



CLARIN-NL Work Plan 2014

1 Introduction

This document provides a description of the activities planned in CLARIN-NL for 2014. Actual budget amounts for the various activities can be found in the separate document with the budget details, and a separate document with the 2014 budget is available.

Important decisions have been marked with the symbol →.

A summary of all decisions to be taken has been provided in section 5. All decisions have been approved by the CLARIN-NL Board in its meeting of April 15 2014.

2 Planned Activities

In 2014 a number of activities that were started up earlier are on-going. The major focus for new activities in 2014 is on two aspects:

- Improving interoperability and certain aspect of user-friendliness
- Dissemination and user involvement

2.1 Technical Implementation

The IIP project has finished. Huygens Institute is still working on its project to become a CLARIN Centre. We will set up a regular CLARIN-NL centre meeting (3 to 4 times a year) to discuss problems and challenges, to exchange information and experiences, and to provide input to the ERIC Centre committee and discuss input from that committee.

All centres have started the certification procedure, and all (except DANS) aim to get certified in 2014.

The *Search & Develop* project has finished. The approach for federated content search developed in this project is different in one small respect from the one developed by CLARIN-D. It has been decided, however, as recommended by the technical experts involved (Daan Broeder, Marc Kemps-Snijders) to conform to the convention adopted by CLARIN-D, so that the Dutch resources for which a so-called federated content search endpoint has been set up can be searched through the [federated search interface set up by CLARIN-D](#).

DBNL has finished its work on becoming a CLARIN Data Provider in January 2014.

The project of the *Koninklijke Bibliotheek* (KB, Royal Library) is still being executed, but has significant delays, in part due to independent (software) changes in the KB.



CLARIN-NL Work Plan 2014

The project for metadata for tools (MD4T) has yielded an initial profile and partial resource descriptions for 5 tools. → In 2014 we will extend the profile and apply it to all software developed in CLARIN-NL so that resource descriptions for all tools and services developed in CLARIN-NL will become available and visible via the [Virtual Language Observatory](#).

The Linked Open Data initiative is rapidly gaining popularity. For this reason, it is desirable if the CMDI Metadata that CLARIN provides can also be represented as Linked Open Data. → A project to achieve this will be carried out in 2014.

The CLAPOP project is working on the CLARIN-NL portal, and aims to have a first version running and with content in May 2014. This portal will provide an integrated entry point to the CLARIN infrastructure, with a focus on the contributions by CLARIN-NL. It is being developed in close cooperation with the European CLARIN portal, which is currently also being developed, to ensure complementarity. Less money was needed for the portal than originally planned, and → CLAPOP's budget has been reduced

→ The support for [CLAM](#) (software to facilitate turning existing tools into RESTful web services) and [FoLiA](#) (proposed standard format for linguistically annotated corpora) is continued in 2014, and actually extended in order to extend CLAM with authentication facilities for web services.

The CLAVAS project has yielded software for unified access to vocabularies, and an editor to maintain vocabularies. → We will use this software to manage some vocabularies used in the metadata and in annotations of resources.

Money had already been reserved in 2013 for a project on harmonizing all (separately and independently created) metadata records. By the end of 2013, a candidate to carry out this project was found, and the project will start by January 1, 2014. A detailed work description is available separately.

Certain tools and services have ad-hoc login procedures rather than federated login, as required by CLARIN. In part, this is due to the fact that these tools have been developed before federated login was implemented by the CLARIN centres. → We have reserved some money to change the ad-hoc logins into CLARIN-compatible federated login. (This concerns Gabmap, BILAND/WAHSP, perhaps PILNAR)

An analysis of the interoperability of the developed functionality has shown that significant improvements of interoperability are needed. The analysis of certain aspects of user-friendliness showed that most tools and services should allow a larger array of file types as input. This can be achieved by more systematically deploying developed converters and wrappers, and by developing some new converters and wrappers. → A project dedicated to these converters and wrappers will be carried out in 2014.



CLARIN-NL Work Plan 2014

The analysis of certain aspects of user-friendliness also showed that in particular the TTNWW software requires significant improvements in this respect. In addition, tools and services should always allow metadata as input and automatically generate metadata for its output. Almost none of the tools and services currently do this. We aim to experiment with doing this for the TTNWW services. → A project to improve the TTNWW functionality in this respect is planned for 2014.

- We aim to further extend the interoperability of specific services by integrating
- TTNWW PoS tagging and INL/OpenSONAR Corpus Search
 - TTNWW/Parsing and the LASSY grammatical relation search application
 - Searching in OpenSONAR combined with Cornetto for filtering search results on the basis of semantic properties

We are currently investigating which of these interoperability activities are technically, logistically, and financially feasible.

For more details on these aspects, we refer to the documents

- *CLARIN-NL Interoperability Usability Overview*, file: CLARIN-NL Interoperability Usability Overview 2014-02-26.pdf
- *Gebruikersvriendelijkheid en interoperabiliteit*, file Gebruikersvriendelijkheid en interoperabiliteit 2014-02-27.pdf
- *Integratieproject - enkele voorstellen*, file: Integratieproject - enkele voorstellen 2014-04-02.pdf

2.2 Data Conversion and Creation

Covered in section 2.3

2.3 Tools and Services

This item also includes *Data Conversion and Creation* (section 2.2) and *User Needs and Usage Cases* (section 2.4)

Calls

All projects from Calls 1 and 2 have been formally closed.

All projects from Call 3, as well as the CKCC project and the MinDisc project, have finished in 2013 but some are not formally closed yet. The goal is to formally close all these projects ultimately by May 2014.

All projects from Call 4 and LAISEANG have started up and most will end mid-2014.

No new calls will be set up in 2014 or 2015.

Other Projects



Cooperation between Netherlands and Flanders

This overall project has effectively finished. Its major subproject TTNWW has finished, though it still has to be formally closed.

ISOCAT Coordinator

→ We target to extend the position of the ISOCAT coordinator for the whole of 2014 (current contract runs until April 1, 2014), but there are problems with continuing the contract with Ineke Schuurman, for which we are currently searching a solution.

UBs as data providers

The experiences with UBU were very positive: they made their digital data available in a CLARIN-compatible manner in a cheap and short-term project. The KB project, however, is continuously delayed. We will not start up any new projects but first await the results of the KB (which may cover much of the university library data anyway)

2.4 User Needs and Usage Cases

Covered in section 2.3

2.5 Advanced LT Services

No plans for this or covered in section 2.3

2.6 Centres of Expertise

The virtual Centre of Expertise on the curation of data (DCS), led by the Radboud University of Nijmegen has ended its activities in 2013. After 2013, no more funds are available from CLARIN-NL to continue this activity, → though a small amount has been reserved to ensure that knowledge and expertise of the DCS is properly documented.

In 2012 it has been decided to support the *NWO-Groot* project *Nederlab* from CLARIN-NL with 150k euro. The EB considers this as a virtual centre in the area of data and software for research on language and culture through time. *Nederlab* has focused on a demonstrator for CLARIAH-SEED in 2013, but will have to make its contribution for CLARIN-NL in 2014. A plan for this is expected Mid April 2014.

2.7 Education and Awareness

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

A lot of data, tools and services have been made available via CLARIN-NL. The portal will become available in 2014, and the temporary stand-in page



CLARIN-NL Work Plan 2014

(<http://www.clarin.nl/node/404>) has been available since September 2013. It is important that potential users are made aware of this and start using all these data, tools and services. Specific activities will be set up in 2014 for this purpose. More specifically we aim to

- Hold lectures explaining CLARIN, the CLARIN-NL results, and specific data, tools or services CLARIN offers, to various CLARIN target user groups (linguists, historians, literary scholars, etc.)
- Record some of these lectures so that they are permanently available on the web. One already exists, see <http://www.clarin.nl/node/1978>. → Some budget has been reserved for covering any costs.
- Hold courses at seasonal schools, inter alia at the LOT 2014 summer school and the 2015 winter school
- Ensure presence and visibility of CLARIN at various conferences, workshops etc. of our target user group

Conferences and workshops to be visited

We aim to visit relevant conferences, workshops, and other events, and make submissions to them where appropriate.

- [THATCamp DH 2014](#), 14-15 jan 2014, Den Haag
- CLIN 2014, Leiden, 17 Jan 2014 (sponsor, stand)
- Various local researcher meetings (e.g. UiL-OTS Uilendag, UiL-OTS Syntax Interface meetings, ELITU, and similar events at other universities)
- Mining Digital Repositories Workshop, Den Haag, 10-11 April 2014
- LREC 2014, Reykjavik (May 28-30, 2014)
- NWO Bessensap, June 6, 2014 (Media Plaza, Utrecht)
- DH Benelux 2014, 12-13 June 2014, Den Haag
- CLARIN Annual meeting, Soesterberg, Oct 2014
- TIN 2015
- Others to be determined during 2014.

Events organized by CLARIN-NL

The EB aims to organize several events itself or with others, inter alia:

| From | To | Description | City | Ctry | Who |
|--------------|----|--------------------------|---------|------|-----|
| Sep 2014 | | CLARIN-NL Call 4 Meeting | Utrecht | NL | EB |
| 4 March 2014 | | Metadata Tutorial | Utrecht | NL | DB |
| June 2014 | | ISOCAT Workshop | Utrecht | NL | IS |



CLARIN-NL Work Plan 2014

Meetings

The EB will organize and/or attend meetings with relevant individuals and organizations. These will surely include

- Meetings with subprojects
- Meetings with CLARIN Centres, and with CLARIN Data Providers
- Meetings with related projects such as Taalportaal, Nederlab, CLARIAH seed money project, etc.
- Meetings with funders
- Meetings with the KNAW, esp., the eHumanities programme
- Meetings with the NL eScience Centre and other technical infrastructure providers (SARA, SURF)
- Monthly National Coordinator Forum (NCF) meetings (usually virtual)
- CLARIN ERIC GA Meeting, Tübingen Nov 2014

Support for other events

CLARIN-NL will continue to support events organized by others that are relevant for infrastructures, digital Humanities, eHumanities. Some relevant events are CLIN 2014 (Leiden, 17 Jan. 2014), a workshop dedicated to audio-visual data (organized by Rene van Horik), and workshops organized by CLARIN-NL subprojects.

Support for travel of Dutch researchers to CLARIN-related events

CLARIN-NL will continue its support of Dutch researchers to attend CLARIN-related events.

CLARIN- NL Newsflashes

CLARIN- NL aims to issue at least 6 Newsflashes in 2014.

CLARIN ERIC Newsflashes

It is likely that a CLARIN ERIC newsflash will start to appear in 2014, and CLARIN-NL will contribute to these (at least twice).

CLARIN ERIC Newsletter

It is likely that a CLARIN ERIC newsletter will start to appear in 2014, and CLARIN-NL will contribute to these (at least twice).

eData & Research

CLARIN-NL has discontinued its membership of the eData & Research Magazine as of 2014. CLARIAH-SEED has joined, so some visibility of CLARIN will very likely be achieved.

Other Publications

The programme director has been invited to be the guest editor of a *Lingua* special issue on linguistic research in the CLARIN infrastructure. The proposal for this special issue has been accepted by the *Lingua* editorial board. Submissions are due



CLARIN-NL Work Plan 2014

April 30th, 2014, after which a reviewing procedure starts. It is planned to have the papers ready for publication in 2014. In 2013, money was already reserved to make this special issue an Open Access journal issue and we will attempt to obtain additional funding for this from NWO and from sources in other countries from which researchers contribute.

The EB will work on getting publications in the *Historisch Nieuwsblad*, *Nederlandse Letterkunde*, NWO-related magazines (*Geestesoog*, *Hypothese*), etc. to reach the humanities researchers.

Movies and Screen Captures

Subprojects have been encouraged to make screen captures to illustrate the functionality of their applications and services.

It was already decided in 2013 to make 3 new movies for the subprojects CKCC, PILNAR and Polimedia. The [CKCC movie](#) already exists, the others are expected Mid 2014.

App

It will be investigated whether making an app for smart phones and tablets is useful from a PR perspective, and whether it is technically and financially feasible. → Some budget is reserved for this investigation and for the development of such an app.

Website

The website is running well, and no big changes are planned. The CLARIN-NL portal will use the CMS that is also used by the web site.

NWO and CLARIN-compatible resources

The EB has addressed the issue of NWO requiring language-related resources resulting from research projects to be CLARIN-compatible since 2012, and will work on an update for documents on this in 2014 (together with DANS).

Outreach to running research projects

The EB will approach running and starting research projects to create awareness of CLARIN-NL and CLARIN-compatibility of data and tools.

Education

Tutorials will be held on various infrastructural topics, e.g. on ISOCAT and CMDI metadata. The EB also aims to visit researchers in existing research projects (outside of CLARIN-NL) to point out the benefits of CLARIN-NL and stimulate and assist the researchers to work in a CLARIN-compatible way.

The pre-study to inventory opportunities for educational opportunities on the basis of the various CLARIN-subprojects has been finished, and educational modules for selected functionality in CLARIN is currently being implemented.



CLARIN-NL Work Plan 2014

Courses will be organized at seasonal schools such as the LOT 2014 summer school and the 2015 winter school. Opportunities for seasonal schools of other research schools are being investigated. → A small amount of money has been reserved for courses at these seasonal schools

Helpdesk

The HelpDesk will continue to run as it has done in 2013. It will be managed by JO and DB.

Final Event

The EB had reserved 30k euro (already in 2013) for a final event to finish the CLARIN-NL project with. This event will very likely be held at the end of 2014 or the beginning of 2015.

2.8 National Coordination Point

The national coordination point continues to operate as before.

It is expected that the amount of work for the National Coordination Point will be reduced in 2014 (no new calls anymore, most money allocated, etc.). For this reason, and because of budgetary constraints, the board approved (in 2013) reducing the effort of the programme director from 1.0FTE to 0.8 FTE as of 2014, and to reduce the effort for the project secretary from 0.5FTE to 0.3FTE as of 2014.

EB The EB has remained unchanged. For the reasons described above, the board approved reducing the effort of the EB for each member (except the programme director) from 0.2 FTE to 0.1 FTE as of 2014.

NAP There were no changes in the composition of the NAP in 2013. One NAP meeting is planned for 2014.

IAP The International Advisory Panel consists of 10 members. No IAP meetings are planned in 2014, though the IAP members will be invited for the final CLARIN-NL event. → The budget for travel and subsistence for governance will be increased to enable the IAP members to attend the Final Event

2.9 EU-Line

The budget for this item will be spent on membership fees the Netherlands is obliged to the CLARIN ERIC as its host, to cover membership for a period of five years, as explicitly agreed upon between the CLARIN-NL board, OCW, and NWO. In 2014 the third part of the fees will have to be paid.

3 2014 Budget



CLARIN-NL Work Plan 2014

The CLARIN-NL budget for 2014 can be found in the document *CLARIN-NL 2014 Budget*.

4 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.

As a result of this, CLARIAH has been put on the roadmap, it has received 1 million euro *seed money*, and it has submitted a revised proposal. A decision on funding this revised proposal is expected in June 2014.

5 Summary of Decisions

- → Metadata for Tools: In 2014 we will extend the profile and apply it to all software developed in CLARIN-NL.
- → A project to represent CMDI metadata as Linked Open Data (RDF triples) will be carried out in 2014
- → The CLAPOP project budget has been reduced.
- The support for [CLAM](#) (software to facilitate turning existing tools into RESTful web services) and [FoLiA](#) (proposed standard format for linguistically annotated corpora) is continued in 2014, and actually extended in order to extend CLAM with authentication facilities for web services.
- → We will use the CLAVAS software to manage some vocabularies used in the metadata and in annotations of resources.
- → We have reserved some money to change the ad-hoc logins into CLARIN-compatible federated login. (This concerns Gabmap, BILAND/WAHSP, perhaps PILNAR).
- → A project dedicated to converters and wrappers for improving interoperability will be carried out in 2014.
- → A project to improve the TTNWW functionality (support for more input formats, for multiple files; for generating metadata) is planned for 2014.
- → The interoperability of specific services is extended by integrating
 - TTNWW PoS tagging and INL/OpenSONAR Corpus Search
 - TTNWW/Parsing and the LASSY grammatical relation search application
 - Searching in OpenSONAR combined with Cornetto for filtering search results on the basis of semantic properties



CLARIN-NL Work Plan 2014

- → We target to extend the position of the ISOCAT coordinator for the whole of 2014.
- → A small amount has been reserved to ensure that knowledge and expertise of the DCS is properly documented.
- → Some budget has been reserved for covering any costs related to recording lectures.
- → Some budget is reserved for this investigation and for the development of an app to provide easy and attractive access to CLARIN functionality.
- → A small amount of money has been reserved for courses at these seasonal schools
- → The budget for travel and subsistence for governance will be increased to enable the IAP members to attend the Final Event