

PROGRESS REPORT 2009-2010 ESFRI PROJECTS

Project name: CLARIN-NL

Project number: 184.021.003

Reporting Period: Jan 1, 2010 t/m December 31, 2010

Report Submitted by (name address data):

Prof. Dr. Jan Odijk
Professor of Language and Speech Technology
CLARIN-NL Programme Director (www.clarin.nl)
UIL-OTS
Utrecht University
Tel (dir): +31 30 253 5745
Tel (secr): +31 30 253 6006
Fax: +31 30 253 6000
Address: Trans 10, 3512 JK Utrecht
The Netherlands
e-mail: j.odijk@uu.nl

PROGRESS REPORT 2009-2010 ESFRI PROJECTS

1. General Developments of the ESFRI Project as a whole in the Reporting Period

Concerning the ESFRI project as a whole, it can be stated that the CLARIN preparatory project is ongoing and on schedule (it is scheduled to run from 2008 until 31 dec 2010). There has been, as in 2009, close collaboration with other infrastructure projects, esp. with DARIAH. One example testifying for this is the [SDH2010 and NEERI conferences](#) organized jointly by CLARIN and DARIAH and held in 2009 and in 2010. The number of participating countries has increased from 23 to 26. All countries involved are working on getting commitments from their national governments to join a CLARIN ERIC to be set up by the end of the preparatory project, and the exact status differs from country to country. The Dutch Minister of Education, Culture and Science, Dr. Ronald Plasterk, has indicated that the Netherlands intends to host the CLARIN ERIC and has invited his colleagues in the other countries to join the CLARIN ERIC. However, in the meantime it has become clear that the CLARIN ERIC will not be realized by the end of the CLARIN preparatory project (end of 2010), and that a more realistic target date for this is by the end of 2011. For this reason the CLARIN European preparatory project has requested and been awarded a project extension with 6 months. The first CLARIN ERIC Steering Committee meetings are expected to take place in the first quarter of 2011.

2. State of Affairs Dutch part of the ESFRI project and modifications, if any

EU-Line

It was planned to spend the money budgeted for this item depending on the set-up of the CLARIN ERIC. However, since it is clear that the CLARIN ERIC will not come into existence before 2011, no money from this budget has been spent. It will be shifted to the next year and be used to pay the ERIC fee for that year and possibly to start up specific CLARIN-EU-oriented activities, e.g. to bridge a gap between the end of the CLARIN preparatory project and the start of the CLARIN ERIC.

NL-Line

Technical

Project Metadata Project (MD)

Coordinator Ir. Daan Broeder (MPI Nijmegen)

Budget 75k euro

Status The Metadata project, which had already started in 2009, has finished and has delivered a CMDI registry which allows one to search for existing metadata components and profiles, or create new ones if needed. The project has created several metadata components and profiles for data that are available in the Netherlands. For the metadata elements contained in the metadata components links have been made to the ISOCAT data category registry, and where needed this registry has been extended, in order to ensure semantic interoperability. More detailed information on this project and its results, including all deliverables, can be found on the CLARIN-NL web site (Metadata project area) and on the CMDI registry site (<http://www.clarin.eu/cmdl>).

Project Infrastructure Implementation Plan (IIP)

PROGRESS REPORT 2009-2010 ESFRI PROJECTS

Coordinator Ir. Daan Broeder (MPI Nijmegen)

Budget 798k euro

Status A revised version of the Infrastructure Implementation Plan (IIP) with a better balance of the responsibilities and work between the participating candidate centres was submitted to the CLARIN-NL board and approved in February 2010. Participants in this subproject are the centres that have announced to become a CLARIN type A/B centre in the Netherlands, i.e. MPI (Nijmegen), Meertens Institute (Amsterdam), INL (Leiden), and DANS (The Hague). The project effectively started in September 2010. It will run for three years.

Project Search & Develop (S&D)

Coordinator Ir. Marc Kemps-Snijders (Meertens Institute Amsterdam)

Budget 576k euro

Status A plan for creating a network of centres in the Netherlands, focusing around search services (called Search & Develop, abbreviated S&D) has been submitted to the Board in early 2010 and approved. Due to staff hiring problems, this project effectively started in September 2010 and it will run for 3 years. A workshop dedicated to setting up a European-wide network of centres around the CLARIN European Demonstrator Case (EDC) has been held on January 25-26, 2010 at the Meertens Institute in Amsterdam, the Netherlands.

In the meantime, new organizations have expressed interest to become a CLARIN-centre. The Huygens Institute expressed this, as well as Beeld & Geluid. Discussions are also on-going with the Koninklijke Bibliotheek (Royal Library), and are being initiated with the Nationaal Archief (National Archive). At the same time, DANS appears to express the ambition to take on a larger role in the CLARIN- centre landscape than they had so far. It will be necessary in 2011 to investigate how all these new centres and the possibly changing roles of existing centres can best be accommodated.

Data, Tools and Services

First Open Call Projects

Most projects of the first open call were approved in 2009, started early in 2010 and have in the meantime finished or are about to finish.

Two projects, ADEPT and WFT-GTB were approved by the CLARIN-NL Board on Feb 15, 2010. Both these projects have been carried out and finished.

- **ADEPT** - Assaying Differences via Edit-Distance of Pronunciation Transcriptions
 - **Project Coordinator:** [Dr. Charlotte Gooskens](#) (Groningen University)
 - **Goal:** The goal of the project is to provide a web application capable of measuring the differences in sets of phonetic (or phonemic) transcriptions via edit distance.
 - **Budget** 57k euro
 - **Scientific impact:** A user-friendly web application version of this existing tool will make it usable for a wide range of humanities researchers such as, phonologists, phoneticians, dialectologists and second-language learning specialists.
 - **Status** The project has effectively finished and has submitted its final reports.
- **AAM-LR** - Automatic Annotation of Multi-modal Language Resources
 - **Project Coordinator:** [Prof.dr. L. Boves](#) (Radboud University Nijmegen)
 - **Goal:** The AAM-LR project aims at building a demonstrator of a web service that will help field researchers to annotate audio- and video-recordings
 - **Budget** 56k euro
 - **Scientific impact:** It will facilitate all field research in which digital audio or video recordings are used.

PROGRESS REPORT 2009-2010 ESFRI PROJECTS

- **Status** The project has effectively finished the core functionality but has requested an extension in order to optimally implement the web service and is now targeted to finish by June 30, 2011.
- **Adelheid** - A Distributed Lemmatizer for Historical Dutch
 - **Project Coordinator:** [Dr. H. van Halteren](#) (Radboud University Nijmegen)
 - **Goal:** This project aims at providing a web-application with which an end user can have historical Dutch text tokenized, lemmatized and part-of-speech tagged.
 - **Budget** 55k euro
 - **Scientific impact:** It will facilitate research on historical texts in fields such as historical linguistics, literature science and history.
 - **Status** The project has effectively finished and is about to submit the final reports. The Second Call project INPOLDER will make use of the results of Adelheid.
- **DUELME-LMF** - Converting DUELME into LMF format
 - **Project Coordinator:** [Prof.dr. J.E.J.M. Odijk](#) (Utrecht University)
 - **Goal:** The goal of the project is develop converters between DUELME and LMF format, and to create a curated DUELME resource fully compliant with standards supported by CLARIN.
 - **Budget** 60k euro
 - **Scientific impact:** The DUELME database makes it possible to address a variety of research questions related to Dutch multi-word expressions in research areas such as computational linguistics and natural language processing, theoretical linguistics and psycholinguistics.
 - **Status:** The project has effectively finished and is about to submit the final reports.
- **INTER-VIEWS** - Curation of Interview Data
 - **Project Coordinator:** [Dr. H. van den Heuvel](#) (Radboud University Nijmegen)
 - **Goal:** The INTER-VIEWS project will make a corpus of interview data available to the community of researchers in the humanities.
 - **Budget** 58k euro
 - **Scientific impact:** This project will contribute to facilitating all research that makes use of interviews, and more specifically research into the Second World War.
 - **Status** The project has effectively finished and is about to submit the final reports.
- **MIMORE** - Microcomparative Morphosyntax Research Tool
 - **Project Coordinator:** [prof. dr. Sjef Barbiers](#) (Meertens Institute and Utrecht University)
 - **Goal:** The demonstrator tool MIMORE will create a common search engine for the databases [DynaSAND](#), [DiDDD](#) and [MAND](#)
 - **Budget** 54k euro
 - **Scientific impact:** The tool is especially relevant to research in variation linguistics. It will allow investigating theoretical questions concerning the language system and language variation, and the geographic distribution of (morpho-)syntactic variables.
 - **Status** The project has effectively finished and will submit its final reports by March 2011..
- **Sign-LinC** - Linking lexical databases and annotated corpora of signed languages
 - **Project Coordinator:** [Dr. O.A. Crasborn](#) (Radboud University Nijmegen)
 - **Goal:** This project aims to link two independently evolved data sets for a signed language: the Corpus NGT and the lexical database of the Dutch Sign Centre.
 - **Budget** 56k euro
 - **Scientific impact:** The linked databases and corpora will facilitate research in the area of sign language linguistics
 - **Status** The project has effectively finished and is about to submit the final reports.
- **TICCLops** - Text-Induced Corpus Clean-up online processing system
 - **Project Coordinator:** Dr. [Martin Reynaert](#) (Tilburg University)
 - **Goal:** This demonstrator project will develop a tool that allows CLARIN users to submit their corpora for fully automatic spelling correction and normalization
 - **Budget** 54k euro

PROGRESS REPORT 2009-2010 ESFRI PROJECTS

- **Scientific impact:** The tool will facilitate all research that makes use of text corpora.
- **Status** The project has effectively finished and is about to submit the final reports.
- **TDS Curator** - A web-services architecture to curate the Typological Database System
 - **Project coordinator:** [Dr. Alexis Dimitriadis](#) (Utrecht University)
 - **Goal:** TDS Curator will make the Typological Database System (TDS) into a sustainable service that conforms to CLARIN infrastructural requirements.
 - **Budget** 57k euro
 - **Scientific impact:** The (TDS) facilitates typological linguistic research by providing integrated access to multiple independently developed typological databases through a common web interface
 - **Status** The project had some changes in partners and has therefore been assigned a later end date (now targeted for May 31, 2011).
- **TQE** - Transcription Quality Evaluation
 - **Project Coordinator:** [Dr. H. Strik](#) (Radboud University Nijmegen)
 - **Goal:** The TQE project aims to create a completely automatic Transcription Quality Evaluation (TQE) tool.
 - **Budget** 57k euro
 - **Scientific impact:** The tool will be useful for validating, obtaining, and selecting phone transcriptions, for detecting phone strings (e.g. words) with deviating pronunciation, and, in general, it can be usefully applied in all research - in various (sub-)fields of humanities and language and speech technology in which audio and phonetic transcriptions are involved.
 - **Status** The technical aspects of the project have largely been finished. An extension was requested to consult the user community and make adaptations depending on user feedback. The project is targeted to finish by June 30, 2011.
- **WFT-GTB** - Integrating the *Wurdboek fan 'e Fryske Taal* into the *Geïntegreerde Taalbank*
 - **Project Coordinator:** [drs. Hindrik Sijens](#) (Fryske Akademy)
 - **Goal:** The project aims to carry out data curation of the *Wurdboek fan de Fryske Taal (WFT)* database and integrate the data in the *Geïntegreerde Taalbank (GTB)* and demonstrate it via the GTB dictionary web application.
 - **Budget** 56k euro
 - **Scientific impact:** Integrating the WFT dictionary in the GTB will facilitate studying formal, semantic and idiomatic aspects of Modern West Frisian by allowing extensive forms of free and structured (in particular thematic) search queries, and it will enable comparative studies with Dutch materials.
 - **Status** The project has effectively finished and demonstrated its results at [Euralex 2010](#) held in Leeuwarden. It is still to submit its final report.

CKCC project The CKCC project (*Circulation of Knowledge and Learned Practices in the 17th-century Dutch Republic*, '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 17th 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).¹

Its proposal for an adaptation of the current project that would make its way of working CLARIN-compatible was approved by the Board in February 2010 (Budget: 59k euro). The project will run until Nov 1, 2012.

¹ http://www.clarin.eu/files/cnl06_web_0.pdf

PROGRESS REPORT 2009-2010 ESFRI PROJECTS

Status The project is running according to schedule.

Cooperation between Netherlands and Flanders A plan has been developed, together with CLARIN Flanders, for a cooperation project between the Netherlands and Flanders. The overall plan 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 Technical) 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. Flanders contributes in this part 452k Euro in accordance with the usual NL:2/3-VL1/3 proportion.
- 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 been approved by Flanders in December 2009 and by the CLARIN-NL Board in February 2010. It has started in March 2010 and will run for app. 2.5 years.

Status Though there were some start-up delays, the project has started and is running well, though a number of changes were needed in the overall planning. It is still targeted to finish by the end of 2012.

Second Open Call An second open call, similar to the first open call of 2009, has been organized in 2010 for data curation and demonstrator projects. It focused on the disciplines of literary studies, history and political studies, communication and media studies, language acquisition and historical linguistics. In the approved projects data and tools developed and in use in the Netherlands will be curated and a series of demonstrators will be developed on the basis of existing technologies and applications to attempt to make them CLARIN-compatible. A major aim with these projects is to test the standards proposed within CLARIN against these data and tools, and to be able to provide evidence-based feedback on the proposed standards and thus to influence their determination in the CLARIN context.

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

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

Each project lasts maximally one year and has a maximum budget of 120k euro. The total available budget for this open call was originally set at 600K euro.

PROGRESS REPORT 2009-2010 ESFRI PROJECTS

As a result of the call 17 project proposals were submitted. Of these, 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 order to accommodate all 9 projects, the budget was increased slightly to 628k euro, which could easily be accommodated in the overall budget.

Detailed reports on the procedure followed for evaluating the second open call by IAP, NAP and EB have been made and have been approved by the Board on December 15th, 2010.

An overview of the awarded 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, and a very short description of the scientific impact of the project. More detailed description of the projects can be found on the CLARIN-NL website²

- **ArthurianFiction** (*Arthurian Fiction in Medieval Europe*)
 - **Project Coordinator:** Dr. Bart Besamusca (Utrecht University)
 - **Goal** ArthurianFiction is a mixed demonstrator and data curation project that aims to carry out data curation for two databases with data for literary research in the area of European Arthurian fiction and to create a demonstrator that will allow users access to the data for searching and editing.
 - **Budget** 47k euro
 - **Scientific Impact** It will enable access to the Arthurian Fiction database to a wide range of humanities researchers such as literary historians in all European languages, cultural historians, medieval book specialists and art historians.
- **C-DSD** (*Curating the Dutch Song Database*)
 - **Project Coordinator:** Dr. E. Stronks (Utrecht University)
 - **Goal** C-DSD is a data curation project that aims to curate the Dutch Song Database (DSD), a database in the field of Literary Studies containing (meta-) data on 140.000 songs and their 15.000 sources from the Middle Ages to the present day.
 - **Budget** 29k euro
 - **Scientific Impact** This project will create new opportunities for (international) collaboration with other projects with similar source materials.
- **COAVA** (*Cognition, Acquisition and Variation Tool*)
 - **Project Coordinator** Dr. Leonie Cornips (Meertens Institute Amsterdam)
 - **Goal** COAVA is a mixed demonstrator and data curation project proposal in which a tool is developed for easily exploring the linguistic characteristics of objects from two databases of language variation (historical dialectology) on the one hand and three first language acquisition databases on the other hand.
 - **Budget** 66k euro
 - **Scientific Impact** It will enable interdisciplinary research into the relation between language acquisition and language variation.
- **INPOLDER** (*Integrated Parser and Lemmatizer Dutch in Retrospect*)
 - **Project Coordinator** Prof. Dr. Ans van Kemenade (Radboud University Nijmegen)
 - **Goal** INPOLDER is a mixed demonstrator and data curation project that aims to provide syntactically analysed corpus material for historical Dutch. It uses the Adelheid tagger for PoS-tagging and a variant of the Penn-Helsinki parser for historical texts for enriching the texts with parse trees.

² <http://www.clarin.nl/node/70>

PROGRESS REPORT 2009-2010 ESFRI PROJECTS

- **Budget** 69k euro
- **Scientific Impact** It will fill the gap in availability of syntactically analysed corpus material for historical Dutch, enabling historical syntactic, morpho-syntactic and lexical research as well as benefiting non-linguistic historical research.
- **IPROSLA** (*Integrating and publishing resources on sign language acquisition*)
 - **Project Coordinator** Prof. dr. P. (Paula) Fikkert (Radboud University Nijmegen)
 - **Goal** IPROSLA is a resource curation project that aims to integrate two different data sets on sign language acquisition by documenting the two data sets with CMDI, and archiving them at the MPI language archive. The data consist of a set of longitudinal data of deaf children from deaf and hearing parents collected at the UvA, and a new collection of longitudinal data collected at the RU from hearing children of deaf parents.
 - **Budget** 67k euro
 - **Scientific Impact** By integrating two datasets in integrated archive and making them visible and accessible, it will enable better and more research on Sign Language.
- **NEHOL** (*Negerhollands Database*)
 - **Project Coordinator** Prof. dr. P.C. Muysken (Radboud University Nijmegen)
 - **Goal** NEHOL is a data curation project that aims to make available to the CLARIN community the data from the Dutch-lexifier Creole language Negerhollands, the now extinct Creole language of the Virgins Islands.
 - **Budget** 43k euro
 - **Scientific Impact** It will enable research into Negerhollands, and thereby into Creole languages and language creolization processes in general.
- **VU-DNC** (*VU Diachronic Newspaper Corpus*)
 - **Project Coordinator** Prof.dr. W. Spooren (Free University Amsterdam)
 - **Goal** VU-DNC is a data curation project that aims to make a unique diachronic corpus of Dutch newspaper articles from five major Dutch newspapers from 1950/1951 and 2002 (2 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.
 - **Budget** 59k euro
 - **Scientific Impact** It will enable research into the recent history and development of the Dutch written language, as well as research into subjectivity and quotations and the lexico-grammatical features encoding them.
- **WAHSP** (*Web-application for historical sentiment mining in public media*)
 - **Project Coordinator** Prof.dr. Toine Pieters (Utrecht University)
 - **Goal** WAHSP is a demonstrator project that aims to provide advanced forms of text mining (more specifically, sentiment mining) in large historical datasets of newspapers and journals in the form of a CLARIN compliant web-application, addressing research questions of historians and policy researchers.
 - **Budget** 120k euro
 - **Scientific Impact** 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.
- **WIP** (*War in Parliament*)
 - **Project Coordinator** Dr. Hinke Piersma (NIOD)
 - **Goal** WIP is a mixed demonstrator and data curation project that aims to make the Dutch Hansard database compliant with the CLARIN infrastructure and to provide an advanced search engine for it in order to make it possible to carry out historical and social science research, in particular with respect to the Second World War.

PROGRESS REPORT 2009-2010 ESFRI PROJECTS

- **Budget** 118k euro
- **Scientific Impact** It will enable historical research by providing tools for text mining large text corpora, applied in this project in particular for references to the Second World War in parliamentary debates.

Directly assigned projects

It was planned to select a few projects that are considered to be of great importance for involving the target user community and fund them directly (outside of the call for projects). The selection would be based on the results of the user survey. However, since the user survey had some delay (see next section), the selection of such projects has also been delayed, and will now be done in 2011.

User Survey

The plan for carrying out a survey of user needs was approved by the Board in February 2010.

The set-up of this user survey is such that it can also serve as a 0-measurement which aims to measure the current state with regard to the use and ease of use of data and tools by researchers from linguistics and the humanities more broadly.

In addition, the survey, which will be carried out by initiating face-to-face dialogues with researchers from targeted research organizations, will provide opportunities to make CLARIN and the advantages it might offer better known in the humanities research community.

The plan has been set up by Dr. Ton van der Wouden, in collaboration with the CLARIN-NL executive board. The plan has been carried out by Dr. Arjan van Hessen and Dr. Jenny Audring.

Partially due to the changes in staff, the user survey has taken longer than originally anticipated. Though the user survey was originally planned to finish before the summer holidays, an intermediate version of the results was submitted just before the summer holidays, and the final version was made available in December 2010 and discussed and approved at the CLARIN-NL Board meeting of December 15th, 2010. The report and its appendix can be found on the CLARIN-NL website, more specifically [here](#).

Centres of Expertise

There are still many legacy data that do not comply to internationally recognized standards and best practices, let alone to CLARIN-specific standards. These data are thus not visible and not easily accessible to researchers. In order to remedy this, the EB requested Dr. Nelleke Oostdijk (Radboud University Nijmegen) to develop a plan for a Data Curation Service (DCS). This service should provide a systematic way of curating data, maximally using the available knowledge, expertise, methodologies and tools already available at the candidate CLARIN-centres and other relevant organizations, and with minimal involvement of the researchers who created the data. The work on this plan started in the fall of 2010 and is expected to end by February 2011. Intermediate results are available on the CLARIN-NL website, [Data Curation Service section](#). The plan is supposed to lead to a virtual centre of expertise on data curation, which, when approved by the CLARIN-NL Board, will start operations in 2011.

PROGRESS REPORT 2009-2010 ESFRI PROJECTS

Education and Awareness

An overview is given of the various activities for stimulating education, training, and awareness carried out in 2009.

Conferences and workshops visited

From	To	Event	City	Country	Who
25/1/2010	26/1/2010	Workshop on CLARIN EDC (European Demonstrator Case)	Amsterdam	NL	HB
05/02/2010	05/02/2010	CLIN 20	Utrecht	NL	JO, AvH
16/05/2010	23/05/2010	LREC 2010	Valletta	MA	JO
11/06/2010	11/06/2010	Symposium "Toekomst voor Digitale Geesteswetenschappen"	Utrecht	NL	JO, AvH
17/06/2010	18/06/2010	NO-CLARIN	Oslo	NO	JO
28/06/2010	28/06/2010	Ed Hovy Symposium	Utrecht	NL	JO
01/07/2010	01/07/2010	DigiHist: Studiedag Digitalisering Historische Handschriften	Leuven	BE	JO
07/07/2010	08/07/2010	Digital Humanities 2010	London	UK	JO
13/09/2010	13/09/2010	BIG Grid User event	Amsterdam	NL	JO, DB
18/10/2010	22/10/2010	SDH2010 + NEERI	Vienna	AT	JO, DB, AvH, HB
02/12/2010	02/12/2010	Conferentie Grootchalige onderzoeksfaciliteiten	Leiden	NL	JO, AvH

Events organized by CLARIN-NL

From	To	Event	City	Country
19/02/2010	19/02/2010	1e CLARIN-NL Meeting	Utrecht	NL
08/01/2010	08/01/2010	Relation Registry Workshop	Nijmegen	NL
25/03/2010	25/03/2010	ISOCAT tutorial	Utrecht	NL
27/05/2010	27/05/2010	Metadata Tutorial	Nijmegen	NL
26/08/2010	26/08/2010	CLARIN-NL Info Session Second Call	Amsterdam	NL
21/09/2010	21/09/2010	ISOCAT workshop	Nijmegen	NL
28/10/2010	28/10/2010	2e CLARIN-NL Meeting	Utrecht	NL

Meetings

From	To	Event	City	Country	Who
05/01/2010	05/01/2010	KNAW / CLARIN-NL	Amsterdam	NL	JO, JN. GB
28/01/2010	28/01/2010	LOT-Bestuursvergadering	Utrecht	NL	HB

PROGRESS REPORT 2009-2010 ESFRI PROJECTS

12/02/2010	12/02/2010	CLARIN – T4ME (META-NET) Meeting (at FLaReNet Forum}	Barcelona	NL	JO
17/03/2010	17/03/2010	Meeting with Bart Besamusca	Utrecht	NL	JO
26/04/2010	26/04/2010	Huygens	Den Haag	NL	JO
28/05/2010	28/05/2010	Preparation Symposium “Toekomst Digitale GeestesWetenschappen”	Utrecht	NL	JO
02/06/2010	02/06/2010	IIP Meeting	Utrecht	NL	JO
23/06/2010	23/06/2010	eData & research Meeting (Martijn de Groot)	Utrecht	NL	JO
24/06/2010	24/06/2010	MIMORE Meeting	Utrecht	NL	JO
30/06/2010	30/06/2010	SURFNet Meeting	Utrecht	NL	JO, DB
24/08/2010	24/08/2010	DEN Meeting	Den Haag	NL	JO, DB, AvH
01/09/2010	01/09/2010	ADEPT Meeting	Groningen	NL	JO
08/09/2010	08/09/2010	CLARIN NTU ERIC	Roosendaal	NL	JO
16/09/2010	16/09/2010	Meeting with Wijnand Mijnhardt about DGO	Utrecht	NL	JO
22/09/2010	22/09/2010	Meeting with José de Kruif	Utrecht	NL	JO, AvH
23/09/2010	23/09/2010	IIP + S&D Meeting	Utrecht	NL	JO
04/10/2010	04/10/2010	TDS Project	Den Haag	NL	JO
06/10/2010	6/10/2010	Libratory	Leiden	NL	JO
13/10/2010	13/10/2010	Metadata ISO	Berlin	DE	JO, DB
26/10/2010	26/10/2010	DGO	Utrecht	NL	JO
09/11/2010	09/11/2010	CMDI-DEN meeting	Nijmegen	NL	JO, DB
10/11/2010	10/11/2010	Beeld & Geluid	Hilversum	NL	JO, AvH
22/11/2010	22/11/2010	DGO	Utrecht	NL	JO
25/11/2010	25/11/2010	IIP + S&D Meeting	Utrecht	NL	AvH
8/12/2010	8/12/2010	Meeting with Steven Krauwer	Utrecht	NL	JO, DB

Helpdesk

A [HelpDesk](#) for CLARIN-NL has been set up and is functioning. Its main goal is to assist researchers in general and participants in CLARIN-NL projects in particular in technical matters. A tracking system has been set up, and a special e-mail address to contact the helpdesk. The Helpdesk website contains sections with a large number of Frequently Asked Questions (FAQ) on technical matters.

All text on the Helpdesk has been written in English, so that all the information found there is also useful for researchers who do not speak Dutch, both in the Netherlands and abroad. We have the ambition to turn the CLARIN-NL helpdesk into a helpdesk for the whole CLARIN infrastructure once the CLARIN ERIC has been set up. To that end it is planned to evaluate what the requirements for a helpdesk are as formulated in the CLARIN preparatory project. We expect the report on these requirements early in 2011.

The project leader for the HelpDesk is Daan Broeder. Paul van Caspel has implemented the system and is the HelpDesk employee. The HelpDesk employee is backed by a range of infrastructure specialists, i.e. persons with excellent knowledge and expertise on specific matters relating to the technical infrastructure, who have been asked to assist the helpdesk employee for more difficult problems and committed to do so.

PROGRESS REPORT 2009-2010 ESFRI PROJECTS

The total budget for the HelpDesk was 30k euro, covering the period from mid 2010 to mid 2011.

Support for other events

Date	Description	City	Country	Form of Support
5/2/2010	CLIN20	Utrecht	NL	Funding
4-5 Aug 2010	Relish Workshop	Nijmegen	NL	Funding travel and subsistence of some participants

Support for travel of Dutch researchers to CLARIN-related events

Date	Description	City	Country	Who	Organization
19-feb 2010	1e CLARIN-NL Meeting	Utrecht	NL	F. de Vriend	Meertens
17/5/2010	LREC Workshop Malta	Valletta	Malta	F. de Vriend	Meertens
12-14/10/2010	SDH2010	Vienna	AT	S. Scagliola	Veteraneninstituut
25-26/11/2010	Workshop <i>Nya perspektiv på dialekter</i>	Helsinki	FI	T. Leinonen	Groningen University

CLARIN- NL Newsflashes

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

- [NewsFlash 12 \(25-11-2010\)](#)
- [NewsFlash 11 \(13-10-2010\)](#)
- [NewsFlash 10 \(23-09-2010\)](#)
- [NewsFlash 9 \(30-08-2010\)](#)
- [NewsFlash 8 \(06-07-2010\)](#)
- [NewsFlash 7 \(27-04-2010\)](#)
- [NewsFlash 6 \(05-03-2010\)](#)
- [NewsFlash 5 \(06-01-2010\)](#)

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

CLARIN-EU Newsflashes

Contributions from CLARIN-NL have been included in the following CLARIN EU Newsflash (<http://www.clarin.eu/clarin-members/members-documents/clarin-newsflash>):

- December 2010 (not accessible online):

To the best of our knowledge, only 3 EU newsflashes appeared in 2010.

CLARIN-EU Newsletter

One can find the CLARIN-EU newsletters here: <http://www.clarin.eu/newsletter>.

PROGRESS REPORT 2009-2010 ESFRI PROJECTS

An article on CLARIN-NL has appeared in CLARIN-EU Newsletter 9-10., which can be found here:

http://www.clarin.eu/system/files/newsletter/cnl09-10_web.pdf :

Website

The website has been extended systematically. Each project was assigned an own project area, with restricted access. In 2010, an external company was hired to make a major update and add some essential functionality. Web forms are fully activated and in use now. Forum functionality has been activated but has not been used so far.

The website still can and must be improved with additional functionality and better lay-out, which is planned for 2011.

Leaflet and Booklet

A leaflet with general information on the CLARIN-NL project and the running subprojects was created early in 2010 and it has been distributed at various events.

In 2010 a start was made with the creation of a booklet with descriptions of the intermediate and/or final results of the projects started up in 2010, and a short description of the Call 2 projects to start up in 2011. The booklet has become available in February 2011.

Paper versions of the leaflet and booklet are available upon request. PDF versions of the leaflet and the booklet can be obtained [here](#).

3. International positioning

The Netherlands is the first country that has committed itself to funding CLARIN nationally, it has expressed that it intends to host the CLARIN ERIC, and the minister of Education, Culture and Science has invited his colleagues from other countries to join the CLARIN ERIC.

In the CLARIN preparatory project the Max Planck Institute for Psycholinguistics (MPI, Nijmegen) plays a leading role for the more technical aspects of the infrastructure (design decisions, specifications and prototype implementations based on inventoried user requirements), and Utrecht University coordinates the preparatory project. It is clear, therefore, that the Netherlands is the leading country in the set-up of the CLARIN infrastructure, both organizationally and technically. This position has only been further strengthened by the national activities that have now been started up and partially finished already. In particular projects dealing with the actual implementation of the technical infrastructure, such as the MetaData project, the infrastructure implementation project (IIP), and the Search & Development project (S&D) are increasing the importance of the Netherlands in Europe and the impact of the technologies it is developing.

4. Governance and management

The national coordination point, set up in 2009, remained unchanged in 2010. Its members are prof.dr. Jan Odijk (1FTE) as Programme Director and Drs. Erica Renckens as project secretary (.5 FTE).

Since Erica Renckens indicated to have other plans for 2011, a replacement for her was looked for and found. Drs. Jolien Scholten is Erica's successor as project secretary, and she has started working since December 1, 2010.

PROGRESS REPORT 2009-2010 ESFRI PROJECTS

The governance structure had been largely implemented in 2009, and has seen a few changes. It 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.

The EB, Board and NAP held a joint initial meeting on May 27, 2009, preceding the CLARIN-NL kick-off meeting.

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. Peter Wittenburg (until March 1, 2010) / Ir. Daan Broeder (since Feb 1, 2010), technical director

Peter Wittenburg indicated that he was too busy to keep his role as technical director in the CLARIN-NL EB, and requested to be dismissed. This was allowed and he was replaced by Ir. Daan Broeder.

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

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

Board The Board consists of 8 members. A full list of its members can be found in Appendix A and on the CLARIN-NL website (<http://www.clarin.nl/node/15>).

The Board has held 3 meetings in 2010 (Feb 2, Jun 29, and Dec 15).

NAP The National Advisory Panel consists of 17 members, providing a good representation of the fields of the humanities, language and speech processing and infrastructure design and implementation. A full list of the members can be found in Appendix B and on the CLARIN-NL website (<http://www.clarin.nl/node/16>). The NAP chair, Dr. Willemijn Heeren, requested to be dismissed from her tasks as chair since she was moving abroad. She was succeeded by Dr. Karina van Dalen-Oskam.

The NAP has held four meetings in 2010. (Feb 16, Apr 8, Jun 17, Nov 30)

IAP Thanks to an intensification of the actions by the EB to invite members for the IAP, seven persons have joined the IAP. The full composition can be found in Appendix C and on the CLARIN-NL Website (<http://www.clarin.nl/node/52>)

One additional member is needed for the IAP. The EB has candidate names but did not make any invitations yet. The IAP has met once in 2010 (Oct 29, 2010, in Utrecht)

Consortium Agreement The consortium agreement was originally signed by 21 partners. The consortium is open to new partners, and two new partners (NIOD and Aletta) have joined the consortium already in 2009. (See <http://www.clarin.nl/node/7>). In 2010, *Dedicon* has joined, and *Erasmus Universiteit* is in the process of signing the consortium agreement. Some participants in the awarded Open Call 2 projects are required to join the consortium, which will happen in 2011. (*ISLA (UvA)*, *Hogeschool Utrecht*) The consortium agreement can be found on the CLARIN-NL website (login required): <http://www.clarin.nl/node/72>.

PROGRESS REPORT 2009-2010 ESFRI PROJECTS

CLARIN-NL Long Term Working Plan The CLARIN-NL Long Term Working Plan as approved by the Board on August 19, 2009 has remained unchanged in 2010. It can be found on the CLARIN-NL website: under Documents: <http://www.clarin.nl/node/47> :

Working Plan 2010 The Working Plan for 2010 has been made and was approved by the Board on Feb 2, 2010. It can be found on the CLARIN-NL website under Documents: <http://www.clarin.nl/node/47> :

Overall, the governance structure as set up functions appropriately and there are no reasons to make any modifications.

5. Financial overview

For 2010:

2010	received budget	expenses	Remaining budget
Total	€ 1,350,000	€ 1,327,489.31	€ 22,510.69

For 2009+2010

2009+2010	received budget	expenses	Remaining budget
Total	€ 2,700,000	€ 1,513,712.31	€ 1,186,287.69

More details on the financial overview can be provided on request.

6. Results of the Facility (research carried out and dissemination of results such as publications, lectures, ...)

Presentations by the CLARIN-NL Governance members

Bennis, H.J. (2009, June 04). *Digitale Geesteswetenschappen; Hoe en Waarom*. Amsterdam, KNAW Symposium Computational Humanities.

Bennis, H.J. (2009, June 26). *Geography, databases, theoretical linguistics and CLARIN-NL*. Copenhagen, Denmark, ICLAVE 5.

Bennis, H.J. (2009, August 27). *CLARIN-NL and Dialect Data*. Alvdalen, Sweden, Grand Meeting ScanDiaSyn.

Bennis, H.J. (2010, June 8). *Linguistics, databases and Open Access*. Trømsø, Norway, Grand Meeting ScanDiaSyn.

Bennis, H.J. (2010, October 19). *Syntactic Variation*. Vienna, Workshop Language Variation and Digital Technology, SDH.

Broeder, D. & P. van Caspel (2010), "CLARIN-NL Helpdesk", presentation held at the CLARIN-NL Second Open Call Info Session, Amsterdam, August 26, 2010. [[pdf](#)]

Hessen, A. van, H. Bennis & T. van der Wouden (2010), "Tools en data voor taalwetenschappers", presentation held at TIN 2010, Utrecht, Feb 6, 2010.

Hessen, A. van (2010), "CLARIN", presentation held at the [LOT Winterschool](#), Utrecht, January, 2010 [[ppt](#)]

Krauwier, S. (2010), "CLARIN-EU: Waar staan we?", presentation held at the CLARIN-NL Second Open Call Info Session, Amsterdam, August 26, 2010 [[pdf](#)]

Odijk, J. (2010), "CLARIN-NL", presentation held at CLIN 20, February 5, 2010, Utrecht [[ppt](#)]

Odijk, J. (2010), "The CLARIN-NL Project", Presentation held at LREC 2010, Valletta, Malta, May 19, 2010. [[ppt](#)]

PROGRESS REPORT 2009-2010 ESFRI PROJECTS

- Odijk, J. (2010), "Taal- en Spraaktechnologie in een Infrastructuur voor Geesteswetenschappers", Presentation held at the symposium *Een toekomst voor digitaal geesteswetenschappelijk onderzoek*, Utrecht, June 11, 2010. [[ppt](#)]. See also [here](#)
- Odijk, J. (2010), "CLARIN-NL: Language Resources and Technology Infrastructure for the Humanities in the Netherlands", presentation held at the NO-CLARIN Meeting, Oslo 18 June 2010. [[ppt](#)]
- Odijk, J. (2010), "CLARIN-NL: What is going on?", presentation held at the CLARIN-NL Second Open Call Info Session, Amsterdam, August 26, 2010. [[pdf](#)]
- Odijk, J. (2010), "CLARIN-NL Second Open Call", presentation held at the CLARIN-NL Second Open Call Info Session, Amsterdam, August 26, 2010. [[pdf](#)]
- Odijk, J. (2010), "The CLARIN-NL Project", CLARIN Newsletter 9-10, March-June 2010, pp 3-4. <http://www.clarin.eu/newsletter/3280> [[pdf](#)]
- Odijk, J. (2010). Contribution to the Panel Session on Open and Persistent Access, SDH 2010, Vienna, October 20, 2010. [[pdf](#)]

Presentations by others

- Bos, E.-J. (2010), "Circulation of Knowledge and Learned Practice", presentation held at the CLARIN-NL Second Open Call Info Session, Amsterdam, August 26, 2010. [[pdf](#)]
- Grégoire, N. (2010), "DuELME LMF project". presentation held at the Workshop "D-Spin meets RELISH: standardizing lexicon structures", April 26, 2010, University of Frankfurt. [[pdf](#)]
- Grégoire, N. (2010), "En Garde Project. The redesign of a Dutch Electronic Lexicon of Multiword Expressions.", presentation held at the workshop "Lexicon Tools en Standards", August 4, 2010, Max Planck Institute, Nijmegen. [[pdf](#)]
- Heuvel, H. van den (2010), "INTER-VIEWS, Curation of interview data", presentation held at the ISOcat Workshop, Nijmegen, September 21, 2010 [[ppt](#)]
- Horik, R. van (2010), "Curation Project 1e Call: INTER-VIEWS", presentation held at the CLARIN-NL Second Open Call Info Session, Amsterdam, August 26, 2010. [[pdf](#)]
- Kemps-Snijders, M., Wright, S.E., Windhouwer, M. (2010), "ISOcat Introduction", presentation held at the ISOcat Tutorial, Utrecht, March 25, 2010 [[ppt](#)]
- Kemps-Snijders, M., Windhouwer, M. (2010), "ISOcat Annotation Tutorial", presentation held at the ISOcat Tutorial, Utrecht, March 25, 2010 [[ppt](#)]
- Kemps-Snijders, M., Windhouwer, M. (2010), "ISOcat Datamodel Tutorial", presentation held at the ISOcat Tutorial, Utrecht, March 25, 2010 [[ppt](#)]
- Kemps-Snijders, M., Windhouwer, M. (2010), "ISOcat Hands-on Tutorial", presentation held at the ISOcat Tutorial, Utrecht, March 25, 2010 [[ppt](#)]
- Odijk, J. (2010), ""Relations between Data Categories, presentation held at the CLARIN Relation Registry Workshop, MPI, Nijmegen, January 8, 2010 [[ppt](#)]
- Odijk, J. (2010), "ISOCAT problems encountered at in DUELME-LMF", presentation held at the ISOcat Workshop, Nijmegen, September 21, 2010 [[ppt](#)]
- Odijk, J. (2010), "Proposed solutions for problems encountered in DUELME-LMF", presentation held at the ISOcat Workshop, Nijmegen, September 21, 2010 [[ppt](#)]
- Ravenek, W. (2010), "Circulation of knowledge and learned practices in the 17th-century Dutch Republic", presentation held at the CLARIN/DARIAH Conference, Vienna, October 19, 2010 [[ppt](#)]
- Schuurman, I. (2010), "CLARIN and ISOcat: Een aantal algemene en meer specifieke, PoS-gerelateerde problemen", presentation held at the ISOcat Workshop, Nijmegen, September 21, 2010 [[ppt](#)]
- Van Uytvanck, D. (2010), "Metadata and DCR", presentation held at the ISOcat Tutorial, Utrecht, March 25, 2010 [[ppt](#)]
- Vriend, F. de (2010), "Creating & Testing CLARIN Metadata Components", presentation held at LREC 2010, Valetta, Malta, May 18, 2010 [[ppt](#)]

PROGRESS REPORT 2009-2010 ESFRI PROJECTS

- Vriend, F. de (2010), "MIMORE", presentation held at the CLARIN-NL Second Open Call Info Session, Amsterdam, August 26, 2010 [[pdf](#)]
- Vriend, F. de, Broeder, D., Depoorter, G., Van Eerten, L. & Van Uytvanck, D. (2010). "Creating & Testing CLARIN Metadata Components", in *Language Resource and Language Technology Standards – state of the art, emerging needs, and future developments Workshop, 7th International Conference on Language Resources and Evaluation*, pp 41-43 [[abstract: pdf](#) & [full version: pdf](#)]
- Windhouwer, M. (2010), "ISOcat status", presentation held at the ISOcat Workshop, Nijmegen, September 21, 2010 [[pptx](#)]
- Windhouwer, M. (2010), "TDS Curator", presentation held at the ISOcat Workshop, Nijmegen, September 21, 2010 [[pptx](#)]
- Windhouwer, M. (2010), "Possible solutions ISOcat", presentation held at the ISOcat Workshop, Nijmegen, September 21, 2010 [[pptx](#)]

Publications

- Odijk, J. (2010) "The CLARIN-NL Project", in Calzolari et al. *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, pp. 48-53. <http://www.lrec-conf.org/proceedings/lrec2010/index.html> [[pdf](#)]

Seasonal School Lectures

- Bennis, H.J. (January 2010) LOT-winterschool, Free University, Amsterdam; PhD-course: *Linguistic Infrastructure and Microvariation*
- Odijk, J. (planned for January 2011), LOT-winterschool, UvA, Amsterdam; PhD Course: *Introduction to Language and Speech Technology*

Workshop Organization

- Bennis, H.J. (Convenor) (2010), Workshop *Language Variation and Digital Technology* at SDH-Conference, Vienna, October 2010

Speakers:

- Hans Bennis (Meertens Institute) "Syntactic Variation"
- Giuseppe Longobardi (University of Trieste) "Syntactic Measuring of Linguistic Distances"
- Frans Gregersen (University of Copenhagen) "Merton, Accountability and the Sociolinguistic Study of Language Variation"

PROGRESS REPORT 2009-2010 ESFRI PROJECTS

Appendix A CLARIN-NL Board Members

Name	Affiliation
Prof.dr. Geert Booij	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	ACLIC, Universiteit van Amsterdam
Prof.dr.ir John Nerbonne	CLCG, Universiteit van Groningen

PROGRESS REPORT 2009-2010 ESFRI PROJECTS

Appendix B CLARIN-NL NAP Members

Name	Affiliation
Prof.dr. Willem Adelaar	LUCL, Universiteit van Leiden
Dr. Sjef Barbiers	Meertens Instituut
Dr. Jeannine Beeken	INL
Prof.dr Antal van den Bosch	Universiteit van Tilburg
Dr. Gosse Bouma	Universiteit van Groningen
Drs. Hennie Brugman	Beeld & Geluid
Dr. Karina van Dalen-Oskam	Huygens Instituut
Dr. Paul Doorenbosch	Koninklijke Bibliotheek
Dr. Willemijn Heeren	Universiteit van Twente
Prof.dr. Kees Hengeveld	ACLIC, Universiteit van Amsterdam
Ir. Marc Kemps-Snijders	Max Planck Instituut voor Psycholinguïstiek, Nijmegen
Dr. Nelleke Oostdijk	CLST, Radboud Universiteit Nijmegen
Prof.dr. Reinier Salverda	Fryske Akademy
Prof.dr. Ted Sanders	UIL-OTS, Universiteit van Utrecht
Prof.dr. Piek Vossen	Vrije Universiteit Amsterdam
Prof.dr. Paul Wouters	VKS
Dr. Joris van Zundert	Huygens Instituut

PROGRESS REPORT 2009-2010 ESFRI PROJECTS

Appendix C CLARIN-NL IAP Members

Name	Affiliation	Country
Prof.dr Scott Farrar	University of Washington, Department of Linguistics	USA
Prof.dr David Hoover	New York University	USA
Prof.dr. Clifford Lynch	Coalition for Networked Information	USA
Prof.dr. Harold Short	King's College London, Centre for Computing in the Humanities	UK
Prof.dr. Benjamin T'sou	City University of Hong Kong	China
Prof.dr Hugo Van hamme	Katholieke Universiteit Leuven	Belgium
Prof.dr. Tamás Váradi	Linguistics Institute, Hungarian Academy of Sciences	Hungary