b>COBWWWEB</b>	<b>Connections Between Women and Writings Within European Borders</b>
<b>Project Coordinator</b>	Dr. Suzan van Dijk (Huygens ING)
<b>Goal</b>	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.
<b>Budget</b>	95k
<b>Centre</b>	HI
<b>Scientific impact</b>	It will enable better research on female writers and the reception of their works
<b>Status</b>	The project is targeted to start early 2013.

<b>OpenSoNaR</b>	<b>Online Personal Exploration and Navigation of SoNaR</b>
<b>Project Coordinator</b>	Dr. Martin Reynaert (Tilburg University)
<b>Goal</b>	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).
<b>Budget</b>	120k
<b>Centre</b>	INL

<b>Scientific impact</b>	It will provide better empirical support for linguistic and literary research
<b>Status</b>	The project is targeted to start early 2013.

<b>SHEBANQ</b>	<b>System for HEBrew Text: ANnotation for Queries and Markup</b>
<b>Project Coordinator</b>	Prof.dr. Wido van Peursen (Free University Amsterdam)
<b>Goal</b>	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.
<b>Budget</b>	101k
<b>Centre</b>	DANS
<b>Scientific impact</b>	It will enable better research in the Hebrew bible texts, and in theology
<b>Status</b>	The project is targeted to start early 2013.

<b>DSS</b>	<b>Dutch Ships and Seamen</b>
<b>Project Coordinator</b>	Prof.dr. Lex Heerma van Voss (Huygens ING)
<b>Goal</b>	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.
<b>Budget</b>	98k
<b>Centre</b>	HI
<b>Scientific impact</b>	It will enable better research into the maritime history of the Netherlands and its predecessors
<b>Status</b>	The project is targeted to start early 2013.

<b>EXILSEA</b>	<b>Exploiting ISOcat's Language Sections in ELAN and ANNEX</b>
<b>Project Coordinator</b>	Dr. Onno Crasborn (Radboud University Nijmegen)
<b>Goal</b>	EXILSEA is a mixed demonstrator and resource curation project from the open call that aims to make annotated audiovisual 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.



<b>Budget</b>	65k
<b>Centre</b>	MPI
<b>Scientific impact</b>	It will enable multilingual support for research into Sign Language
<b>Status</b>	The project is targeted to start early 2013.

<b>ColTime</b>	<b>Collaboration on time-based resources</b>
<b>Project Coordinator</b>	Dr. Onno Crasborn (Radboud University Nijmegen)
<b>Goal</b>	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.
<b>Budget</b>	68k
<b>Centre</b>	MPI
<b>Scientific impact</b>	It will enable better research on time-based resources
<b>Status</b>	The project is targeted to start early 2013.

<b>RemBench</b>	<b>A Digital Workbench for Rembrandt Research</b>
<b>Project Coordinator</b>	Prof.dr. Volker Manuth (Radboud University Nijmegen)
<b>Goal</b>	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.
<b>Budget</b>	121k
<b>Centre</b>	HI
<b>Scientific impact</b>	It will enable better research on the Rembrandt Documents
<b>Status</b>	The project is targeted to start early 2013.

### *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 started in March 2010, and was supposed to run for 2.5 years. It has requested and received an extension until Dec 31, 2012. The project has effectively finished, though there are still some software problems (memory leaks), esp. with regard to the speech modules, which are being investigated. A keynote speech on TTNWW will be held at CLIN 2013 (Jan 18, 2013) in Enschede. Final reports have not been submitted yet (March 2013: still not submitted), [Link to workflow system](#)<sup>12</sup>.

#### *ISOCAT Coordinator*

The work by the ISOCAT Coordinator (Ineke Schuurman) started up in 2011 has continued in 2012. A separate report on the ISOCAT Coordinator activities in 2012 is available..

#### *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. For this reason, Maarten van Gompel has been active in 2012 in a project financed by CLARIN-NL to maintain and further extend CLAM, and to provide support to its users. (1/10 appointment).

---

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



### Support for EPIC follow-up

The EPIC consortium aims to be a service provider for Persistent Identifiers for research data. SARA, one of the EPIC partners, together with Cultural Heritage institutes, needed raising 100k euro as of mid 2012 to have these services run. They might apply for co funding by CLARIN-NL to also widen the service offer. The EB was open for a concrete proposal, budget and list of other (candidate) co-funders and will evaluate the proposal in terms of its relevance for CLARIN(-NL). If approved will fund this from the call budget. However, no concrete proposal has been submitted in 2012.

### Centres of Expertise (WBS 1.6)

The virtual 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 run as planned and budgeted in 2012.

While curating these data, the DCS will investigate, plan and budget for additional candidates to be curated. The plan and budget for each selected resource will be submitted to the EB for approval before the data curation activities start. An actual overview of the data curation results and plans can be found on the [DCS project area](#)<sup>13</sup> on the CLARIN-NL website.

The EB has decided to continue the DCS activities in 2013, which still has to be approved by the Board (March 2013: approved by the Board).

The EB has requested the DCS to create a guideline for curation of data, so that other projects can benefit from this. This guideline is available [here](#)<sup>14</sup>.

As expected, the EB has concluded that an activity of harmonizing all (separately and independently created) metadata records may be required as of early 2013. It has been decided to reserve money for this in the 2013 budget (march 2013: this has been approved by the board). Though it differs from data curation per se, it is not unnatural to have the DCS carry out this task, though this still has to be decided upon.

As to a second Centre of Expertise, Steven Krauwer (SK) had suggested to set up a *CLARIN Annotation Support Centre* together with CLARIN-D. No concrete proposal for this has been received, and already early in 2012 it was decided to consider [Nederlab](#)<sup>15</sup> as the second (virtual) CLARIN-NL Centre of Expertise, more specifically on data and tools for the study of the Dutch language and culture across time. 150k Euro has been made available for *Nederlab* (80k from the Closed Call 4 budget and 70k from the Centres of Expertise budget). The major activities of *Nederlab* for CLARIN-NL will take place in 2014.

---

<sup>13</sup> <http://www.clarin.nl/node/147>

<sup>14</sup> <http://www.clarin.nl/group/415>

<sup>15</sup> <http://www.nederlab.nl/>

### Education and Awareness (WBS 1.7)

An overview is given of the various activities for stimulating education, training, and awareness carried out in 2012.

#### Conferences and workshops visited

From	To	Event	City	Country	Who
14/12/2012	14/12/2012	Oratie Rens Bod	Amsterdam	NL	JO, HB, AvH
11/12/2012	11/12/2012	Humanities Lecture Frans Wieringa	Utrecht	NL	JO
03/12/2012	04/12/2012	CATCH+ / DEN Digital Heritage Conference	Rotterdam	NL	HB
29/10/2012	29/10/2012	CLARIN's Turn to the literary text	Den Haag	NL	JO, KvDO, ES
24/10/2012	28/10/2012	CLARIN Integration Conference	Sofia	BU	EB + others
23/10/2012	23/10/2012	Kick-off CLICK-NL	Eindhoven	NL	HB
23/10/2012	23/10/2012	Digital Humanities Lecture Linguistic Research and The CLARIN Infrastructure	Utrecht	NL	JO
22/10/2012	23/10/2012	International Workshop on Open Access	Rotterdam	NL	HB
19/10/2012	19/10/2012	Kick-off Digital Humanities Amsterdam, KNAW	Amsterdam	NL	HB
18/10/2012	18/10/2012	PhD Defense Folkert de Vriend	Nijmegen	NL	HB
08/10/2012	08/10/2012	Presentation Project 'Gekaapte Brieven'	Rotterdam	NL	HB
03/10/2012	04/10/2012	SURF Relatiedagen	Noordwijkerhout	NL	JO
07/09/2012	07/09/2012	Afscheidscollege Geert Booij	Leiden	NL	JO, HB
07/09/2012	07/09/2012	Symposium Geert Booij	Leiden	NL	JO, HB
03/09/2012	03/09/2012	Opening Academisch jaar	Utrecht	NL	JO, SK
03/09/2012	03/09/2012	Opening letterenbibliotheek	Utrecht	NL	JO, SK
31/08/2012	31/08/2012	Presentation Network Cultural	Hilversum	NL	HB



		Heritage / CLICK-NL			
30/08/2012	30/08/2012	CLARIN informatiesessie Call 4	Amsterdam	NL	JO
20/06/2012	21/06/2012	META-FORUM 2012	Brussels	BE	JO
04/06/2012	04/06/2012	Workshop Open Access, Museon	The Hague	NL	HB
20/05/2012	27/05/2012	LREC	Istanbul	TU	JO, DB, AvH
23/03/2012	23/03/2012	Meeting with CLARIN-CZ/LINDAT (Jan Hajič, Eva Hajičová)	Prague	CZ	HB
22/03/2012	22/03/2012	Lecture in Prague	Prague	CZ	HB
06/03/2012	06/03/2012	CLARIN Kick-off	Hilversum	NL	JO
31/01/2012	01/01/2012	Lustrumcongres Alfa-Informatica en oratie Gertjan van Noord en Johan Bos	Groningen	NL	JO
20/01/2012	20/01/2012	CLIN	Tilburg	NL	JO, AvH

#### *Events organized by CLARIN-NL*

From	To	Description	City	Ctry	Who
23/10/2012	23/10/2012	Digital Humanities Lecture Linguistic Research and The CLARIN Infrastructure	Utrecht	NL	JO
30/08/2012	30/08/2012	CLARIN informatie sessie Call 4	Amsterdam	NL	EB
06/03/2012	06/03/2012	CLARIN Kick-off	Hilversum	NL	EB
13/09/2012	13/09/2012	CMDI Tutorial	Nijmegen	NL	DB
10/10/2012	10/10/2012	ISOCAT Tutorial 2012 (III)	Utrecht	NL	IS
19/06/2012	19/06/2012	ISOCAT Tutorial 2012 (II)	Utrecht	NL	IS
20/03/2012	20/03/2012	ISOCAT Tutorial 2012 (I)	Utrecht	NL	IS

### Publications

Here's a (not fully complete) overview:

- Broeder, D., Uytvanck, D. van, Gavriliadou, M., Trippel, T. and Windhouwer, M. (2012), "Standardizing a Component Metadata Infrastructure". In Calzolari, N., Choukri, K., Declerck, T., Doğan, M.U., Maegaard, B., Mariani, J., Odijk, J. and Piperidis, S. (eds.), Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12). Istanbul, Turkey: European Language Resources Association (ELRA), pp.1387-1390. [[pdf](#)]
- Broeder, D. Uytvanck, D. van, Senft, G. (2012), "Citing on-line Language Resources". In Calzolari, N., Choukri, K., Declerck, T., Doğan, M.U., Maegaard, B., Mariani, J., Odijk, J. and Piperidis, S. (eds.), Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12). Istanbul, Turkey: European Language Resources Association (ELRA), pp. 1391-1394. [[pdf](#)]
- Elbers, W., Broeder, D. and Uytvanck, D. van (2012), "Proper Language Resource Centers". In Calzolari, N., Choukri, K., Declerck, T., Doğan, M.U., Maegaard, B., Mariani, J., Odijk, J. and Piperidis, S. (eds.), Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12). Istanbul, Turkey: European Language Resources Association (ELRA), pp. 3260-3263. [[pdf](#)]
- Heuvel, H. van den, Sanders, E., Rutten, R., Scagliola, S. and Witkamp, P. (2012), "An Oral History Annotation Tool for INTER-VIEWS". In Calzolari, N., Choukri, K., Declerck, T., Doğan, M.U., Maegaard, B., Mariani, J., Odijk, J. and Piperidis, S. (eds.), Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12). Istanbul, Turkey: European Language Resources Association (ELRA), pp. 215-218. [[pdf](#)]
- Odijk, J.E.J.M. (2012), "Recent Developments in CLARIN-NL". In Calzolari, N., Choukri, K., Declerck, T., Dogan, M.U., Maegaard B., Mariani J., Odijk, J. and Piperidis, S. (eds.), Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12) (pp. 1055-1060). Istanbul, Turkey: European Language Resources Association (ELRA), pp. 1055-1060. [[url](#)] [[pdf](#)]
- Stehouwer, H., Durco, M. Auer, E. and Broeder, D. (2012), "Federated Search: Towards a Common Search Infrastructure". In Calzolari, N., Choukri, K., Declerck, T., Dogan, M.U., Maegaard B., Mariani J., Odijk, J. and Piperidis, S. (eds.), Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12). Istanbul, Turkey: European Language Resources Association (ELRA), pp. 3255-3259. [[pdf](#)]
- Uytvanck, D. van, Stehouwer, H. and Lampen, L. (2012), "Semantic metadata mapping in practice: the Virtual Language Observatory". In Calzolari, N., Choukri, K., Declerck, T., Dogan, M.U., Maegaard B., Mariani J., Odijk, J. and Piperidis, S. (eds.), Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12). Istanbul, Turkey: European Language Resources Association (ELRA), pp. 1029-1034. [[pdf](#)]
- Windhouwer, M. (2012), "RELcat: a Relation Registry for ISOcat data categories". In Calzolari, N., Choukri, K., Declerck, T., Dogan, M.U., Maegaard B., Mariani J., Odijk, J. and Piperidis, S. (eds.), Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12). Istanbul, Turkey: European Language Resources Association (ELRA), pp. 3661-3664. [[pdf](#)]
- Zhang, J. (2012), "Supporting Serendipitous and Focused Search". In Wilson, M.L., Russell-Rose, T., Larsen, B. and Kalbach, J. (eds.), Proceedings of the European Workshop on Human-

Computer Interaction and Information Retrieval (EuroHCIR2012). Tilburg, The Netherlands: CEUR Workshop Proceedings, pp. 79-82. [[pdf](#)]

Zhang, J., Kemps-Snijders, M. and Bennis, H. (2012), "The CMDI MI Search Engine: Access to Language Resources and Tools Using Heterogeneous Metadata Schemas". In Zaphiris, P. et al. (eds.), Proceedings of Theoretic and Practice Digital Libraries Conference (TPDL 2012). Springer Berlin/Heidelberg, Germany: Lecture Notes in Computer Science (LNCS), volume 7489, pp. 492-495. [[pdf](#)]

### *Presentations*

Here's a (not fully complete) overview:

Cornips, L., Kemps-Snijders, M., Snijders, M., Swanenberg, J. & De Vriend, F. de (2011, november 17). *Bridging the gap between first language acquisition and historical linguistics with the help of digital humanities*. University of Copenhagen, Supporting Digital Humanities. [[pdf](#)]

Crasborn, O. (2012), "Sign languages in Europe", poster presentation at the META FORUM 2012 meeting, Brussels, June 20-21, 2012. [[pdf](#)]

Crasborn, O. (2012), "Sign language technologies", poster presentation at the META FORUM 2012 meeting, Brussels, June 20-21, 2012. [[pdf](#)]

Gooskens, C. (2012), "Intelligibility of dialects and related languages", course at the [LOT Winterschool 2012](#), "[Pronunciation differences \(in dialectology and elsewhere\)](#)", Tilburg, January 9-13, 2012. [[pdf](#)]

Heeringa, W. (2012), "Validation", course at the [LOT Winterschool 2012](#), "[Pronunciation differences \(in dialectology and elsewhere\)](#)", Tilburg, January 9-13, 2012. [[pdf](#)]

Hessen, A. van (2012), "Language Resources and Technology Infrastructure for the Humanities and the Social Sciences in the Netherlands", presentation held at the NWO Seminar on Open Access data: your view on "do's and don'ts" for research funders, The Hague, June 4, 2012. [[pdf](#)]

Nerbonne, J. (2012), "Introduction to measuring pronunciation differences, application in dialectometry", course at the [LOT Winterschool 2012](#), "[Pronunciation differences \(in dialectology and elsewhere\)](#)", Tilburg, January 9-13, 2012. [[pdf](#)]

Nerbonne, J. (2012), "Analyzing Linguistic Distances", course at the [LOT Winterschool 2012](#), "[Pronunciation differences \(in dialectology and elsewhere\)](#)", Tilburg, January 9-13, 2012. [[pdf](#)]

Odiijk, J.E.J.M. (2012), "Recent Developments in CLARIN-NL". Presentation at the Eight International Conference on Language Resources and Evaluation (LREC'12) May 23rd, 2012. Istanbul, Turkey. [[pptx](#)]

Odiijk, J.E.J.M. (2012, October 19), Proposal for a Usage Scenario and Theme for presentation at the PMMD conference, Amsterdam. [[doc](#)]

Odiijk, J.E.J.M. (2012, October 23). *Linguistic Research and the CLARIN Infrastructure*. Utrecht, Digital Humanities Lecture. [[ppt](#)]

Pieters, T. and Snelder, S. (2012), "Towards a flexible and stable CLARIN-supported *open-source* web-application for historical sentiment mining in public media", presentation held at Drinks and Drugs in Asia Conference, Shanghai, 22-24 juni, 2012. [[link](#)]



Schuurman, I. (2012), "ISOcat usage in CLARIN NL/VL: results and challenges", presentation held at the ISO-TC 37 DCR meeting, Madrid, June 26, 2012. [[pdf](#)]

Schuurman, I. & M. Windhouwer (2012), ISOcat workshop 2012 part 2, presentation held at the ISOCAT Call 3 Follow-up Workshop, Utrecht October10, 2012) [[ppt](#)]

Schuurman, I. & M. Windhouwer (2012), Call 3 ISOcat follow-up, presentation held at the ISOCAT Call 3 Follow-up Workshop, Utrecht October10,2012) [[pptx](#)]

Swanenberg, J., Cornips, L.M.E.A., Vriend, F. de & Heeringa, W.J. (2012, juli 24). *Is what we have acquired early, less vulnerable to variation? A comparison between data from dialectlexicography and data from first language acquisition*. Wenen, VIIIth Congress of the International Society for Dialectology and Geolinguistics (SIDG). [[pptx](#)]

Wieling, M. and Nerbonne, J. (2012), "Segment Distances & Foreign Accents", course at the [LOT Winterschool 2012](#), "[Pronunciation differences \(in dialectology and elsewhere\)](#)", Tilburg, January 9-13, 2012. [[pdf](#)] [[sound files](#)]

### Meetings

From	To	Event	City	Country	Who	With Whom
18/12/2012	18/12/2012	Meeting CLARIN educatie	Utrecht	NL	JO	Gerrit Bloothoof
11/12/2012	11/12/2012	Bijeenkomst 'Technologie & digitalisering in de wetenschap' op Ministerie van OCW	Den Haag	NL	JO	
30/10/2012	30/10/2012	Digital Humanities curriculum	Utrecht	NL	JO	Gerrit Bloothoof
23/10/2012	23/10/2012	FESLI meeting	Amsterdam	NL	JO	FESLI team
19/10/2012	19/10/2012	Brainstorming PMMD conferentie	Amsterdam	NL	JO	JN et al.
16/10/2012	16/10/2012	Meeting Leo Lentz	Utrecht	NL	JO	Leo Lentz
16/10/2012	16/10/2012	Meeting software metadata	Utrecht	NL	JO	EW
26/09/2012	26/09/2012	Ambassadeursbijeenkomst	Utrecht	NL	JO	UU Congressbureau
14/09/2012	14/09/2012	FESLI meeting	Amsterdam	NL	JO	FESLI team
12/09/2012	12/09/2012	Meeting CLARIAH evaluation	Groningen	NL	JO	Frans Swartz
03/09/2012	03/09/2012	CLARIN ERIC celebration	Utrecht	NL	JO, SK	R.J Smits, CvB



						UU
20/09/2012	20/09/2012	Presentation SURF report	Utrecht	NL	HB	
27/08/2012	27/08/2012	Meeting Computational Humanities	Amsterdam	NL	JO	Rens Bod
22/08/2012	22/08/2012	ISOCAT meeting	Utrecht	NL	JO	IS, DB
21/08/2012	21/08/2012	Meeting Bountouridis	Utrecht	NL	JO	Bountouridis
14/08/2012	14/08/2012	Meeting Metadata4Tools	Utrecht	NL	JO	EW
14/08/2012	14/08/2012	Meeting UU financien	Utrecht	NL	JO	Jan Welling
09/08/2012	09/08/2012	Meeting UBU	Utrecht	NL	JO	Johan Tilstra
26/06/2012	26/06/2012	FESLI meeting	Amsterdam	NL	JO	FESLI team
12/06/2012	12/06/2012	Taalportaal Meeting	Leiden	NL	JO	Taalportaal team
07/06/2012	07/06/2012	NWO bijeenkomst over call CI Topsector	Amsterdam	NL	JO	NWO
05/06/2012	05/06/2012	Cornetto-LMF-RDF Kickoff meeting	Amsterdam	NL	JO	Piek Vossen en team
30/05/2012	30/05/2012	Clariah Meeting	Den Haag	NL	JO	CLARIAH consortium
16/05/2012	16/05/2012	Meeting UBU	Utrecht	NL	JO	Johan Tilstra
08/05/2012	08/05/2012	Clariah Meeting	Den Haag	NL	JO	CLARIAH consortium
18/04/2012	19/04/2012	CLARIN ERIC GA en Coördinators meeting	Den Haag	NL	JO	ERIC GA, NCF
12/04/2012	12/04/2012	Meeting Els Kloek	Den Haag	NL	JO	Els Kloek
05/04/2012	05/04/2012	GrNe Kick-off	Leiden	NL	JO	GrNe team
04/04/2012	04/04/2012	Meeting Huygens ING IVM Centre plan	Utrecht	NL	JO, DB	Ronald Haantjes, Gertjan Filarski
23/03/2012	23/03/2012	MEETING with CLARIN-CZ / LINDAT	Prague	CZ	HB	Jan Hajič, Eva Hajičová

09/03/2012	09/03/2012	Nederlab Site Visit	Amsterdam	NL	JO, HB	Nederlab-team, NWO-commissie
02/03/2012	02/03/2012	Infrastructuur toekenningsbijeenkomst	Utrecht	NL	JO	Halbe Zijlstra
18/01/2012	18/01/2012	GrNe bespreking	Utrecht	NL	JO	M. Treijtel, J. de Jong
18/01/2012	18/01/2012	PILNAR bespreking	Utrecht	NL	JO	Paul Post
09/01/2012	09/01/2012	Clariah Interview	Utrecht	NL	JO, PD, JN. JB	NWO-Commissie
04/01/2012	04/01/2012	Clariah Meeting	Den Haag	NL	JO, PD, JN. JB	Hans Chang

#### Support for other events

From	To	Event	City	Form of Support
29/10/2012	29/10/2012	CLARIN's Turn to the literary text	Den Haag	Subsidy of €2.761
31/01/2012	01/01/2012	Lustrumcongres Alfainformatica en oratie Gertjan van Noord e	Groningen	Subsidy of €1.500
20/01/2012	20/01/2012	CLIN	Tilburg	Subsidy of €2500
30-Nov-12	30-Nov-12	A Shared LMF Core for Sign Language Lexicons	Ravenstein	Subsidy of €6840

#### Support for travel of Dutch researchers to CLARIN-related events

Date	Description	City	Country	Who	Organization
2012-05-21	ISOcat workshop at LREC 2012	Istanbul	TU	Franca Wesseling	MI

2012-05-21	LREC	Istanbul	TU	AvH	UU
2012-07-16	DH 2012	Hamburg	DE	AvH	UU
2012-09-23	International Conference on Theory and Practice of Digital Libraries (TPDL 2012)	Nicosia	CY	Junte Zhang	MI
2012-08-24	EuroHCIR 2012: The 2nd European Workshop on Human-Computer Interaction and Information Retrieval	Nijmegen	NL	Junte Zhang	MI
2012-10-26	CLARIN Annual Meeting	Sofia	BU	EB	multiple
2012-10-26	CLARIN Annual Meeting	Sofia	BU	Marc Kemps-Snijders	MI
2012-10-26	CLARIN Annual Meeting	Sofia	BU	Ineke Schuurman	UU
2012-10-26	CLARIN Annual Meeting	Sofia	BU	Menzo Windhouwer	MPI
2012-10-26	CLARIN Annual Meeting	Sofia	BU	Wim Kok	INL
2012-10-26	CLARIN Annual Meeting	Sofia	BU	Rik Hoekstra	HI
2012-06-19	TKE (Terminology and Knowledge Engineering) Conference	Madrid	ES	Marc Kemps-Snijders	MI
2012-05-21	LREC Tutorial ISOCAT	Istanbul	TU	Ineke Schuurman	UU
2012-06-26	ISO meeting Madrid	Madrid	ES	Ineke Schuurman	UU

### *CLARIN- NL Newsflashes*

Eight CLARIN- NL Newsflashes have been sent out to all registered persons on the following dates (2 more than targeted):

- [NewsFlash 26 \(21-12-2012\)](#).
- [NewsFlash 25 \(24-10-2012\)](#)
- [NewsFlash 24 \(28-09-2012\)](#)
- [NewsFlash 23 \(16-08-2012\)](#)



- [NewsFlash 22 \(02-07-2012\)](#)
- [NewsFlash 21 \(18-06-2012\)](#)
- [NewsFlash 20 \(26-04-2012\)](#)
- [NewsFlash 19 \(24-02-2012\)](#)
- 

All Newsflashes can be accessed via <http://www.clarin.nl/node/82>.

#### *CLARIN-EU Newsflashes*

In 2012, no CLARIN (ERIC) newsflashes have appeared.

#### *CLARIN-EU Newsletter*

In 2012 no CLARIN (ERIC) Newsletters have appeared.

#### *eData & Research*

CLARIN-NL has joined the eData & Research Magazine. CLARIN-NL finds the existence of this magazine important, and aims to make itself more visible by actively participating in its board and editorial team and co-financing it. Several articles on CLARIN-NL have appeared in eData&Research, and CLARIN-NL has continued its participation in 2012. On behalf of CLARIN-NL, Erica Renckens participates in the editorial team and Jan Odijk in the board.

The following articles have been published in eData & Research in 2012. They 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. See <http://www.clarin.nl/node/185> for a list of such videos.

Project	Link to Movie / Screen capture
GabMap	<a href="http://www.youtube.com/watch?v=anElatt5fFM&amp;feature=plcp">http://www.youtube.com/watch?v=anElatt5fFM&amp;feature=plcp</a>
Interviews / Veteran Tapes	<a href="http://www.youtube.com/watch?v=VpCXh_NVVso">http://www.youtube.com/watch?v=VpCXh_NVVso</a>
WFT-GTB	<a href="http://www.youtube.com/watch?v=fevAPknCaDo&amp;feature=player_embedded">http://www.youtube.com/watch?v=fevAPknCaDo&amp;feature=player_embedded</a>

The EB has also commissioned the production of 6 short movies on CLARIN-NL projects (to Dyzlo). Below one can find the CLARIN-NL projects that are 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	<a href="http://www.youtube.com/watch?v=TBDGrqKYbIA">http://www.youtube.com/watch?v=TBDGrqKYbIA</a>
SignLinc	<a href="http://www.youtube.com/watch?v=je3ZKPw6w5s">http://www.youtube.com/watch?v=je3ZKPw6w5s</a>
Arthurian Fiction	<a href="http://www.youtube.com/watch?v=t0y-ecrVD5w">http://www.youtube.com/watch?v=t0y-ecrVD5w</a>
WFT-GTB	<a href="http://www.youtube.com/watch?v=aSWKJucbWY8">http://www.youtube.com/watch?v=aSWKJucbWY8</a>
WIP	<a href="http://www.youtube.com/watch?v=tEqAlH2o2rc&amp;feature=plcp">http://www.youtube.com/watch?v=tEqAlH2o2rc&amp;feature=plcp</a>
Search & Develop	<a href="http://www.youtube.com/watch?v=6CkCfaJ7rao&amp;feature=plcp">http://www.youtube.com/watch?v=6CkCfaJ7rao&amp;feature=plcp</a>

The movies have been made available in streaming mode via a VIMEO and/or a YouTube channel, and the originals have been made available on solid state memory.

### *Website*

A website has been set up for CLARIN-NL. The design and maintenance of the website has been carried out by a company (Clearwind), which also has redesigned the website in 2012, as planned. At the end of 2012, it has been decided to make major upgrade for the website from Drupal 6 to Drupal 7 (in combination with the CLARIN ERIC website)

The CLARIN-NL website is a project-internal website. We also aimed to add an external website, with nice lay-out and pictures, with information on the project targeted at people external to the project, and accompanied with regular PR activities (news messages, showcases, etc.). We have commissioned some work on this (design, editorial work), which has been carried out. This external part of the website will go live after the transition to Drupal 7.

It was also planned to build, in 2012, a portal to the CLARIN infrastructure, which will integrate all the loose bits and pieces (components) that have been and are being developed in the various projects. This portal will give access to increasingly rich and sophisticated infrastructure functionality, where the introduction of each new component will be accompanied by PR-activities. This, however has been delayed due to the delay in the CLAOP plan, and will now be realized only as of 2013.

### *Education*

John Nerbonne and colleagues held a LOT Winterschool course, [Pronunciation differences \(in dialectology and elsewhere\)](#)", including the use of the ADEPT tools (Gabmap).

Various tutorials have been given on CMDI and ISOCAT (see overview events organized by CLARIN-NL).

A [Metadata workshop](#)<sup>16</sup> has been organized at LREC (co-organizer: Daan Broeder), as well as an [ISOCAT Tutorial](#)<sup>17</sup> (organized by Ineke Schuurman).

---

<sup>16</sup> <http://www.lrec-conf.org/proceedings/lrec2012/workshops/11.LREC2012%20Metadata%20Proceedings.pdf>

<sup>17</sup> <http://www.lrec-conf.org/proceedings/lrec2012/index.html>



### Helpdesk

The CLARIN-NL Helpdesk<sup>18</sup> has been set up and has run all year. The amount of requests is rather low, and therefore the Helpdesk was easily managed in 2012 by Jan Odijk and Daan Broeder.

The EB is has proposed to CLARIN-D to cooperate on the Helpdesk, and to combine FAQ sections. Neither a discussion on this nor any concrete actions have been taken in 2012.

A document explaining the concept “CLARIN-compatible” has been written<sup>19</sup> In 2011 (originally specifically targeted at KNAW researchers) The document so far exists only in Dutch, and it was planned to make an English version of it, which, however, still has to be done. It was planned to hold a meeting with the relevant researchers, inviting them to relevant tutorials, etc. to ensure that they indeed understand what it means to produce and deliver resources in a CLARIN-compatible manner, and actually apply this knowledge, but this still has to be done and is in the plan for 2013.

### National Coordination Point

The national coordination point has remained unchanged in 2012. Its members are prof.dr. Jan Odijk (1FTE) as Programme Director and Drs. Jolien Scholten as project secretary (.5 FTE). Drs. Jolien Scholten has announced she will take up a different position as of March 1, 2013, so a replacement will have to be found for her.

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 from 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.

A few members have left the board in 2012:

- Prof.dr. Geert Booij, who also was the CLARIN Board chair (UvL)

---

<sup>18</sup> <http://trac.clarin.nl/trac>

<sup>19</sup> <http://trac.clarin.nl/trac/attachment/wiki/WikiStart/CLARIN%20compatible%20NL%20110805.pdf>



- Prof.dr Aafke Hulk (NIAS, UvA)

New members have joined the CLARIN Board In 2012:\

- Prof.dr. Els Stronks (UU)
- Ir. Peter Wittenburg (MPI)

As of Sep 1, 2012, prof.dr. John Nerbonne (RUG) is the chair of the CLARIN Board.

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 3 meetings in 2012.

**NAP** The National Advisory Panel consisted of 17 members at the start of 2012, and of 16 members at the end of 2012, 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

A few members left the NAP in 2012:

- Prof.dr. Ted Sanders
- Prof.dr. Sally Wyatt
- Prof.dr. Willem Adelaar

Two new members joined the NAP in 2012:

- Prof.dr. Arjen Versloot (UvA)
- Prof.dr. Toine Pieters (UU)

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> ).

The NAP has held 2 meetings in 2012. In one of these meetings the focus was on evaluating the proposals submitted in the fourth call.

**IAP** The International Advisory Panel consisted of 8 members at the start of 2012, and of 11 members since Sep 1, 2012.. 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>).

As was planned, three new members joined the IAP in 2012:

- 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)

The IAP has held 1 meeting in 2012.

**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 2012** The Working Plan for 2012 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 2012 Finances

A detailed report on the CLARIN-NL finances in 2012 can be found in the 2012 Financial Report.

## 5 Other Activities

As a response to the invitation of the *Commissie voor de update van de Nationale Roadmap voor Grootschalige 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. This proposal has been submitted and it has received excellent and good-excellent reviews by external experts. The consortium has written a rebuttal to the comments of the external experts, and it has presented and defended the proposal in an interview with the Committee on January 9<sup>th</sup>, 2012. The project has been put on the Roadmap and has received 1 million euro as 'seed money', to maintain the dynamics of the consortium and to prepare a new, improved, proposal. Plans have been worked out for this CLARIAH seed money project. Details about it can be found on the CLARIAH website ([www.clariah.nl](http://www.clariah.nl))



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