Common Language Resources and Technology Infrastructure

CLARIN NL

Fact Book

May 2016

WWW.CLARIN.NL

Contents

CLARIN-NL Objectives 6

Success Criteria 7

Global Budget 9

Governance 11

Board 11

Executive Board 11

National Advisory Panel 11

International Advisory Panel 12

Project office 12

CLARIN Centres 13

CLARIN Data Providers 14

11-027 EdiTE 14

11-030 Sound and Vision as CLARIN Data provider 14

11-036 KB as Data Provider 15

13-008 Utrecht University Library 15

Calls 16

Texts (integrally taken from original publications) 16

Call 1 – Open Call (2009) 16

Call 2 – Open Call (2010) 26

Call 3 (2011) 37

Call 4 (2012) 59

Call3 (Open Call) 93

Call 3 (Closed Call) 107

Call 4 (Open Call) 120

Call 4 (Closed Call) 140

Projects 160

Call 1 160

Call 2 166

Call 3 171

Call 4 179

2013 Projects 184

13-012 Lassy 184

2014 Projects 185
14-003 CMDI2RDF 185
14-010 PaQu 185
14-011 OpenConvert 186
14-012 AutoSearch 186
14-013 CORNETTO-LMF-II 186
2015 Projects 187
15-001 TPC 187
15-002 MDF 187
15-003 CorpusSearchWeb 188
15-004 CARE 188
15-005 OpenCGN 189
Infrastructure Implementation Projects 190
10-002 TTNWW 190
09-001 MD 190
10-003 IIP 191
12-003 CLARIN centre infrastructure Huygens ING 191
10-004 S&D 192
MD4T 192
CLAM-S (11-032, 13-001, 13-013) 192
ISOCAT Coordinator (11-033) 193
11-002 CLAVAS 193
13-011 MCC 194
14-014 OpenSKOS 194
Centres of Expertise 195
Data Curation Centre (10-025, 11-005) 195
Centre for the study of the Dutch language and culture across time (13-003) 195
User Survey 197
13-005 EduClarin (and 13-014, 14-005 – 14-009) 197
Events 199
Organized 199
2009 199
2010 199

2011 199
2012 199
2013 199
2014 199
2015 200
Supported 200
2009 200
2010 200
2011 201
2012 201
2013 201
2014 201
2015 202
Attended 203
2009 203
2010 203
2011 203
2012 204
2013 205
2014 206
2015 207
Meetings 207
2009 207
2010 208
2011 208
2012 209
2013-2015 210
Media 211
Movies 211
Newsflash 212
Press 213
2010 213
2011 214

2012 214
2013 215
2014 216
2015 217
Publications 217
2010 217
2011 218
2012 219
2013 220
2014 221
2015 227
2016 228
Presentations 229
2009 229
2010 231
2011 235
2012 239
2013 242
2014 245
2015 248
2016 251
Awards, Prizes 251
2013 251
PoliMedia wins the Veni LinkedUp Challenge 251
Polimedia is finalist of the Semantic Web Challenge. 251
2014 252
SHEBANQ wins the Digital Humanities Award 2014 on the category *Best Tool or Suite of Tools* 252
2015 252
PoliMedia wins LODLAM Open Data Prize 252
CLARIN Young Scientist Award for Maarten van Gompel 252

CLARIN-NL Objectives

CLARIN-NL aims to design, construct, validate, and exploit a research infrastructure that is needed to provide a sustainable and persistent eScience working environment for researchers in the Humanities, and Linguistics in particular, who want to make use of language resources and the technology to use these resources for their research. This infrastructure will provide these researchers with a wide variety of resources and services, intelligent access methods for exploring the resources and innovative ways of combining different resources into virtual collections, so that information hidden in unstructured textual and multimedia documents can be disclosed. Inter-operability of independently developed resources and services will be key for a properly functioning infrastructure. The infrastructure will be easy to use for non-technical researchers. Targeted dissemination activities, educational programmes and training sessions will enable a whole generation of researchers and students to acquaint themselves with this new research methodology and the potential for ground breaking research it offers, creating an advanced scientific environment in the Netherlands that will attract top-researchers and students from abroad.

CLARIN-NL forms the Netherlands national counterpart of the CLARIN enterprise on the European level (CLARIN-EU). It therefore resembles and complements the preparatory project that is currently being executed on the European level (CLARIN-prep). Many of the activities and sub-projects within CLARIN-NL implement activities in the Netherlands that in the programme of work for CLARIN-prep are envisaged to take place in every participating country and that will be funded through the national contributions to CLARIN. Such activities include (1) the design and implementation of the infrastructure technology; (2) application projects in which technology providers and the intended users integrate local repositories and set up local services for prototypical test installations as initial demonstrators, enabling evidence-based contributions to the discussion on standards and best practices for inter-operability, and to contribute to the survey of requirements for the infrastructure technology; (3) the preparation of an essential data collection and service set for the locally relevant languages (ideally on the basis of existing tools and data) that allows for testing and validation of proposed standards, services and tools in the experimental prototype; and (4) the integration of advanced infrastructure services.

Since it is not possible to assign all these tasks to participants right from the start, CLARIN-NL has been set up as a mixture between a programme and a project.

CLARIN-NL, however, also contains a range of activities that aim to further strengthen the leading position the Netherlands currently has in CLARIN-EU (both the principal coordinator and the technical coordinator for infrastructure technology are based in the Netherlands). It has a separate line of activities aimed to position the Netherlands prominently in CLARIN-EU also beyond CLARIN-prep, and to extend its leading position further by initiating, in an early stage, projects with selected international partners to develop, in a multilingual setting, showcase demonstrators of the infrastructure and the services it offers, as well as by setting up at least two centres of expertise

The CLARIN-NL proposal covers a period of 6 years, partitioned in three phases of two years: the preparation phase, the construction phase, and the first two years of the exploitation phase. Though the infrastructure is primarily aimed at language and humanities researchers, it offers various

opportunities for usage in other domains and by other users, both for commercial applications as well as for important developments in society.

Success Criteria

The general ambition of the CLARIN-NL project is to be a major contributor both organisationally and technically to the design, specification, construction and exploitation of a European-wide CLARIN infrastructure that is actually used by its intended users and where its use has become a normal mode of operation for them as reflected by the incorporation of training and education in the curricula of linguistics and humanities studies.

This general ambition can be translated into more specific success criteria, subdivided into an EU-Line and an NL-Line (as was also done for the actions):

EU-Line

- Has the governance structure as recommended by the CLARIN preparatory project been implemented, and is it functioning successfully?
- Has the main CLARIN office for administrative and logistic support for the governance bodies been set up in the Netherlands?
- Has the main European CLARIN Technical Centre to build and maintain the technical infrastructure been set up?
- Has the central CLARIN coordination point been set up for development and maintenance of standards, harmonization of IPR issues, and education, dissemination and promotion?
- Has the Netherlands set up an international example infrastructure with one or two other leading partners?
    - Is this example infrastructure nationally and internationally recognized as exemplary?
    - Is this example infrastructure successfully used by the intended users (linguistic and humanities researchers)?
    - Is this example infrastructure sufficiently known among the intended users?

NL-Line

- Technical Infrastructure
    - Has the intended technical infrastructure actually been constructed?
    - Is the distributed nature of technical infrastructure indeed invisible to users?
    - Is the performance of the technical infrastructure sufficiently good?
- Data Infrastructure
    - Are the data identified as “essential” included in the technical infrastructure?
    - Can they be easily found and accessed by users, and are they properly documented?
    - Have NL-specific IPR issues been adequately dealt with? Have procedures been defined to deal with IPR issues for new data?
    - Are there guidelines and procedures in place as well as supporting tools to easily incorporate new data into the infrastructure?
- Language technology infrastructure

  - Have the tools and services identified as “essential” been included into the technical infrastructure?
  - Can they be easily found and accessed by users, and is there proper documentation to assess their appropriateness for a given task?
  - Has actual interoperability among tools/services and between tools and data been achieved in the technical infrastructure?
- User Needs
  - Are the data and tools identified as “essential” indeed the data and tools the intended users need most?
  - Is the technical infrastructure indeed used by the intended users?
    - Are the intended users happy in using the technical infrastructure?
    - Do the interfaces promote and stimulate working with it or do they pose obstacles?
    - Are the interfaces user –friendly and self-explanatory where possible?
    - How many researchers actually use the infrastructure on a regular basis or occasionally? (Target: 40% of the intended users regularly use it; 70% use it at least occasionally)
- Centres of Expertise
  - Have one or two centres of expertise been created and are they operating successfully?
  - Are they recognized as centres of expertise, both nationally and internationally?
- Dissemination, education, awareness
  - Is the existence of the technical infrastructure known to the intended users (target: 80% know of its existence)
    - Have all relevant players been sufficiently informed to be able to participate in designing and constructing the infrastructure?
    - Has enough training and education on using the infrastructure been given? Are still ample opportunities offered to get such trainings?
    - Has training in the use and actual use of the infrastructure been incorporated in the regular curricula of linguistics and humanities studies, or are there concrete plans to do so? (Target: incorporated in 60% of the curricula after 6 years)
  - National Coordination Point: Has a national coordination point been set up and is it functioning successfully?
  - Has a business model been developed that guarantees the long term sustainability of the CLARIN infrastructure?

Global Budget

As stated at the start of the project in 2009:

CLARIN-NL 2009-2014 all amounts in M€	preparation 2009_2010			construction 2011_2012			exploitation 2013_2014			all periods 2009_2014		
EU-level	*labour*	*other*	*total*	*labour*	*other*	*total*	*labour*	*other*	*total*	*labour*	*other*	*total*
Management and coordination	0.02	0.03	0.05	0.11	0.05	0.16	0.24	0.06	0.30	0.37	0.14	0.51
Technical coordination	0.01	0.02	0.03	0.05	0.03	0.08	0.12	0.04	0.16	0.17	0.09	0.26
Linguistic coordination	0.01	0.01	0.01	0.05	0.03	0.08	0.12	0.04	0.16	0.17	0.08	0.25
Outreach coordination	0.01	0.01	0.02	0.05	0.03	0.08	0.12	0.04	0.16	0.17	0.08	0.26
Internationalisation	0.00	0.01	0.01	0.01	0.01	0.02	0.02	0.02	0.04	0.04	0.04	0.08
Total EU level in M€	0.04	0.08	0.12	0.26	0.15	0.41	0.62	0.20	0.82	0.93	0.43	1.35
NL-level	*labour*	*other*	*total*	*labour*	*other*	*total*	*labour*	*other*	*total*	*labour*	*other*	*total*
Technical construction	0.30	0.05	0.35	0.48	0.07	0.55	0.19	0.02	0.21	0.97	0.14	1.11
Data conversion	0.48	0.06	0.54	0.54	0.10	0.64	0.29	0.05	0.34	1.31	0.21	1.52
Tools and services	0.54	0.07	0.61	0.60	0.08	0.68	0.17	0.04	0.21	1.31	0.19	1.50
User needs and usage cases	0.43	0.06	0.50	0.59	0.09	0.68	0.18	0.02	0.20	1.20	0.17	1.37
Advanced LT services	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Expertise centres	0.06	0.01	0.07	0.20	0.06	0.27	0.19	0.07	0.26	0.46	0.15	0.61
Dissemination and training	0.12	0.08	0.20	0.19	0.09	0.28	0.12	0.09	0.21	0.43	0.26	0.69
Coordination & management	0.23	0.08	0.30	0.22	0.07	0.29	0.20	0.06	0.26	0.65	0.20	0.85
Total NL level in M€	2.16	0.41	2.57	2.82	0.57	3.39	1.34	0.34	1.69	6.32	1.33	7.65
Grand total in M€	2.20	0.49	2.69	3.08	0.72	3.80	1.97	0.54	2.51	7.25	1.75	9.00

Current situation:

Financial Overview 2009-2015	final	actual	actual	actual	actual		budget	budget	budget		
		2009	2010	2011	2012		2013	2014	2015		Total
Technical construction	**1.1**		€ 208,845.31	€ 222,000.00	€ 199,668.80		€ 172,624.00	€ 244,500.00	€ 0.00		**€ 1,047,638.11**
Data conversion & creation	**1.2**		€ 171,984.00	€ 35,339.32	€ 14,083.00		€ 0.00	€ 0.00	€ 0.00		**€ 221,406.32**
Tools and services	**1.3**		€ 717,981.00	€ 604,595.10	€ 1,365,422.57		€ 1,511,004.59	€ 327,418.50	€ 0.00		**€ 4,526,421.76**
User needs and usage cases	**1.4**	€ 24,790.89	€ 38,643.57	€ 119.40	€ 0.00		€ 0.00	€ 0.00	€ 0.00		**€ 63,553.86**
Advanced LT services	**1.5**				€ 0.00		€ 0.00	€ 0.00	€ 0.00		**€ 0.00**
Expertise centres	**1.6**		€ 6,750.00	€ 2,250.00	€ 60,000.00		€ 187,690.00	€ 37,500.00	€ 0.00		**€ 294,190.00**
Dissemination and training	**1.7**	€ 18,410.15	€ 36,031.25	€ 75,583.12	€ 96,002.54		€ 148,385.88	€ 99,346.81	€ 5,000.00		**€ 478,759.75**
Coordination & management	**1.8**	€ 144,811.07	€ 190,380.43	€ 174,998.57	€ 249,370.33		€ 188,958.00	€ 134,715.20	€ 0.00		**€ 1,083,233.60**
EU Level	**1.9**			€ 4,223.08	€ 322,127.29		€ 255,000.00	€ 260,100.00	€ 535,910.00		**€ 1,377,360.37**
Totaal		**€ 188,012.11**	**€ 1,370,615.56**	**€ 1,119,108.59**	**€ 2,306,674.53**		**€ 2,463,662.47**	**€ 1,103,580.51**	**€ 540,910.00**		**€ 9,092,563.77**
Cumulatief		**€ 188,012.11**	**€ 1,558,627.67**	**€ 2,677,736.26**	**€ 4,984,410.79**		**€ 7,448,073.26**	**€ 8,551,653.77**	**€ 9,092,563.77**		**TRUE**
Inkomsten Other		€ 0.00	€ 0.00	€ 0.00	€ 10,000.00		€ 72,563.77	€ 0.00	€ 0.00		**€ 82,563.77**
Inkomsten NWO		€ 0.00	€ 2,700,000.00	€ 1,900,000.00	€ 1,900,000.00		€ 1,250,000.00	€ 800,000.00	€ 460,000.00		**€ 9,010,000.00**
Inkomsten Totaal		**€ 0.00**	**€ 2,700,000.00**	**€ 1,900,000.00**	**€ 1,910,000.00**		**€ 1,322,563.77**	**€ 800,000.00**	**€ 460,000.00**		**€ 9,092,563.77**
Ink Cumulatief		**€ 0.00**	**€ 2,700,000.00**	**€ 4,600,000.00**	**€ 6,510,000.00**		**€ 7,832,563.77**	**€ 8,632,563.77**	**€ 9,092,563.77**		**TRUE**
Income Cum - Expenses Cum		-€ 188,012.11	€ 1,141,372.33	€ 1,922,263.74	€ 1,525,589.21		€ 384,490.51	€ 80,910.00	€ 0.00		

Governance

The CLARIN-NL organizational structure consists of the following bodies:

Board

Prof. dr. Geert Booij – Leiden University (until 2012)
Prof. dr. Lou Boves – Radboud University Nijmegen
Dr. Peter Doorn – Data Archiving and Networking Services
Prof. dr. Martin Everaert – Utrecht University
Prof. dr. Jaap van den Herik – Tilburg University
Prof. dr. Aafke Hulk – University of Amsterdam (until 2012)
Prof. dr. John Nerbonne – University of Groningen
Prof. dr. Els Stronks – Utrecht University (since 2012)
Peter Wittenburg – Max Planck Institute for Psycholinguistics (since 2012)

Executive Board

Prof. dr. Hans Bennis – Meertens Institute
Ir. Daan Broeder (since 2010)
Dr. Arjan van Hessen – Utrecht University
Prof. dr. Jan Odijk – Utrecht University
Peter Wittenburg - Max Planck Institute for Psycholinguistics (until 2010)

National Advisory Panel

Prof. dr. Willem Adelaar – Leiden University (until 2012)
Prof. dr. Sjef Barbiers – Meertens Institute
Dr. Jeannine Beeken – Institute for Dutch Lexicology
Prof. dr. Antal van den Bosch – Radboud University Nijmegen
Dr. Gosse Bouma – University of Groningen
Drs. Hennie Brugman – Meertens Institute
Prof. dr. Karina van Dalen-Oskam – Huygens ING
Drs. Paul Doorenbosch – National Library of the Netherlands
Dr. Willemijn Heeren – Leiden University
Ir. Marc Kemps-Snijders – Meertens Institute
Prof. dr. Kees Mandemakers – Erasmus University Rotterdam, International Institute of Social History
Prof. dr. Nelleke Oostdijk – Radboud University Nijmegen
Prof. dr. Toine Pieters – Utrecht University (since 2012)
Prof. dr. Reinier Salverda – Fryske Akademy
Prof. dr. Ted Sanders – Utrecht University (until 2012)
Prof. dr. Arjen Versloot – Fryske Akademy, University of Amsterdam (since 2012)
Prof. dr. Piek Vossen – VU University Amsterdam
Prof. dr. Sally Wyatt – KNAW eHumanities (until 2012)
Drs. Joris van Zundert – Huygens ING

International Advisory Panel

Prof. dr. Jan Hajič – Charles University (Prague, Czech Republic) (since 2012)
Prof. dr. Koenraad De Smedt – University of Bergen (Bergen, Norway) (since 2012)
Prof. dr. David Hoover – New York University (New York, USA)
Dr. Tamás Váradi – Hungarian Academy of Sciences (Budapest, Hungary)
Prof. dr. Harold Short – King's College London (London, United Kingdom)
Prof. dr. Erhard Hinrichs – Eberhard Karls University Tübingen (Tübingen, Germany) (since 2012)
Prof. dr. Clifford Lynch – Coalition for Networked Information (Washington DC, USA)
Prof. dr. Hugo Van hamme – KU Leuven (Leuven, Belgium)
Prof. dr. Benjamin T'sou – The Hong Kong Institute of Education (Hong Kong, China)
Dr. Scott Farrar – University of Washington (Washington DC, USA)

Project office

Since the start in 2009 the CLARIN-NL project office has been situated in the city centre of Utrecht at the Utrecht Institute of Linguistics OTS at Utrecht University. In January 2011 the project office moved from Janskerkhof 13 to Trans 10, due to a reorganisation of the Faculty of Humanities.

The project secretary has been staffed by:
Erica Renckens, MA (July 2009 – December 2010)
Jolien Scholten, MA (January 2011 – March 2013)
Arwin van der Zwan, MA (March 2013 – now)

CLARIN Centres

CLARIN Centres offer various types of services to the research community. They will host their own data and tools as well as those from "other parties". These Centres take care for the sustainability of data and tools and the permanent access to them.

Centre	Focus on resources	Resource examples
Meertens Institute	Relevant for the study of function, meaning and coherence of cultural expressions and resources relevant for the structural, dialectological and sociolinguistic study of language variation within the Dutch language.	Typological databases for Dutch dialects, descriptions of Dutch songs, databases of Dutch names, etc.
Max Planck Institute for Psycholinguistics	Related to the study of psychological, social and biological foundations of language.	Documentation of endangered languages, resources for sign languages, phonetic resources for the study of phone perception, speech error databases, tools for creating and annotating resources, etc.
Huygens ING	Related to the study of history and literature of the Netherlands.	Historical and literary manuscripts and their annotations, tools to annotate such manuscripts and create scholarly editions, etc.
Institute for Dutch Lexicology	That are relevant to the lexicological study of the Dutch language and on resources relevant for research in and development of language and speech technology.	Lexicons, lexical databases, text corpora, speech corpora, language and speech technology tools, etc.
Data Archiving and Networking Services	sustained access to digital research data. The DANS data archive contains thousands of datasets in the fields of humanities including oral history, archaeology, geospatial sciences and behavioural and social sciences. The data archive has acquired the Data Seal of Approval. Regarding the conditions under which access to the data is granted, the DANS motto is "Open Access when possible, Restricted Access when necessary".	Text resources, databases, spreadsheets, audio and audio-visual resources.

CLARIN Data Providers

In 2011, CLARIN-NL introduced a new centre type, namely CLARIN Data Provider. CLARIN-NL decided to define this new type of centre, because there is a range of centres in the Netherlands (and elsewhere) that by their very nature and mission archive large amounts of dedicated data. Moreover, these organizations give the research community access to (most of) their data. CLARIN-NL wants these data to be visible and accessible in a transparent and standardized manner in the CLARIN infrastructure. However, due to their nature, these organizations will not host data or tools from "third parties", which prevent them from becoming a full CLARIN Centre.

A couple of organizations have been invited to become CLARIN Data Provider. Some of them already started working towards becoming a CLARIN Data Provider, others are preparing their plans. They include the *Instituut voor Beeld en Geluid*, the *Koninklijke Bibliotheek*, the *Digitale Bibliotheek der Nederlandse Letteren (DBNL)*, the *Nationaal Archief*, and university libraries starting with the *Universiteitsbibliotheek Utrecht*.

11-027 EdiTE

Edited Texts in Editions
2012-07-01 – 2012-12-31
Project coordinator: Dr. R. van Stipriaan (Dutch Library for Dutch Literature (DBNL))

Participant	Budget
Dutch Library for Dutch Literature (DBNL)	€30,000
Total	**€30,000**

Abstract
Since the early days of DBNL, around 2000, modern editions of historical texts are very central to its digitization programs. Editions contain, by its nature, texts from at least two different periods; the edited text stems from a more or less precisely discernible moment in history; and it contains also text from the editor him/herself (the so-called paratext). Within historical search, the hybrid character of textual editions case serious problems. In the project Edited Texts in Editions DBNL will take two steps to solve these problems, the first step will be the isolation of the edited historical text within the whole of the edition, the second step will be the definition of a strong relation of the edited text with the metadata of the original text. By now, the DBNL-database consists of more than 1000 editions, most of them of scholarly origin. After the proposed data curation the content of all these text will be fit for any historical inquiry.

11-030 Sound and Vision as CLARIN Data provider

Steps for Sound and Vision to become a CLARIN Data Provider
2011-12-01 – 2012-02-29
Project coordinator: Dr. Johan Oomen

Participant	Budget
Netherlands Institute for Sound and Vision	€30,000
Total	**€30,000**

Abstract
This project enables the Netherlands Institute for Sound and Vision to make the necessary steps to comply with the CLARIN requirements to become a CLARIN centre Data Provider.

11-036 KB as Data Provider

Steps for the National Library of the Netherlands to become a CLARIN Data Provider
2012-03-01 – 2012-12-31
Project coordinator: Drs. H. Matthezing

Participant	Budget
National Library of the Netherlands	€30,000
Total	**€30,000**

Abstract
For the KB the aim of this project is to meet the requirements to be a CLARIN Data Provider. As the KB cannot meet the complete and extended set of the CLARIN Data Provider requirements at this moment, the project will aim to meet only a minimum necessary set of requirements. The KB does expect to meet the full range of requirements in the near future and is already working on them as part of its own policy.
Ideally, the KB would like to make all its metadata and digital content accessible through the CLARIN infrastructure, copyright will however limit the actual number of accessible collections. For the selected collections, metadata will be available for harvest in the CMDI format. A quality assessment will be done of the repository, pertaining to the EDBO-collection.

13-008 Utrecht University Library

Steps for the Utrecht University Library to become a CLARIN Data Provider
2013-02-01 – 2013-11-15
Project coordinator: Johan Tilstra

Participant	Budget
Utrecht University Library	€12,350
Total	**€12,350**

Abstract
University Library Utrecht will make available the metadata of digital data that it has available in CMDI-format and make the CMDI metadata available for harvesting by CLARIN metadata aggregators such as the CLARIN VLO.

Calls

Texts (integrally taken from original publications)

Call 1 – Open Call (2009)

CLARIN-NL launches an open call for project proposals. This call is open from Tuesday June 2, 2009. The total budget for this call is limited to a maximum of € 600,000. This call is specifically open for proposals targeting resource curation projects or demonstrator projects. Only those proposals that specifically target this priority will be eligible. The maximum budget per project is 60,000€.

1 Introduction

The first phase of the CLARIN-NL project focuses on the specification and design of the infrastructure. Obviously, the CLARIN infrastructure should be designed in such a way that it can incorporate the data and tools currently used by humanities researchers to address their research questions. On the other hand, the CLARIN infrastructure can only be successful if these data and tools meet certain requirements with regard to standards and interoperability, not only with regard to the data and tools themselves but also with regard to their visibility and accessibility. Many of the data and tools currently in use do not meet these requirements, or meet them only partially. One of the aims of the first call is to make existing data and tools compliant with the requirements on standards and interoperability currently advocated in CLARIN.

However, the list of standards, best practices and interoperability requirements currently advocated by CLARIN has not been completely fixed yet. There are good reasons for this: (1) there may be crucial data or tools for which none of the currently advocated standards or best practices is suited; (2) the currently advocated standards and best practices may be incomplete, insufficiently specific, inconvenient or even incompatible with crucial data and tools. Therefore, a second aim of the first call is to test whether and to which extent the currently existing data and tools can be made compatible with the standards and best practices currently advocated in CLARIN, and to get a good overview of any incompatibilities as well as suggestions for adaptations of these standards.

These matters can best be investigated by actually attempting the necessary conversions and adaptations. When applied to tools and other software this will result in CLARIN-compliant web applications that can also serve as showcases of the kind of functionality CLARIN aims to offer. A third aim of the first call is therefore to make sure that the applications developed become available in demonstrators which can help promote the particular applications but also the CLARIN infrastructure as a whole.

Finally, by investigating these matters using specific data and tools that are currently in use a lot can be learnt about requirements the CLARIN infrastructure should meet or desiderata that it should offer. The fourth aim of this call is therefore obtaining a detailed list of such requirements and desiderata.

2 Project Types

In the first call, CLARIN-NL therefore solicits (1) projects that carry out resource (data or tools) curation, and (2) demonstrator projects.

Curation projects

CLARIN-NL aims to support the curation of digital language resources (data and tools) so that these resources can participate in the CLARIN infrastructure, more easily be accessed by interested researchers via online methods and become part of appealing new applications. Many language resources are neither visible nor accessible. Visibility is mainly achieved by standardized metadata that are being harvested by service providers. Accessibility has many different aspects:

1. The resource needs to be stored at computers that are accessible.
2. The resource needs to be identified in a persistent manner.
3. The resource needs to be interpretable, which requires a format that adheres to best practices and it requires references to registries were the used concepts are defined.

Demonstrator projects

CLARIN-NL aims to support projects that create appealing showcases of functionality that the CLARIN infrastructure should support. Such projects should make available web applications that can be used as demonstrators of functionality that supports addressing research questions of the CLARIN-NL intended user group. The development of these demonstrators will also be used to inventory a list of requirements the CLARIN infrastructure should meet and desiderata it preferably should offer.

3 Goals

The goal of a curation project is:

- Adapting specific resources so that they are visible, uniquely referable and accessible via the web, and properly documented.

The goal of a demonstrator project is:

- Creating a documented web application starting from an existing tool or application that can be used as a demonstrator and function as a showcase of the type of functionality CLARIN will incorporate and support. Within the web application there must be a clear separation between the web-based user interface and the core component. For the latter an API must be defined and documented.

Important goals common to both types of projects are:

- Applying standards and best practices and makes use of the suggested CLARIN architecture and agreements to understand their limitations and the requirements for extensions.
- Establishing requirements and desiderata for the CLARIN infrastructure

In both project types the use of CLARIN-supported standards and best practices is essential.

The selection of CLARIN-supported standards and best practices is currently ongoing and it is the intention that the projects supported in this call contribute to this. A preliminary set of candidates for CLARIN-supported standards and best practices is available, and though it is by no means final we will refer to this list in this document by the term “CLARIN-standards”.

4 Roles

Four roles of persons involved in the projects can be distinguished: the *user*, the *data provider* (DP), the *technology provider* (TP), and the *infrastructure specialist* (IS). The *user* is a researcher from a linguistics or humanities institute who aims to investigate one or more specific research questions. The *DP* has a certain set of digital language-related data at his/her disposal (the 'research data') that can be used for addressing the research questions of the user. The *TP* has a certain technology (e.g. language or speech technology) at its disposal and a thorough understanding of this technology (e.g. because it was developed by the TP). This technology offers functionality that makes it possible to (better) address the user's research question by applying this functionality to the research data. The *infrastructure specialist* (IS) is a specialist who has a deep understanding of the CLARIN service-oriented architecture and its requirements, and/or a specialist in data and tool format standards supported in CLARIN. In many cases the different roles of user, DP, TP and IS will be played by persons from different organizations, but they may originate from a single organization, and occasionally even be played by a single person. CLARIN-NL can offer assistance in bringing the right experts together, if desired. CLARIN-NL will assign IS specialist to awarded projects.

4.1 User

The user is a researcher from a linguistics or humanities institute who aims to investigate one or more specific research questions. The project proposal should clearly describe the research question(s) of the user, and the research question(s) must be in the domain of research in the humanities in general and the study of language in particular. The project must be led by the user.

4.2 Data Provider

The DP has a certain set of digital language-related data at his/her disposal (the 'research data') that can be used for addressing the research questions of the user. The project proposal should clearly describe the research data the DP has at his/her disposal that can be used to address the research question(s), and how they can be used for this purpose. The research data must be existing digital language or language-related data. No new research data should be created in the project. The DP must have the right to make the research data available on a CLARIN server running at a dedicated centre. If the data are in a format that is not currently on the list of CLARIN standards, a resource curation project is in order. Otherwise, the data can be used in a demonstrator project. The project proposal should contain a detailed description of the research data, its current state and format, the plans to convert it if needed, justification for using different formats if applicable, and a detailed plan for dealing with the data and its metadata (see below). Any restrictions on the use of the data as well as any ethical issues that apply or may arise must be properly documented in the proposal.

4.3 The Technology Provider

The TP has a certain technology at his/her disposal that can be used as a basis for the development of a web-based application (possibly web-services based) and concomitant demonstrator, or that can be used for resource curation. Since the research data are language data, the technology will in most case be language or speech technology.

The proposal should contain a detailed description of the available technology and its current status. It should make clear that the TP has a thorough understanding of this technology and describe how the TP obtained this understanding (e.g. because the TP developed the technology).

The intended use of the technology in the project should be described, as well as any extensions or modifications that have to be made to the technology in the project, and a plan to achieve this.

The TP must have the right to use this technology and indicate how it will be used in the project.

4.4 Infrastructure Specialist

The infrastructure specialist (IS) is a specialist who has a deep understanding of the CLARIN service-oriented architecture and its requirements, and/or a specialist in data, metadata and tool format standards and best practices supported in CLARIN. The IS will assist the user and the TP to turn the research data and/or the technology into CLARIN-supported formats and architectures. The IS will advise and assist the project partners and may also contribute by providing reusable CLARIN components to the project partners. The IS will assist selected projects to turn its web application into a web service, and will assist the project with testing the project results.

A research data resource often consists of information of various kinds contained in multiple folders and multiple files of varying types. The information contained in such a resource can include documentation, source data, annotations of the source data, aggregate statistics tables on the source data and/or annotations, etc. A web service should find out in a fully automated manner whether the research data selected by the user are appropriate input for the web-service, and, if so, that it is applied to the right information (e.g. to the source data but not to the documentation). To achieve this it will integrate wrappers that read and write metadata and provenance information provided by the CLARIN infrastructure. Any requirements or desiderata that follow from this for metadata and data contents and formats should be properly documented in the documents with the requirements and desiderata for the CLARIN infrastructure.

5 Project Types

5.1 Resource Curation

Resource curation involves a number of different aspects:

1. The resource should be brought into a format that adheres to widely accepted standards and best practices currently considered as likely candidates by CLARIN.
2. Proper metadata descriptions need to be created and made available. They must be compliant with the CLARIN component metadata infrastructure (CMDI) and it should be possible to harvest and access them.
3. Metadata descriptions should include persistent identifiers that can be resolved and the CLARIN requirements should hold for the PID system.
4. As far as possible, the linguistic encoding must be related to the data category registry, i.e. an auxiliary resource should be created and made available that includes formally represented mappings between the categories specific to the research data and those that are registered in ISOcat or with new entries registered in the user space of ISOcat that you create if existing categories do not match.
5. Provide proper documentation of the resource.

The plan for a curation project should describe in detail how these different aspects are going to be addressed in the project.

The results of these aspects should be tested by the project participants. The IS will assist with some of these tests. Setting up tests for this should be included in the project plan and the results of these tests will be included in the project's success criteria. Example tests are e.g. a metadata harvesting test and formal procedures such as testing against an XML Schema.

The resulting resource and its metadata must be made available on a server of a recognized CLARIN centre. The project proposal must specify which (candidate) CLARIN centre this will be and concrete arrangements must have been made with this centre.

Depending on the type of resource (text corpora, annotated corpora, lexica, audio, etc.), different expertise from the IS will be needed. If the project is awarded, CLARIN NL will select and assign one or more IS specialists to the project.

5.2 Demonstrator

In a demonstrator project a demonstrator is developed using a documented web-based application based on a technology that the TP currently has at his/her disposal. The development is carried out in close cooperation with the user and the IS.

The project proposal must contain a detailed description of the targeted functionality, including input and output specifications, and how it can contribute to addressing the user's research questions. More generic functionality, i.e. functionality that can serve multiple different research questions from linguistics and humanities research will be preferred over less generic or completely idiosyncratic functionality. See below for additional criteria related to the functionality that will be used to rank proposals. Since a demonstrator project is short in duration and is relatively small, this functionality must already be available to the TP, though perhaps not in the form of a web-based application, and it perhaps only operates on data formats other than the ones listed in the CLARIN standards. The project proposal should contain a detailed description of the functionality in its current state, the targeted web-based application and its components, and a plan to achieve this. The application includes a web-based user interface that takes care of user interactions and method invocations to the core component. An Application Programming Interface (API) to the core component must be provided and documented. The TP must have the rights to make the targeted core component as well as the web-application available on a CLARIN server running at a dedicated centre.

The core component of the web-application must at least be able to operate on the research data and yield output in the formats agreed upon between user and TP. It is a pre if it can apply to other formats from the CLARIN standard list and yield additional output formats. The web-application and its core component will be used to obtain requirements and specifications of the architectural framework that is being worked out in CLARIN and may be used to test it.

The demonstrator consists, as a minimum, of a web application, the research data, and a demonstration scenario. A demonstration scenario is a detailed description of example (sequences of) actions a user can take to have the application applied to the research data and the corresponding system responses in order to get a representative picture of the functionality offered. A movie or sequence of screen captures to illustrate the functionality is nice to have. The application will have to be installed on a CLARIN server, and the project proposal must contain a plan for doing this. It is the intention to have the demonstrator applications available for the lifetime of the CLARIN-

NL project (2009-2014), so occasional support may be needed from the original developers even after the demonstrator project has finished.

The application must be tested with at least one of the common web browsers on the client side (MS IE, Firefox). Agreements about additional technical details (operating system, programming language, workspace requirements, etc) need to be made with the dedicated centre where the services should be executed.

Any vendor, platform or operating system dependent aspects of the application must be made explicit in the proposal and properly documented in the project.

The web-application and its core component should be properly documented, for users (user documentation), for application developers who want to use the core component (documentation of the API), and for technology developers who want to modify or extend the basic functionality of the application.

Auxiliary Resources
The web-application may require data and other software (auxiliary resources) while running.

It must be documented which auxiliary data (e.g. a lexicon) and software (e.g. a library, converters) are needed during runtime for the application. The TP must have the right to make these auxiliary data and software available on a CLARIN server. Any restrictions on their usage (including costs) should be properly documented in the project proposal and in the documentation of the resulting application.

The application and the core component must be able to run on a dedicated CLARIN-server. An application or core component that can run only on a specific (non CLARIN) server (e.g. because it contains auxiliary resources that cannot be made available otherwise) is not acceptable.

6 Metadata

For the web application and its core component, the research data and all runtime auxiliary data used in the application, metadata descriptions must be made in accordance with the CLARIN metadata standard (CMDI). Since the CMDI is currently under development and we hope that the projects of this call can contribute to this development, metadata should be created using the IMDI metadata format. Metadata in IMDI format will be harvestable and can be integrated in the future CMDI-compatible metadata repositories. If the IMDI format is not applicable or suitable, this should be thoroughly justified and documented and discussed with an IS with excellent IMDI expertise. If unavoidable, deviations from this standard are allowed. Any such deviation must be properly documented and be included in the CLARIN Requirements and Desiderata document.

7 Requirements and Desiderata for CLARIN infrastructure

One important result of both demonstrator and curation projects is a document or series of documents describing requirements and desiderata for the CLARIN infrastructure resulting from the experiences gained with the curation of the research data and/or tools, and with the development of the application, its core component and web-services derived from it. These requirements and

desiderata can concern many aspects. The following is a non-exhaustive list of aspects that should be considered:

- Requirements for data formats and encoding standards
- Web-service wrappers
- Metadata elements and formats
- Processing requirements
- Memory requirements
- Network Bandwidth requirements
- User workspace requirements
- API requirements (e.g. Calling conventions)
- IPR / restricted use / ethical issues requirements
- Documentation requirements
- Repository Requirements
- Requirements for registering and resolving PIDs
- Requirements related to semantic interoperability

8 Evaluation Criteria

Proposals for projects will be evaluated and ranked according to criteria listed in the CLARIN-NL Long Term Working Plan. For convenience, these are also listed here:

- Quality
  - Clarity and originality of the project proposal, in particular of the problem and the proposed approach
  - Suitability of the method and plan for the problem at hand
  - Feasibility of the project targets: can they be realized within the specified amount of time and with the instruments proposed?
  - Adequate balance between requested instruments and funds and proposed targets
  - Clearly specified and realistic work plan
  - Conformance to established standards and protocols as supported within CLARIN, or contribute to the development such standards and protocols.
- Project Participants
  - Competence of the participating partners (including their past performance);
  - Balanced cooperation and task assignments within the project. Justification of the composition of the team.
  - Availability of the infrastructure required for the project to be successful
  - Embedding of the work in other research programmes or projects, and/or additional funding from other funding sources is an advantage
- User-orientation of the project
  - Does the project address needs of the targeted infrastructure users (linguists and humanities researchers)?
  - Is there cooperation with or support from the targeted (future) infrastructure users?
  - Is the resulting tool/ service user-friendly, i.e. will non-technical linguistic and humanities researchers be able to use it?

- Is dissemination of the results to the targeted users and (where appropriate) training of them planned?

- Contribution to CLARIN-NL as a whole
  - Conformance to the goals of CLARIN-NL in particular and CLARIN in general and the priorities set within them
  - Contribution to knowledge transfer and network creation. In particular, cooperation between the intended users (linguists and humanities researchers) and technology and service providers (researchers in language and speech technology, computer science, etc.) is an advantage.
- Intellectual Property Rights and Synergy
  - Each proposal must contain clear statements about the situation of the IPR of the data and tools/technologies used, and a detailed plan to resolve any open issues.
  - The project participants have the obligation and must therefore have the rights to incorporate the core data and tools used in a project into the CLARIN infrastructure (this is a sine qua non). There has to be a clear specification and justification of the use of any data or tools needed in the project that cannot be incorporated into the CLARIN infrastructure.
  - Each proposal must show that the submitters have adequate and up-to-date knowledge of data, tools and services that are already available, so that any duplication of effort can be avoided.
- Formal compliance
  - A proposal must meet the formal requirements imposed by the CLARIN-NL organization for proposals, such as
    - conformance to the prescribed format and proposal template
    - submission before the set deadline, using the means prescribed
    - conformance to the prescribed language of the proposal

In addition, more generic applications and data, i.e. applications and data that can serve multiple different research questions will be preferred over less generic applications and data. In this call, a project that does not meet the IPR-requirements stated or is insufficiently clear about it will be considered formally noncompliant.

9 Duration

The duration of the project must be justified. The default maximum duration is 6 months. Any duration longer than 6 months requires thorough justification.

10 Budget

The project budget must be in accordance with the tasks to be carried out, and this must be justified in the project proposal. The maximum budget is 60k€ (the approximate costs of 2 FTEs for half a year).

11 Intellectual Property Rights (IPR)

Ownership of all original data and software remains with the original owners.

An agreement must be in place between the owners of the original data and software and the project participants on the IPR of the adapted data and software before the submission date of a proposal if the owners of the original data and software are not identical to the project participants. If applicable, a copy of this agreement must be uploaded together with the project proposal. Otherwise ownership of the created adaptations and extensions will be with the creator(s).

The project participants have the obligation and therefore must have the rights to make the research data, the application, its core component, and any runtime auxiliary data or software available on a CLARIN server for use by researchers having access to the CLARIN infrastructure. This is a sine qua non. Any proposal not satisfying this requirement or being insufficiently clear about this matter will be considered to be formally noncompliant and will be rejected on these grounds.

The project proposal should describe all issues related to IPR and present solution for them. The relations between the partners in a project must be agreed upon in a consortium agreement before the start of the project.

Practical details

The CLARIN-NL First Open Call is open from Tuesday June 2, 2009. The total budget for this call is limited to a maximum of € 600,000. This call is specifically open for proposals targeting resource curation projects or demonstrator projects. Only those proposals that specifically target this priority will be eligible.

Full proposals must be submitted in English and in PDF format to the CLARIN-NL electronic proposal submission system using the prescribed template (which can be found on the CLARIN-NL website). The deadline for submitting full proposals in this call has been set for Monday August 17, 2009 13:00 hours CET.

The CLARIN-NL electronic proposal submission system can be accessed as of Monday June 22, 2009 via the CLARIN-NL website.

Who can apply?

Applications can be submitted only by researchers affiliated to CLARIN-NL participants that have signed the CLARIN-NL consortium agreement. The list of CLARIN-NL participants is available on the CLARIN-NL website. The main applicant and coordinator of the project must be a researcher from linguistics or humanities more broadly.

CLARIN-NL is in principle open to new participants. If your institute is not currently a CLARIN-NL participant, contact the CLARIN-NL office if you want your organization to become a participant in CLARIN-NL well in advance of the submission deadline.

Applicants who are planning to submit a proposal are strongly advised to contact the CLARIN-NL Office for an eligibility check of their plans.

Eligible costs

- Personnel costs directly related to the project up to a maximum of 60,000€, in accordance with the Akkoord NWO-VSNU 2008 (and any additions to it).

- A fee of maximally 3.000 € per FTE per year (or a pro rata part for less than 1 FTE per year) for covering travel and subsistence costs

Evaluation procedure full proposals
All eligible full proposals submitted in this call will be presented to a panel of international experts in the humanities, language and speech technology and infrastructures (International Advisory Panel, IAP). The composition of the IAP will be published on the CLARIN-NL website as soon as it is available. If the proposals require this, the CLARIN-NL Executive Board can decide to involve additional experts in the evaluation.

This international panel will assess and rank the eligible applications based on the assessment criteria relevant to this call and may formulate a set of recommendations for improving the individual proposals. The panel's assessment and recommendations will be presented to the National Advisory Panel (NAP). The members of the NAP who are not directly involved in the submitted proposals will assess the applications and the IAP's assessment and recommendations and determine the order of priority of the eligible proposals. On the basis of the IAP's advice and the NAP's advice, the CLARIN-NL Board will finally determine which projects will be funded.

Projects should start within two months after the applicant has received the formal notification of funding.

CLARIN-NL consortium agreement
More information as to which legal rules apply for this specific CLARIN-NL granting scheme are laid down in the CLARIN-NL consortium agreement, which can be found on the CLARIN-NL website.

Timetable

Activity	Date
CLARIN-NL Call Open	Tuesday June 2, 2009
CLARIN-NL Kickoff Meeting	Wednesday May 27, 2009
Deadline Proposal Submission	Monday August 17, 2009 13:00hrs
Assessment IAP	Wednesday September 16, 2009
Assessment NAP	Friday October 16, 2009
Decision on Funding by Board	Monday November 2, 2009

CLARIN-NL Organization
The CLARIN-NL project is funded by NWO.

The CLARIN-NL project is coordinated by the Programme Director, prof.dr. J.E.J.M. Odijk, who is a member of the CLARIN-NL Executive Board.

The International Advisory Panel (IAP) is a group of international experts in the areas of humanities, in particular linguistics, language and speech technology, and infrastructures for scientific research.

The National Advisory Panel is a group of national researchers representative for the fields of linguistics and humanities, language and speech technologies and infrastructures for scientific research.

The CLARIN-NL Board consists of national senior researchers with great expertise in governance and/or relevant technical expertise

The composition of these CLARIN-NL governance bodies can be found on the CLARIN-NL website.

CLARIN-NL Office
Contact the CLARIN-NL office for any questions related to this call, e.g.

- Additional clarification
- Advice on eligibility of your plans
- Assistance with finding experts, data or technology required
- Etc.

Jan Odijk
Tel: +31 30 253 6006
Fax: +31 30 253 6000
Postal Address: Janskerkhof 13, 3511 BL Utrecht
Visiting Address: Achter de Dom 24, 3512 JP Utrecht
E-mail: clarinnl@uu.nl (or j.odijk@uu.nl)
CLARIN-NL Website: http://www.hum.uu.nl/clarin-nl/.

Call 2 – Open Call (2010)

CLARIN-NL launches the second open call for project proposals. This call is open from Tuesday July 6, 2010. The total budget for this call is limited to a maximum of € 600,000. This call is specifically open for proposals targeting resource curation projects or demonstrator projects. Such projects take existing data or tools as a basis and attempt to apply CLARIN-supported standards and best practices to make the data and tools CLARIN-compliant. Examples of such projects awarded in 2009 can be found on the CLARIN-NL website. Only those proposals that specifically target this priority will be eligible. The maximum budget per project is 120,000€.

1 Introduction

The first phase of the CLARIN-NL project focuses on the specification and design of the infrastructure. Obviously, the CLARIN infrastructure should be designed in such a way that it can incorporate the data and tools currently used by humanities researchers to address their research questions. On the other hand, the CLARIN infrastructure can only be successful if these data and tools meet certain requirements with regard to standards and interoperability, not only with regard to the data and tools themselves but also with regard to their visibility and accessibility. Many of the data and tools currently in use do not meet these requirements, or meet them only partially. One of the aims of the second call is to make existing data and tools compliant with the requirements on standards and interoperability currently advocated in CLARIN.

However, the list of standards, best practices and interoperability requirements currently advocated by CLARIN has not been completely fixed yet. There are good reasons for this: (1) there may be crucial data or tools for which none of the currently advocated standards or best practices is suited; (2) the currently advocated standards and best practices may be incomplete, insufficiently specific, inconvenient or even incompatible with crucial data and tools. Therefore, a second aim of the second

call is to test whether and to which extent the currently existing data and tools can be made compatible with the standards and best practices currently advocated in CLARIN, and to get a good overview of any incompatibilities as well as suggestions for adaptations of these standards.

These matters can best be investigated by actually attempting the necessary conversions and adaptations. When applied to tools and other software this will result in CLARIN-compliant web applications that can also serve as showcases of the kind of functionality CLARIN aims to offer. A third aim of the second call is therefore to make sure that the applications developed become available in demonstrators which can help promote the particular applications but also the CLARIN infrastructure as a whole.

Finally, by investigating these matters using specific data and tools that are currently in use a lot can be learnt about requirements the CLARIN infrastructure should meet or desiderata that it should offer. The fourth aim of this call is therefore obtaining a detailed list of such requirements and desiderata.

2 Project Types

In the second call, CLARIN-NL therefore solicits (1) projects that carry out resource (data or tools) curation, and (2) demonstrator projects.

Curation projects

CLARIN-NL aims to support the curation of digital language resources (data and tools) so that these resources can participate in the CLARIN infrastructure, more easily be accessed by interested researchers via online methods and become part of appealing new applications. Many language resources are neither visible nor accessible. Visibility is mainly achieved by standardized metadata that are being harvested by service providers. Accessibility has many different aspects:

1. The resource needs to be stored at computers that are accessible.
2. The resource needs to be identified in a persistent manner.
3. The resource needs to be interpretable, which requires a format that adheres to best practices and it requires references to registries were the used concepts are defined.

Demonstrator projects

CLARIN-NL aims to support projects that create appealing showcases of functionality that the CLARIN infrastructure should support. Such projects should make available web applications that can be used as demonstrators of functionality that supports addressing research questions of the CLARIN-NL intended user group. The development of these demonstrators will also be used to inventory a list of requirements the CLARIN infrastructure should meet and desiderata it preferably should offer.

3 Goals

The goal of a curation project is:

- Adapting specific resources so that they are visible, uniquely referable and accessible via the web, and properly documented.

The goal of a demonstrator project is:

- Creating a documented web application starting from an existing tool or application that can be used as a demonstrator and function as a showcase of the type of functionality CLARIN will incorporate and support. Within the web application there must be a clear separation between the web-based user interface and the core component. For the latter an API must be defined and documented.

Important goals common to both types of projects are:

- Applying standards and best practices and makes use of the suggested CLARIN architecture and agreements to understand their limitations and the requirements for extensions.
- Establishing requirements and desiderata for the CLARIN infrastructure

In both project types the use of CLARIN-supported standards and best practices is essential.

The selection of CLARIN-supported standards and best practices is currently ongoing and it is the intention that the projects supported in this call contribute to this. A preliminary set of candidates for CLARIN-supported standards and best practices is available, and though it is by no means final we will refer to this list in this document by the term "CLARIN-standards".

4 Roles

Four roles of persons involved in the projects can be distinguished: the *user*, the *data provider* (DP), the *technology provider* (TP), and the *infrastructure specialist* (IS). The *user* is a researcher from a linguistics or humanities institute who aims to investigate one or more specific research questions. The *DP* has a certain set of digital language-related data at his/her disposal (the 'research data') that can be used for addressing the research questions of the user. The *TP* has a certain technology (e.g. language or speech technology) at its disposal and a thorough understanding of this technology (e.g. because it was developed by the TP). This technology offers functionality that makes it possible to (better) address the user's research question by applying this functionality to the research data. The *infrastructure specialist* (IS) is a specialist who has a deep understanding of the CLARIN service-oriented architecture and its requirements, and/or a specialist in data and tool format standards supported in CLARIN. IS specialists usually get involved in the project via the CLARIN-NL Helpdesk. In many cases the different roles of user, DP, TP and IS will be played by persons from different organizations, but they may originate from a single organization, and occasionally even be played by a single person. CLARIN-NL can offer assistance in bringing the right experts together, if desired.

4.1 User

The user is a researcher from a linguistics or humanities institute who aims to investigate one or more specific research questions. The project proposal should clearly describe the research question(s) of the user, and the research question(s) must be in the domain of research in the humanities in general and the study of language in particular. The project must be led by the user.

4.2 Data Provider

The DP has a certain set of digital language-related data at his/her disposal (the 'research data') that can be used for addressing the research questions of the user. The project proposal should clearly describe the research data the DP has at his/her disposal that can be used to address the research question(s), and how they can be used for this purpose. The research data must be existing digital

language or language-related data. No new research data should be created in the project. The DP must have the right to make the research data available on a CLARIN server running at a dedicated centre. If the data are in a format that is not currently on the list of CLARIN standards, a resource curation project is in order. Otherwise, the data can be used in a demonstrator project. The project proposal should contain a detailed description of the research data, its current state and format, the plans to convert it if needed, justification for using different formats if applicable, and a detailed plan for dealing with the data and its metadata (see below). Any restrictions on the use of the data as well as any ethical issues that apply or may arise must be properly documented in the proposal.

4.3 The Technology Provider

The TP has a certain technology at his/her disposal that can be used as a basis for the development of a web-based application (possibly web-services based) and concomitant demonstrator, or that can be used for resource curation. Since the research data are language data, the technology will in most cases be language or speech technology.

The proposal should contain a detailed description of the available technology and its current status. It should make clear that the TP has a thorough understanding of this technology and describe how the TP obtained this understanding (e.g. because the TP developed the technology).

The intended use of the technology in the project should be described, as well as any extensions or modifications that have to be made to the technology in the project, and a plan to achieve this.

The TP must have the right to use this technology and indicate how it will be used in the project.

4.4 CLARIN-NL Helpdesk and Infrastructure Specialist

CLARIN-NL has set up a Helpdesk that project participants can turn to for all kinds of technical questions related to their project. If the relevant questions cannot be addressed by the Helpdesk functionality (FAQ section, etc.) or staff, the Helpdesk will involve infrastructure specialists. The infrastructure specialist (IS) is a specialist who has a deep understanding of the CLARIN service-oriented architecture and its requirements, and/or a specialist in data, metadata and tool format standards and best practices supported in CLARIN. The Helpdesk, and where needed IS specialists, will advise and assist the project partners.

5 Project Types

5.1 Resource Curation

Resource curation involves a number of different aspects:

1. The resource should be brought into a format that adheres to widely accepted standards and best practices currently considered as likely candidates by CLARIN.
2. Proper metadata descriptions need to be created and made available. They must be compliant with the CLARIN component metadata infrastructure (CMDI) and it should be possible to harvest and access them.
3. Metadata descriptions should include persistent identifiers that can be resolved and the CLARIN requirements should hold for the PID system.
4. The linguistic encoding must be related to the data category registry, i.e. data categories used must be mapped to corresponding ISOCAT data categories where they exist in a

formal way (e.g. via an XML Schema) and new data categories must be added to ISOCAT if they do not exist there yet.

5. Provide proper documentation of the resource, at least in English.

The plan for a curation project should describe in detail how these different aspects are going to be addressed in the project.

The results of these aspects should be tested by the project participants. Setting up tests for this should be included in the project plan and the results of these tests will be included in the project's success criteria. Example tests are e.g. a metadata harvesting test and formal procedures such as testing against an XML Schema.

The resulting resource and its metadata must be made available on a server of a recognized CLARIN centre. The project proposal must specify which (candidate) CLARIN centre this will be and concrete arrangements must have been made with this centre.

5.2 Demonstrator

In a demonstrator project a demonstrator is developed using a documented web-based application based on a technology that the TP currently has at his/her disposal. The development is carried out in close cooperation with the user

The project proposal must contain a detailed description of the targeted functionality, including input and output specifications, and how it can contribute to addressing the user's research questions. More generic functionality, i.e. functionality that can serve multiple different research questions from linguistics and humanities research will be preferred over less generic or completely idiosyncratic functionality. See below for additional criteria related to the functionality that will be used to rank proposals. Since a demonstrator project is short in duration and is relatively small, this functionality must already be available to the TP, though perhaps not in the form of a web-based application, and it perhaps only operates on data formats other than the ones listed in the CLARIN standards. The project proposal should contain a detailed description of the functionality in its current state, the targeted web-based application and its components, and a plan to achieve this. The application includes a web-based user interface that takes care of user interactions and method invocations to the core component. An Application Programming Interface (API) to the core component must be provided and documented. The TP must have the rights to make the targeted core component as well as the web-application available on a CLARIN server running at a dedicated centre.

The core component of the web-application must at least be able to operate on the research data and yield output in the formats agreed upon between user and TP. It is a pre if it can apply to other formats from the CLARIN standard list and yield additional output formats. The web-application and its core component will be used to obtain requirements and specifications of the architectural framework that is being worked out in CLARIN and may be used to test it.

A research data resource often consists of information of various kinds contained in multiple folders and multiple files of varying types. The information contained in such a resource can include documentation, source data, annotations of the source data, aggregate statistics tables on the source data and/or annotations, etc. Any tool (which eventually will very likely be integrated in the

infrastructure as web service) should find out in a fully automated manner whether the research data selected by the user are appropriate input for it, and, if so, that it is applied to the right information (e.g. to the source data but not to the documentation). To achieve this it will integrate wrappers that read and write metadata and provenance information provided by the CLARIN infrastructure. Any requirements or desiderata that follow from this for metadata and data contents and formats should be properly documented in the documents with the requirements and desiderata for the CLARIN infrastructure.

The demonstrator consists, as a minimum, of a web application, the research data, and a demonstration scenario. A demonstration scenario is a detailed description of example (sequences of) actions a user can take to have the application applied to the research data and the corresponding system responses in order to get a representative picture of the functionality offered. A movie or sequence of screen captures to illustrate the functionality is nice to have. The application will have to be installed on a CLARIN server, and the project proposal must contain a plan for doing this. It is the intention to have the demonstrator applications available for the lifetime of the CLARIN-NL project (2009-2014), so occasional support may be needed from the original developers even after the demonstrator project has finished.

The application must be tested with at least one of the common web browsers on the client side (MS IE, Firefox). Agreements about additional technical details (operating system, programming language, workspace requirements, etc) need to be made with the dedicated centre where the services should be executed.

Any vendor, platform or operating system dependent aspects of the application must be made explicit in the proposal and properly documented in the project.

The web-application and its core component should be properly documented, for users (user documentation), for application developers who want to use the core component (documentation of the API), and for technology developers who want to modify or extend the basic functionality of the application. The documentation, as well as the software user interface, must be provided at least in English.

Auxiliary Resources
The web-application may require data and other software (auxiliary resources) while running.

It must be documented which auxiliary data (e.g. a lexicon) and software (e.g. a library, converters) are needed during runtime for the application. The TP must have the right to make these auxiliary data and software available on a CLARIN server. Any restrictions on their usage (including costs) should be properly documented in the project proposal and in the documentation of the resulting application.

The application and the core component must be able to run on a dedicated CLARIN-server. An application or core component that can run only on a specific (non CLARIN) server (e.g. because it contains auxiliary resources that cannot be made available otherwise) is not acceptable.

6 Metadata

For the web application and its core component, the research data and all runtime auxiliary data used in the application, metadata descriptions must be made in accordance with the CLARIN metadata standard (CMDI). CMDI provides a flexible component-based framework for dealing with metadata; the data and tools of the projects of this call may require the development of new CMDI components or the adaptation of existing components and thus can contribute to the further development of the CMDI framework. Any required or desirable extension or modifications of the CMDI framework must be properly documented and be included in the CLARIN Requirements and Desiderata document.

7 Requirements and Desiderata for CLARIN infrastructure

One important result of both demonstrator and curation projects is a document or series of documents describing requirements and desiderata for the CLARIN infrastructure resulting from the experiences gained with the curation of the research data and/or tools, and with the development of the application, its core component and web-services derived from it. These requirements and desiderata can concern many aspects. The following is a non-exhaustive list of aspects that should be considered:

- Requirements for data formats and encoding standards
- Web-service wrappers
- Metadata elements and formats
- Processing requirements
- Memory requirements
- Network Bandwidth requirements
- User workspace requirements
- API requirements (e.g. Calling conventions)
- IPR / restricted use / ethical issues requirements\
- Documentation requirements
- Repository Requirements
- Requirements for registering and resolving PIDs
- Requirements related to semantic interoperability

8 Evaluation Criteria

Proposals are elicited from humanities topics and disciplines, with a priority for the following (sub)disciplines:

1. Literary studies
2. History and political studies
3. Communication and media studies (focused on language data and tools)
4. 1st and 2nd language acquisition
5. Historical linguistics

A proposal must describe a project that is compatible with the requirements mentioned in this call; in particular it must be a resource curation and/or a demonstrator project.

Proposals for projects will furthermore be evaluated and ranked according to the following more general criteria:

- Quality
  - Clarity and added value of the project proposal, in particular of the problem and the proposed approach
  - Suitability of the method and plan for the problem at hand
  - Feasibility of the project targets: can they be realized within the specified amount of time and with the instruments proposed?
  - Adequate balance between requested instruments and funds and proposed targets
  - Clearly specified and realistic work plan
  - Conformance to established standards and protocols as supported within CLARIN, or contribute to the development such standards and protocols.
- Project Participants
  - Competence of the participating partners (including their past performance);
  - Balanced cooperation and task assignments within the project. Justification of the composition of the team.
  - Availability of the infrastructure required for the project to be successful
  - Embedding of the work in other research programmes or projects, and/or additional funding from other funding sources is an advantage
- User-orientation of the project
  - Does the project address needs of the targeted infrastructure users (linguists and humanities researchers)?
  - Projects on tools or data that are widely in use in the targeted user community will score higher than projects focusing on lesser used tools and data.
  - Is there cooperation with or support from the targeted (future) infrastructure users?
  - Is the resulting tool / service user-friendly, i.e. will non-technical linguistic and humanities researchers be able to use it?
  - Is dissemination of the results to the targeted users and (where appropriate) training of them planned?
- Contribution to CLARIN-NL as a whole
  - Conformance to the goals of CLARIN-NL in particular and CLARIN in general and the priorities set within them
  - Contribution to knowledge transfer and network creation. In particular, cooperation between the intended users (linguists and humanities researchers) and technology and service providers (researchers in language and speech technology, computer science, etc.) is an advantage.
- Intellectual Property Rights and Synergy
  - Each proposal must contain clear statements about the situation of the IPR of the data and tools/technologies used, and a detailed plan to resolve any open issues.
  - The project participants have the obligation and must therefore have the rights to incorporate the core data and tools used in a project into the CLARIN infrastructure (this is a sine qua non). There has to be a clear specification and justification of the use of any data or tools needed in the project that cannot be incorporated into the CLARIN infrastructure.

- Each proposal must show that the submitters have adequate and up-to-date knowledge of data, tools and services that are already available, so that any duplication of effort can be avoided.

- Formal compliance
  - A proposal must meet the formal requirements imposed by the CLARIN-NL organization for proposals, such as
    - conformance to the prescribed format and proposal template
    - submission before the set deadline, using the means prescribed
    - conformance to the prescribed language of the proposal

In addition, more generic applications and data, i.e. applications and data that can serve multiple different research questions will be preferred over less generic applications and data. Projects that are part of or fit in with international cooperation with partners from CLARIN in other countries will be preferred (of course, foreign organizations will have to find their own funding). Finally, a project that does not meet the IPR-requirements stated or is insufficiently clear about it will be considered formally noncompliant.

9 Duration

The duration of the project must be justified. The default maximum duration is 12 months. Any duration longer than 12 months requires thorough justification.

10 Budget

The project budget must be in accordance with the tasks to be carried out, and this must be justified in the project proposal. The maximum budget is 120k€.

11 Intellectual Property Rights (IPR)

Ownership of all original data and software remains with the original owners.

An agreement must be in place between the owners of the original data and software and the project participants on the IPR of the adapted data and software before the submission date of a proposal if the owners of the original data and software are not identical to the project participants. If applicable, a copy of this agreement must be uploaded together with the project proposal. Otherwise ownership of the created adaptations and extensions will be with the creator(s).

The project participants have the obligation and therefore must have the rights to make the research data, the application, its core component, and any runtime auxiliary data or software available on a CLARIN server for use by researchers having access to the CLARIN infrastructure. This is a sine qua non. Any proposal not satisfying this requirement or being insufficiently clear about this matter will be considered to be formally noncompliant and will be rejected on these grounds.

The project proposal should describe all issues related to IPR and present solution for them. The relations between the partners in a project must be agreed upon in a consortium agreement before the start of the project.

Practical details

The CLARIN-NL Second Open Call is open from Tuesday July 6, 2010. The total budget for this call is limited to a maximum of € 600,000. This call is specifically open for proposals targeting resource curation projects or demonstrator projects. Only those proposals that specifically target this priority will be eligible.

Full proposals must be submitted in English and in PDF format to the CLARIN-NL electronic proposal submission system using the prescribed template (which can be found on the CLARIN-NL website). The deadline for submitting full proposals in this call has been set for Wednesday September 29, 2010 13:00 hours CET.

The CLARIN-NL electronic proposal submission system can be accessed as of Tuesday September 7, 2010 via the CLARIN-NL website. If the electronic proposal submission system would not work, send proposals via e-mail to clarinnl@uu.nl.

Who can apply?

Applications can be submitted only by researchers affiliated to CLARIN-NL participants that have signed the CLARIN-NL consortium agreement. The list of CLARIN-NL participants is available on the CLARIN-NL website. The main applicant and coordinator of the project must be a researcher from linguistics or humanities more broadly.

CLARIN-NL is in principle open to new participants. If your institute is not currently a CLARIN-NL participant, contact the CLARIN-NL office if you want your organization to become a participant in CLARIN-NL well in advance of the submission deadline.

Applicants who are planning to submit a proposal are strongly advised to contact the CLARIN-NL Office for an eligibility check of their plans, or for assistance in finding suitable partners.

Eligible costs

- Personnel costs directly related to the project, in accordance with the Akkoord NWO-VSNU 2008 (and any additions to it).
- A fee of maximally 3.000 € per FTE per year (or a pro rata part for less than 1 FTE per year) for covering travel and subsistence costs
- The requested funding cannot exceed 120,000€

Evaluation procedure full proposals

All eligible full proposals submitted in this call will be presented to a panel of international experts in the humanities, language and speech technology and infrastructures (International Advisory Panel, IAP). The composition of the IAP can be found on the CLARIN-NL website. If the proposals require this, the CLARIN-NL Executive Board can decide to involve additional experts in the evaluation.

This international panel will assess and rank the eligible applications based on the assessment criteria relevant to this call and may formulate a set of recommendations for improving the individual proposals. The panel's assessment and recommendations will be presented to the National Advisory Panel (NAP). The members of the NAP who are not directly involved in the submitted proposals will assess the applications and the IAP's assessment and recommendations and determine the order of

priority of the eligible proposals. On the basis of the IAP's advice and the NAP's advice, the CLARIN-NL Board will finally determine which projects will be funded.

Projects should start within two months after the applicant has received the formal notification of funding.

CLARIN-NL consortium agreement
More information as to which legal rules apply for this specific CLARIN-NL granting scheme are laid down in the CLARIN-NL consortium agreement, which can be found on the CLARIN-NL website (http://www.clarin.nl/node/72).

Timetable

Activity	Date
CLARIN-NL Second Call Open	2010-07-06
CLARIN-NL Second Call Information Session	2010-08-26 (afternoon)
Deadline Proposal Submission	2010-09-29 13:00hrs

The exact deadlines of the assessments by the IAP and NAP, and the final decision by the board will be communicated later via the CLARIN-NL website. The target is to make the final decisions before the end of 2010.

CLARIN-NL Organization
The CLARIN-NL project is funded by NWO.

The CLARIN-NL project is coordinated by the Programme Director, prof.dr. J.E.J.M. Odijk, who is a member of the CLARIN-NL Executive Board. Drs. Erica Renckens is the CLARIN-NL project secretary.

The International Advisory Panel (IAP) is a group of international experts in the areas of humanities, in particular linguistics, language and speech technology, and infrastructures for scientific research.

The National Advisory Panel is a group of national researchers representative for the fields of linguistics and humanities, language and speech technologies and infrastructures for scientific research.

The CLARIN-NL Board consists of national senior researchers with great expertise in governance and/or relevant technical expertise

The composition of these CLARIN-NL governance bodies can be found on the CLARIN-NL website.

CLARIN-NL Office
Contact the CLARIN-NL office for any questions related to this call, e.g.

- Additional clarification
- Advice on eligibility of your plans
- Assistance with finding experts, data or technology required
- Assistance with finding project partners
- Etc.

Jan Odijk/Erica Renckens
Tel: +31 30 253 6730
Fax: +31 30 253 6000
Postal Address: Janskerkhof 13, 3511 BL Utrecht
Visiting Address: Achter de Dom 24, 3512 JP Utrecht
E-mail: clarinnl@uu.nl
CLARIN-NL Website: http://www.clarin.nl.

Call 3 (2011)
Open Call

CLARIN-NL launches the third open call for project proposals. This call is open from Friday July 3, 2011. The total budget for this call is limited to a maximum of € 600,000. This call is specifically open for proposals targeting resource curation projects or demonstrator projects. Such projects take existing data or tools as a basis and attempt to apply CLARIN-supported standards and best practices to make the data and tools CLARIN-compliant. Examples of such projects awarded in 2009 and 2010 can be found on the CLARIN-NL website. Only those proposals that specifically target this priority will be eligible. The maximum budget per project is 120,000€.

1 Introduction

The first phase of the CLARIN-NL project focuses on the specification and design of the infrastructure. Obviously, the CLARIN infrastructure should be designed in such a way that it can incorporate the data and tools currently used by humanities researchers to address their research questions. On the other hand, the CLARIN infrastructure can only be successful if these data and tools meet certain requirements with regard to standards and interoperability, not only with regard to the data and tools themselves but also with regard to their visibility and accessibility. Many of the data and tools currently in use do not meet these requirements, or meet them only partially. One of the aims of the third call is to make existing data and tools compliant with the requirements on standards and interoperability currently advocated in CLARIN.

However, the list of standards, best practices and interoperability requirements currently advocated by CLARIN has not been completely fixed yet. There are good reasons for this: (1) there may be crucial data or tools for which none of the currently advocated standards or best practices is suited; (2) the currently advocated standards and best practices may be incomplete, insufficiently specific, inconvenient or even incompatible with crucial data and tools. Therefore, a second aim of the third call is to test whether and to which extent the currently existing data and tools can be made compatible with the standards and best practices currently advocated in CLARIN, and to get a good overview of any incompatibilities as well as suggestions for adaptations of these standards.

These matters can best be investigated by actually attempting the necessary conversions and adaptations. When applied to tools and other software this will result in CLARIN-compliant web applications that can also serve as showcases of the kind of functionality CLARIN aims to offer. A third aim of the third call is therefore to make sure that the applications developed become available in demonstrators which can help promote the particular applications but also the CLARIN infrastructure as a whole.

Finally, by investigating these matters using specific data and tools that are currently in use a lot can be learnt about requirements the CLARIN infrastructure should meet or desiderata that it should offer. The fourth aim of this call is therefore obtaining a detailed list of such requirements and desiderata.

2 Project Types

In the third call, CLARIN-NL therefore solicits (1) projects that carry out resource (data or tools) curation, and (2) demonstrator projects.

Curation projects

CLARIN-NL aims to support the curation of digital language resources (data and tools) so that these resources can participate in the CLARIN infrastructure, more easily be accessed by interested researchers via online methods and become part of appealing new applications. Many language resources are neither visible nor accessible. Visibility is mainly achieved by standardized metadata that are being harvested by service providers. Accessibility has many different aspects:

1. The resource needs to be stored at computers that are accessible.
2. The resource needs to be identified in a persistent manner.
3. The resource needs to be interpretable, which requires a format that adheres to best practices and it requires references to registries were the used concepts are defined.

Demonstrator projects

CLARIN-NL aims to support projects that create appealing showcases of functionality that the CLARIN infrastructure should support. Such projects should make available web applications that can be used as demonstrators of functionality that supports addressing research questions of the CLARIN-NL intended user group. The development of these demonstrators will also be used to inventory a list of requirements the CLARIN infrastructure should meet and desiderata it preferably should offer.

3 Goals

The goal of a curation project is:

- Adapting specific resources so that they are visible, uniquely referable and accessible via the web, and properly documented.

The goal of a demonstrator project is:

- Creating a documented web application starting from an existing tool or application that can be used as a demonstrator and function as a showcase of the type of functionality CLARIN will incorporate and support. Within the web application there must be a clear separation between the web-based user interface and the core component. For the latter an API must be defined and documented.

Important goals common to both types of projects are:

- Applying standards and best practices and makes use of the suggested CLARIN architecture and agreements to understand their limitations and the requirements for extensions.
- Establishing requirements and desiderata for the CLARIN infrastructure

In both project types the use of CLARIN-supported standards and best practices is essential.

The selection of CLARIN-supported standards and best practices is currently ongoing and it is the intention that the projects supported in this call contribute to this. A preliminary set of candidates for CLARIN-supported standards and best practices is available, and though it is by no means final we will refer to this list in this document by the term "CLARIN-standards".

4 Roles

Four roles of persons involved in the projects can be distinguished: the *user*, the *data provider* (DP), the *technology provider* (TP), and the *infrastructure specialist* (IS). The *user* is a researcher from a linguistics or humanities institute who aims to investigate one or more specific research questions. The *DP* has a certain set of digital language-related data at his/her disposal (the 'research data') that can be used for addressing the research questions of the user. The *TP* has a certain technology (e.g. language or speech technology) at its disposal and a thorough understanding of this technology (e.g. because it was developed by the TP). This technology offers functionality that makes it possible to (better) address the user's research question by applying this functionality to the research data. The *infrastructure specialist* (IS) is a specialist who has a deep understanding of the CLARIN service-oriented architecture and its requirements, and/or a specialist in data and tool format standards supported in CLARIN. IS specialists usually get involved in the project via the CLARIN-NL Helpdesk. In many cases the different roles of user, DP, TP and IS will be played by persons from different organizations, but they may originate from a single organization, and occasionally even be played by a single person. CLARIN-NL can offer assistance in bringing the right experts together, if desired.

4.1 User

The user is a researcher from a linguistics or humanities institute who aims to investigate one or more specific research questions. The project proposal should clearly describe the research question(s) of the user, and the research question(s) must be in the domain of research in the humanities in general and the study of language in particular. The project must be led by the user.

4.2 Data Provider

The DP has a certain set of digital language-related data at his/her disposal (the 'research data') that can be used for addressing the research questions of the user. The project proposal should clearly describe the research data the DP has at his/her disposal that can be used to address the research question(s), and how they can be used for this purpose. The research data must be existing digital language or language-related data. No new research data should be created in the project. The DP must have the right to make the research data available on a CLARIN server running at a dedicated centre. If the data are in a format that is not currently on the list of CLARIN standards, a resource curation project is in order. Otherwise, the data can be used in a demonstrator project. The project proposal should contain a detailed description of the research data, its current state and format, the plans to convert it if needed, justification for using different formats if applicable, and a detailed plan for dealing with the data and its metadata (see below). Any restrictions on the use of the data as well as any ethical issues that apply or may arise must be properly documented in the proposal.

4.3 The Technology Provider

The TP has a certain technology at his/her disposal that can be used as a basis for the development of a web-based application (possibly web-services based) and concomitant demonstrator, or that can

be used for resource curation. Since the research data are language data, the technology will in most cases be language or speech technology.

The proposal should contain a detailed description of the available technology and its current status. It should make clear that the TP has a thorough understanding of this technology and describe how the TP obtained this understanding (e.g. because the TP developed the technology).

The intended use of the technology in the project should be described, as well as any extensions or modifications that have to be made to the technology in the project, and a plan to achieve this.

The TP must have the right to use this technology and indicate how it will be used in the project.

4.4 CLARIN-NL Helpdesk and Infrastructure Specialist

CLARIN-NL has set up a Helpdesk that project participants can turn to for all kinds of technical questions related to their project. If the relevant questions cannot be addressed by the Helpdesk functionality (FAQ section, etc.) or staff, the Helpdesk will involve infrastructure specialists. The infrastructure specialist (IS) is a specialist who has a deep understanding of the CLARIN service-oriented architecture and its requirements, and/or a specialist in data, metadata and tool format standards and best practices supported in CLARIN. The Helpdesk, and where needed IS specialists, will advise and assist the project partners.

5 Project Types

5.1 Resource Curation

Resource curation involves a number of different aspects:

1. The resource should be brought into a format that adheres to widely accepted standards and best practices currently considered as likely candidates by CLARIN.
2. Proper metadata descriptions need to be created and made available. They must be compliant with the CLARIN component metadata infrastructure (CMDI) and it should be possible to harvest and access them.
3. Metadata descriptions should include persistent identifiers that can be resolved and the CLARIN requirements should hold for the PID system.
4. The linguistic encoding must be related to the data category registry, i.e. data categories used must be mapped to corresponding ISOCAT data categories where they exist in a formal way (e.g. via an XML Schema) and new data categories must be added to ISOCAT if they do not exist there yet.
5. Provide proper documentation of the resource, at least in English.

The plan for a curation project should describe in detail how these different aspects are going to be addressed in the project.

The results of these aspects should be tested by the project participants. Setting up tests for this should be included in the project plan and the results of these tests will be included in the project's success criteria. Example tests are e.g. a metadata harvesting test and formal procedures such as testing against an XML Schema.

The resulting resource and its metadata must be made available on a server of a recognized CLARIN centre. The project proposal must specify which (candidate) CLARIN centre this will be and concrete arrangements must have been made with this centre.

5.2 Demonstrator

In a demonstrator project a demonstrator is developed using a documented web-based application based on a technology that the TP currently has at his/her disposal. The development is carried out in close cooperation with the user

The project proposal must contain a detailed description of the targeted functionality, including input and output specifications, and how it can contribute to addressing the user's research questions. More generic functionality, i.e. functionality that can serve multiple different research questions from linguistics and humanities research will be preferred over less generic or completely idiosyncratic functionality. See below for additional criteria related to the functionality that will be used to rank proposals. Since a demonstrator project is short in duration and is relatively small, this functionality must already be available to the TP, though perhaps not in the form of a web-based application, and it perhaps only operates on data formats other than the ones listed in the CLARIN standards. The project proposal should contain a detailed description of the functionality in its current state, the targeted web-based application and its components, and a plan to achieve this. The application includes a web-based user interface that takes care of user interactions and method invocations to the core component. An Application Programming Interface (API) to the core component must be provided and documented. The TP must have the rights to make the targeted core component as well as the web-application available on a CLARIN server running at a dedicated centre.

The core component of the web-application must at least be able to operate on the research data and yield output in the formats agreed upon between user and TP. It is a pre if it can apply to other formats from the CLARIN standard list and yield additional output formats. The web-application and its core component will be used to obtain requirements and specifications of the architectural framework that is being worked out in CLARIN and may be used to test it.

A research data resource often consists of information of various kinds contained in multiple folders and multiple files of varying types. The information contained in such a resource can include documentation, source data, annotations of the source data, aggregate statistics tables on the source data and/or annotations, etc. Any tool (which eventually will very likely be integrated in the infrastructure as web service) should find out in a fully automated manner whether the research data selected by the user are appropriate input for it, and, if so, that it is applied to the right information (e.g. to the source data but not to the documentation). To achieve this it will integrate wrappers that read and write metadata and provenance information provided by the CLARIN infrastructure. A popular wrapper, used a lot and supported in CLARIN-NL is CLAM, developed by Tilburg University. Any requirements or desiderata that follow from this for metadata and data contents and formats should be properly documented in the documents with the requirements and desiderata for the CLARIN infrastructure.

The demonstrator consists, as a minimum, of a web application, the research data, and a demonstration scenario. A demonstration scenario is a detailed description of example (sequences

of) actions a user can take to have the application applied to the research data and the corresponding system responses in order to get a representative picture of the functionality offered. A movie or sequence of screen captures to illustrate the functionality is nice to have. The application will have to be installed on a CLARIN server, and the project proposal must contain a plan for doing this. It is the intention to have the demonstrator applications available for the lifetime of the CLARIN-NL project (2009-2014), so occasional support may be needed from the original developers even after the demonstrator project has finished.

The application must be tested with at least one of the common web browsers on the client side (MS IE, Firefox). Agreements about additional technical details (operating system, programming language, workspace requirements, etc) need to be made with the dedicated centre where the services should be executed.

Any vendor, platform or operating system dependent aspects of the application must be made explicit in the proposal and properly documented in the project.

The web-application and its core component should be properly documented, for users (user documentation), for application developers who want to use the core component (documentation of the API), and for technology developers who want to modify or extend the basic functionality of the application. The documentation, as well as the software user interface, must be provided at least in English.

Auxiliary Resources

The web-application may require data and other software (auxiliary resources) while running.

It must be documented which auxiliary data (e.g. a lexicon) and software (e.g. a library, converters) are needed during runtime for the application. The TP must have the right to make these auxiliary data and software available on a CLARIN server. Any restrictions on their usage (including costs) should be properly documented in the project proposal and in the documentation of the resulting application.

The application and the core component must be able to run on a dedicated CLARIN-server. An application or core component that can run only on a specific (non CLARIN) server (e.g. because it contains auxiliary resources that cannot be made available otherwise) is not acceptable.

6 Metadata

For the web application and its core component, the research data and all runtime auxiliary data used in the application, metadata descriptions must be made in accordance with the CLARIN metadata standard (CMDI). CMDI provides a flexible component-based framework for dealing with metadata; the data and tools of the projects of this call may require the development of new CMDI components or the adaptation of existing components and thus can contribute to the further development of the CMDI framework. Any required or desirable extension or modifications of the CMDI framework must be properly documented and be included in the CLARIN Requirements and Desiderata document.

7 Requirements and Desiderata for CLARIN infrastructure

One important result of both demonstrator and curation projects is a document or series of documents describing requirements and desiderata for the CLARIN infrastructure resulting from the experiences gained with the curation of the research data and/or tools, and with the development of the application, its core component and web-services derived from it. These requirements and desiderata can concern many aspects. The following is a non-exhaustive list of aspects that should be considered:

- Requirements for data formats and encoding standards
- Web-service wrappers
- Metadata elements and formats
- Processing requirements
- Memory requirements
- Network Bandwidth requirements
- User workspace requirements
- API requirements (e.g. Calling conventions)
- IPR / restricted use / ethical issues requirements
- Documentation requirements
- Repository Requirements
- Requirements for registering and resolving PIDs
- Requirements related to semantic interoperability

8 Evaluation Criteria

Proposals are elicited from humanities topics and disciplines, with a priority for the following (sub)disciplines:

1. Literary studies
2. Communication and media studies (focused on the new media such as texting (sms), chat, blogs, twitter, etc.)
3. History
4. Linguistics, in particular
    a. Syntax
    b. Phonology
    c. Semantics
    d. Descriptive linguistics and language documentation
    e. Discourse

A proposal must describe a project that is compatible with the requirements mentioned in this call; in particular it must be a resource curation and/or a demonstrator project.

Proposals for projects will furthermore be evaluated and ranked according to the following more general criteria:

- Quality
    - Clarity and added value of the project proposal, in particular of the problem and the proposed approach

  - Suitability of the method and plan for the problem at hand
  - Feasibility of the project targets: can they be realized within the specified amount of time and with the instruments proposed?
  - Adequate balance between requested instruments and funds and proposed targets
  - Clearly specified and realistic work plan
  - Conformance to established standards and protocols as supported within CLARIN, or contribute to the development such standards and protocols.
- Project Participants
  - Competence of the participating partners (including their past performance);
  - Balanced cooperation and task assignments within the project. Justification of the composition of the team.
  - Availability of the infrastructure required for the project to be successful
  - Embedding of the work in other research programmes or projects, and/or additional funding from other funding sources is an advantage
- User-orientation of the project
  - Does the project address needs of the targeted infrastructure users (linguists and humanities researchers)?
  - Projects on tools or data that are widely in use in the targeted user community will score higher than projects focusing on lesser used tools and data.
  - Is there cooperation with or support from the targeted (future) infrastructure users?
  - Is the resulting tool / service user-friendly, i.e. will non-technical linguistic and humanities researchers be able to use it?
  - Is dissemination of the results to the targeted users and (where appropriate) training of them planned?
- Contribution to CLARIN-NL as a whole
  - Conformance to the goals of CLARIN-NL in particular and CLARIN in general and the priorities set within them
  - Contribution to knowledge transfer and network creation. In particular, cooperation between the intended users (linguists and humanities researchers) and technology and service providers (researchers in language and speech technology, computer science, etc.) is an advantage.
- Intellectual Property Rights and Synergy
  - Each proposal must contain clear statements about the situation of the IPR of the data and tools/technologies used, and a detailed plan to resolve any open issues.
  - The project participants have the obligation and must therefore have the rights to incorporate the core data and tools used in a project into the CLARIN infrastructure (this is a sine qua non). There has to be a clear specification and justification of the use of any data or tools needed in the project that cannot be incorporated into the CLARIN infrastructure.
  - Each proposal must show that the submitters have adequate and up-to-date knowledge of data, tools and services that are already available, so that any duplication of effort can be avoided.
- Formal compliance

- A proposal must meet the formal requirements imposed by the CLARIN-NL organization for proposals, such as
  - conformance to the prescribed format and proposal template
  - submission before the set deadline, using the means prescribed
  - conformance to the prescribed language of the proposal

In addition, more generic applications and data, i.e. applications and data that can serve multiple different research questions will be preferred over less generic applications and data. Projects that are part of or fit in with international cooperation with partners from CLARIN in other countries will be preferred (of course, foreign organizations will have to find their own funding). Finally, a project that does not meet the IPR-requirements stated or is insufficiently clear about it will be considered formally noncompliant.

9 Duration

The duration of the project must be justified. The default maximum duration is 12 months. Any duration longer than 12 months requires thorough justification.

10 Budget

The project budget must be in accordance with the tasks to be carried out, and this must be justified in the project proposal. The maximum budget is 120k€.

11 Intellectual Property Rights (IPR)

Ownership of all original data and software remains with the original owners.

An agreement must be in place between the owners of the original data and software and the project participants on the IPR of the adapted data and software before the submission date of a proposal if the owners of the original data and software are not identical to the project participants. If applicable, a copy of this agreement must be uploaded together with the project proposal. Otherwise ownership of the created adaptations and extensions will be with the creator(s).

The project participants have the obligation and therefore must have the rights to make the research data, the application, its core component, and any runtime auxiliary data or software available on a CLARIN server for use by researchers having access to the CLARIN infrastructure. This is a sine qua non. Any proposal not satisfying this requirement or being insufficiently clear about this matter will be considered to be formally noncompliant and will be rejected on these grounds.

The project proposal should describe all issues related to IPR and present solutions for them. The relations between the partners in a project must be agreed upon in a consortium agreement before the start of the project.

Practical details

The CLARIN-NL Third Open Call is open from Friday July 3, 2011. The total budget for this call is limited to a maximum of € 600,000. This call is specifically open for proposals targeting resource curation projects or demonstrator projects. Only those proposals that specifically target this priority will be eligible.

Full proposals must be submitted in English and in PDF format to the CLARIN-NL electronic proposal submission system using the prescribed template (which can be found on the CLARIN-NL website). The deadline for submitting full proposals in this call has been set for Friday September 30, 2011 13:00 hours CET.

The CLARIN-NL electronic proposal submission system can be accessed as of July 3, 2011 via the CLARIN-NL website. If the electronic proposal submission system would not work, contact the CLARIN-NL Office clarinnl@uu.nl.

Who can apply?
Applications can be submitted only by researchers affiliated to CLARIN-NL participants that have signed the CLARIN-NL consortium agreement. The list of CLARIN-NL participants is available on the CLARIN-NL website. The main applicant and coordinator of the project must be a researcher from linguistics or humanities more broadly.

CLARIN-NL is in principle open to new participants. If your institute is not currently a CLARIN-NL participant, contact the CLARIN-NL office if you want your organization to become a participant in CLARIN-NL well in advance of the submission deadline.

Applicants who are planning to submit a proposal are strongly advised to contact the CLARIN-NL Office for an eligibility check of their plans, or for assistance in finding suitable partners.

Eligible costs

- Personnel costs directly related to the project, in accordance with the Akkoord NWO-VSNU 2008 (and any additions to it).
- A fee of maximally 3.000 € per FTE per year (or a pro rata part for less than 1 FTE per year) for covering travel and subsistence costs
- The requested funding cannot exceed 120,000€

Evaluation procedure full proposals
All eligible full proposals submitted in this call will be presented to a panel of international experts in the humanities, language and speech technology and infrastructures (International Advisory Panel, IAP). The composition of the IAP can be found on the CLARIN-NL website. If the proposals require this, the CLARIN-NL Executive Board can decide to involve additional experts in the evaluation.

This international panel will assess and rank the eligible applications based on the assessment criteria relevant to this call and may formulate a set of recommendations for improving the individual proposals. The panel's assessment and recommendations will be presented to the National Advisory Panel (NAP). The members of the NAP who are not directly involved in the submitted proposals will also assess the applications and the IAP's assessment and recommendations and make its own recommendations. On the basis of the IAP's advice and the NAP's advice, the CLARIN-NL Board will finally determine which projects will be funded.

Projects should start within three months after the applicant has received the formal notification of funding.

CLARIN-NL consortium agreement

More information as to which legal rules apply for this specific CLARIN-NL granting scheme are laid down in the CLARIN-NL consortium agreement, which can be found on the CLARIN-NL website (http://www.clarin.nl/node/72).

Timetable

Activity	**Date**
CLARIN-NL Third Call Open	Sunday July 3, 2011
CLARIN-NL Third Call Information Session	Thursday August 25, 2011 (afternoon)
Deadline Proposal Submission	Friday September 30, 2011 13:00hrs CET

The exact deadlines of the assessments by the IAP and NAP, and the final decision by the board will be communicated later via the CLARIN-NL website. The target is to make the final decisions before the end of 2011.

CLARIN-NL Organization

The CLARIN-NL project is funded by NWO.

The CLARIN-NL project is coordinated by the Programme Director, prof.dr. J.E.J.M. Odijk, who is a member of the CLARIN-NL Executive Board. Drs. Jolien Scholten is the CLARIN-NL project secretary.

The International Advisory Panel (IAP) is a group of international experts in the areas of humanities, in particular linguistics, language and speech technology, and infrastructures for scientific research.

The National Advisory Panel is a group of national researchers representative for the fields of linguistics and humanities, language and speech technologies and infrastructures for scientific research.

The CLARIN-NL Board consists of national senior researchers with great expertise in governance and/or relevant technical expertise

The composition of these CLARIN-NL governance bodies can be found on the CLARIN-NL website.

CLARIN-NL Helpdesk

Contact the CLARIN-NL Helpdesk for any technical questions related to this call, e.g.

- Questions on metadata and CMDI
- Questions on data categories en ISOCAT
- Questions on standards supported in CLARIN
- Questions on web services and CLAM.
- Etc.

e-mail: helpdesk@clarin.nl

CLARIN-NL Centres

A list of the current candidate CLARIN centres and their contact persons can be found on the CLARIN-NL website: http://www.clarin.nl/node/130

CLARIN-NL Office

Contact the CLARIN-NL office for any organizational or practical questions related to this call, e.g.

- Additional clarification
- Advice on eligibility of your plans
- Assistance with finding experts, data or technology required
- Assistance with finding project partnersA
- ssistance with selecting a CLARIN Centre.
- Etc.

Jan Odijk/Jolien Scholten
Tel: +31 30 253 6279
Fax: +31 30 253 6000
Address: Trans 10, 3512 JK Utrecht
E-mail: clarinnl@uu.nl
CLARIN-NL Website: http://www.clarin.nl.

Closed call

CLARIN-NL launches in its third call a Closed Call for project proposals. This called is only open for researchers who have been explicitly invited to submit a project proposal for the closed call by the CLARIN-NL Board. The call is open from Friday July 3, 2011. The total budget for this closed call is limited to a maximum of € 400,000. The call is specifically open for proposals targeting resource curation projects or demonstrator projects. Such projects take existing data or tools as a basis and attempt to apply CLARIN-supported standards and best practices to make the data and tools CLARIN-compliant. Examples of such projects awarded in 2009 and 2010 can be found on the CLARIN-NL website. Only those proposals that specifically target this priority will be eligible. The maximum budget per project is 80,000€.

1 Introduction

The first phase of the CLARIN-NL project focuses on the specification and design of the infrastructure. Obviously, the CLARIN infrastructure should be designed in such a way that it can incorporate the data and tools currently used by humanities researchers to address their research questions. On the other hand, the CLARIN infrastructure can only be successful if these data and tools meet certain requirements with regard to standards and interoperability, not only with regard to the data and tools themselves but also with regard to their visibility and accessibility. Many of the data and tools currently in use do not meet these requirements, or meet them only partially. One of the aims of the third call is to make existing data and tools compliant with the requirements on standards and interoperability currently advocated in CLARIN.

However, the list of standards, best practices and interoperability requirements currently advocated by CLARIN has not been completely fixed yet. There are good reasons for this: (1) there may be

crucial data or tools for which none of the currently advocated standards or best practices is suited; (2) the currently advocated standards and best practices may be incomplete, insufficiently specific, inconvenient or even incompatible with crucial data and tools. Therefore, a second aim of the third call is to test whether and to which extent the currently existing data and tools can be made compatible with the standards and best practices currently advocated in CLARIN, and to get a good overview of any incompatibilities as well as suggestions for adaptations of these standards.

These matters can best be investigated by actually attempting the necessary conversions and adaptations. When applied to tools and other software this will result in CLARIN-compliant web applications that can also serve as showcases of the kind of functionality CLARIN aims to offer. A third aim of the third call is therefore to make sure that the applications developed become available in demonstrators which can help promote the particular applications but also the CLARIN infrastructure as a whole.

Finally, by investigating these matters using specific data and tools that are currently in use a lot can be learnt about requirements the CLARIN infrastructure should meet or desiderata that it should offer. The fourth aim of this call is therefore obtaining a detailed list of such requirements and desiderata.

2 Project Types

In the third call, CLARIN-NL therefore solicits (1) projects that carry out resource (data or tools) curation, and (2) demonstrator projects.

Curation projects

CLARIN-NL aims to support the curation of digital language resources (data and tools) so that these resources can participate in the CLARIN infrastructure, more easily be accessed by interested researchers via online methods and become part of appealing new applications. Many language resources are neither visible nor accessible. Visibility is mainly achieved by standardized metadata that are being harvested by service providers. Accessibility has many different aspects:

1. The resource needs to be stored at computers that are accessible.
2. The resource needs to be identified in a persistent manner.
3. The resource needs to be interpretable, which requires a format that adheres to best practices and it requires references to registries were the used concepts are defined.

Demonstrator projects

CLARIN-NL aims to support projects that create appealing showcases of functionality that the CLARIN infrastructure should support. Such projects should make available web applications that can be used as demonstrators of functionality that supports addressing research questions of the CLARIN-NL intended user group. The development of these demonstrators will also be used to inventory a list of requirements the CLARIN infrastructure should meet and desiderata it preferably should offer.

3 Goals

The goal of a curation project is:

- Adapting specific resources so that they are visible, uniquely referable and accessible via the web, and properly documented.

The goal of a demonstrator project is:

- Creating a documented web application starting from an existing tool or application that can be used as a demonstrator and function as a showcase of the type of functionality CLARIN will incorporate and support. Within the web application there must be a clear separation between the web-based user interface and the core component. For the latter an API must be defined and documented.

Important goals common to both types of projects are:

- Applying standards and best practices and makes use of the suggested CLARIN architecture and agreements to understand their limitations and the requirements for extensions.
- Establishing requirements and desiderata for the CLARIN infrastructure

In both project types the use of CLARIN-supported standards and best practices is essential.

The selection of CLARIN-supported standards and best practices is currently ongoing and it is the intention that the projects supported in this call contribute to this. A preliminary set of candidates for CLARIN-supported standards and best practices is available, and though it is by no means final we will refer to this list in this document by the term “CLARIN-standards”.

4 Roles

Four roles of persons involved in the projects can be distinguished: the *user*, the *data provider* (DP), the *technology provider* (TP), and the *infrastructure specialist* (IS). The *user* is a researcher from a linguistics or humanities institute who aims to investigate one or more specific research questions. The *DP* has a certain set of digital language-related data at his/her disposal (the ‘research data’) that can be used for addressing the research questions of the user. The *TP* has a certain technology (e.g. language or speech technology) at its disposal and a thorough understanding of this technology (e.g. because it was developed by the TP). This technology offers functionality that makes it possible to (better) address the user’s research question by applying this functionality to the research data. The *infrastructure specialist* (IS) is a specialist who has a deep understanding of the CLARIN service-oriented architecture and its requirements, and/or a specialist in data and tool format standards supported in CLARIN. IS specialists usually get involved in the project via the CLARIN-NL Helpdesk. In many cases the different roles of user, DP, TP and IS will be played by persons from different organizations, but they may originate from a single organization, and occasionally even be played by a single person. CLARIN-NL can offer assistance in bringing the right experts together, if desired.

4.1 User

The user is a researcher from a linguistics or humanities institute who aims to investigate one or more specific research questions. The project proposal should clearly describe the research question(s) of the user, and the research question(s) must be in the domain of research in the humanities in general and the study of language in particular. The project must be led by the user.

4.2 Data Provider

The DP has a certain set of digital language-related data at his/her disposal (the 'research data') that can be used for addressing the research questions of the user. The project proposal should clearly describe the research data the DP has at his/her disposal that can be used to address the research question(s), and how they can be used for this purpose. The research data must be existing digital language or language-related data. No new research data should be created in the project. The DP must have the right to make the research data available on a CLARIN server running at a dedicated centre. If the data are in a format that is not currently on the list of CLARIN standards, a resource curation project is in order. Otherwise, the data can be used in a demonstrator project. The project proposal should contain a detailed description of the research data, its current state and format, the plans to convert it if needed, justification for using different formats if applicable, and a detailed plan for dealing with the data and its metadata (see below). Any restrictions on the use of the data as well as any ethical issues that apply or may arise must be properly documented in the proposal.

4.3 The Technology Provider

The TP has a certain technology at his/her disposal that can be used as a basis for the development of a web-based application (possibly web-services based) and concomitant demonstrator, or that can be used for resource curation. Since the research data are language data, the technology will in most cases be language or speech technology.

The proposal should contain a detailed description of the available technology and its current status. It should make clear that the TP has a thorough understanding of this technology and describe how the TP obtained this understanding (e.g. because the TP developed the technology).

The intended use of the technology in the project should be described, as well as any extensions or modifications that have to be made to the technology in the project, and a plan to achieve this.

The TP must have the right to use this technology and indicate how it will be used in the project.

4.4 CLARIN-NL Helpdesk and Infrastructure Specialist

CLARIN-NL has set up a Helpdesk that project participants can turn to for all kinds of technical questions related to their project. If the relevant questions cannot be addressed by the Helpdesk functionality (FAQ section, etc.) or staff, the Helpdesk will involve infrastructure specialists. The infrastructure specialist (IS) is a specialist who has a deep understanding of the CLARIN service-oriented architecture and its requirements, and/or a specialist in data, metadata and tool format standards and best practices supported in CLARIN. The Helpdesk, and where needed IS specialists, will advise and assist the project partners.

5 Project Types

5.1 Resource Curation

Resource curation involves a number of different aspects:

1. The resource should be brought into a format that adheres to widely accepted standards and best practices currently considered as likely candidates by CLARIN.

2. Proper metadata descriptions need to be created and made available. They must be compliant with the CLARIN component metadata infrastructure (CMDI) and it should be possible to harvest and access them.
3. Metadata descriptions should include persistent identifiers that can be resolved and the CLARIN requirements should hold for the PID system.
4. The linguistic encoding must be related to the data category registry, i.e. data categories used must be mapped to corresponding ISOCAT data categories where they exist in a formal way (e.g. via an XML Schema) and new data categories must be added to ISOCAT if they do not exist there yet.
5. Provide proper documentation of the resource, at least in English.

The plan for a curation project should describe in detail how these different aspects are going to be addressed in the project.

The results of these aspects should be tested by the project participants. Setting up tests for this should be included in the project plan and the results of these tests will be included in the project's success criteria. Example tests are e.g. a metadata harvesting test and formal procedures such as testing against an XML Schema.

The resulting resource and its metadata must be made available on a server of a recognized CLARIN centre. The project proposal must specify which (candidate) CLARIN centre this will be and concrete arrangements must have been made with this centre.

5.2 Demonstrator

In a demonstrator project a demonstrator is developed using a documented web-based application based on a technology that the TP currently has at his/her disposal. The development is carried out in close cooperation with the user

The project proposal must contain a detailed description of the targeted functionality, including input and output specifications, and how it can contribute to addressing the user's research questions. More generic functionality, i.e. functionality that can serve multiple different research questions from linguistics and humanities research will be preferred over less generic or completely idiosyncratic functionality. See below for additional criteria related to the functionality that will be used to rank proposals. Since a demonstrator project is short in duration and is relatively small, this functionality must already be available to the TP, though perhaps not in the form of a web-based application, and it perhaps only operates on data formats other than the ones listed in the CLARIN standards. The project proposal should contain a detailed description of the functionality in its current state, the targeted web-based application and its components, and a plan to achieve this. The application includes a web-based user interface that takes care of user interactions and method invocations to the core component. An Application Programming Interface (API) to the core component must be provided and documented. The TP must have the rights to make the targeted core component as well as the web-application available on a CLARIN server running at a dedicated centre.

The core component of the web-application must at least be able to operate on the research data and yield output in the formats agreed upon between user and TP. It is a pre if it can apply to other formats from the CLARIN standard list and yield additional output formats. The web-application and

its core component will be used to obtain requirements and specifications of the architectural framework that is being worked out in CLARIN and may be used to test it.

A research data resource often consists of information of various kinds contained in multiple folders and multiple files of varying types. The information contained in such a resource can include documentation, source data, annotations of the source data, aggregate statistics tables on the source data and/or annotations, etc. Any tool (which eventually will very likely be integrated in the infrastructure as web service) should find out in a fully automated manner whether the research data selected by the user are appropriate input for it, and, if so, that it is applied to the right information (e.g. to the source data but not to the documentation). To achieve this it will integrate wrappers that read and write metadata and provenance information provided by the CLARIN infrastructure. A popular wrapper, used a lot and supported in CLARIN-NL is CLAM, developed by Tilburg University. Any requirements or desiderata that follow from this for metadata and data contents and formats should be properly documented in the documents with the requirements and desiderata for the CLARIN infrastructure.

The demonstrator consists, as a minimum, of a web application, the research data, and a demonstration scenario. A demonstration scenario is a detailed description of example (sequences of) actions a user can take to have the application applied to the research data and the corresponding system responses in order to get a representative picture of the functionality offered. A movie or sequence of screen captures to illustrate the functionality is nice to have. The application will have to be installed on a CLARIN server, and the project proposal must contain a plan for doing this. It is the intention to have the demonstrator applications available for the lifetime of the CLARIN-NL project (2009-2014), so occasional support may be needed from the original developers even after the demonstrator project has finished.

The application must be tested with at least one of the common web browsers on the client side (MS IE, Firefox). Agreements about additional technical details (operating system, programming language, workspace requirements, etc) need to be made with the dedicated centre where the services should be executed.

Any vendor, platform or operating system dependent aspects of the application must be made explicit in the proposal and properly documented in the project.

The web-application and its core component should be properly documented, for users (user documentation), for application developers who want to use the core component (documentation of the API), and for technology developers who want to modify or extend the basic functionality of the application. The documentation, as well as the software user interface, must be provided at least in English.

Auxiliary Resources

The web-application may require data and other software (auxiliary resources) while running.

It must be documented which auxiliary data (e.g. a lexicon) and software (e.g. a library, converters) are needed during runtime for the application. The TP must have the right to make these auxiliary data and software available on a CLARIN server. Any restrictions on their usage (including costs)

should be properly documented in the project proposal and in the documentation of the resulting application.

The application and the core component must be able to run on a dedicated CLARIN-server. An application or core component that can run only on a specific (non CLARIN) server (e.g. because it contains auxiliary resources that cannot be made available otherwise) is not acceptable.

6 Metadata

For the web application and its core component, the research data and all runtime auxiliary data used in the application, metadata descriptions must be made in accordance with the CLARIN metadata standard (CMDI). CMDI provides a flexible component-based framework for dealing with metadata; the data and tools of the projects of this call may require the development of new CMDI components or the adaptation of existing components and thus can contribute to the further development of the CMDI framework. Any required or desirable extension or modifications of the CMDI framework must be properly documented and be included in the CLARIN Requirements and Desiderata document.

7 Requirements and Desiderata for CLARIN infrastructure

One important result of both demonstrator and curation projects is a document or series of documents describing requirements and desiderata for the CLARIN infrastructure resulting from the experiences gained with the curation of the research data and/or tools, and with the development of the application, its core component and web-services derived from it. These requirements and desiderata can concern many aspects. The following is a non-exhaustive list of aspects that should be considered:

- Requirements for data formats and encoding standards
- Web-service wrappers
- Metadata elements and formats
- Processing requirements
- Memory requirements
- Network Bandwidth requirements
- User workspace requirements
- API requirements (e.g. Calling conventions)
- IPR / restricted use / ethical issues requirements
- Documentation requirements
- Repository Requirements
- Requirements for registering and resolving PIDs
- Requirements related to semantic interoperability

8 Evaluation Criteria

A proposal submitted in the Closed Call will not compete with other proposals. Each proposal is evaluated in accordance with the criteria described here, and if this evaluation is positive, the project will be funded.

A proposal must describe a project that is compatible with the requirements mentioned in this call; in particular it must be a resource curation and/or a demonstrator project.

Proposals for projects will furthermore be evaluated according to the following more general criteria:

- Quality
    - Clarity and added value of the project proposal, in particular of the problem and the proposed approach
    - Suitability of the method and plan for the problem at hand
    - Feasibility of the project targets: can they be realized within the specified amount of time and with the instruments proposed?
    - Adequate balance between requested instruments and funds and proposed targets
    - Clearly specified and realistic work plan
    - Conformance to established standards and protocols as supported within CLARIN, or contribute to the development such standards and protocols.
- Project Participants
    - Competence of the participating partners (including their past performance);
    - Balanced cooperation and task assignments within the project. Justification of the composition of the team.
    - Availability of the infrastructure required for the project to be successful
    - Embedding of the work in other research programmes or projects, and/or additional funding from other funding sources is an advantage
- User-orientation of the project
    - Does the project address needs of the targeted infrastructure users (linguists and humanities researchers)?
    - Projects on tools or data that are widely in use in the targeted user community will score higher than projects focusing on lesser used tools and data.
    - Is there cooperation with or support from the targeted (future) infrastructure users?
    - Is the resulting tool / service user-friendly, i.e. will non-technical linguistic and humanities researchers be able to use it?
    - Is dissemination of the results to the targeted users and (where appropriate) training of them planned?
- Contribution to CLARIN-NL as a whole
    - Conformance to the goals of CLARIN-NL in particular and CLARIN in general and the priorities set within them
    - Contribution to knowledge transfer and network creation. In particular, cooperation between the intended users (linguists and humanities researchers) and technology and service providers (researchers in language and speech technology, computer science, etc.) is an advantage.
- Intellectual Property Rights and Synergy
    - Each proposal must contain clear statements about the situation of the IPR of the data and tools/technologies used, and a detailed plan to resolve any open issues.
    - The project participants have the obligation and must therefore have the rights to incorporate the core data and tools used in a project into the CLARIN infrastructure (this is a sine qua non). There has to be a clear specification and justification of the

use of any data or tools needed in the project that cannot be incorporated into the CLARIN infrastructure.
    - Each proposal must show that the submitters have adequate and up-to-date knowledge of data, tools and services that are already available, so that any duplication of effort can be avoided.
- Formal compliance
    - A proposal must meet the formal requirements imposed by the CLARIN-NL organization for proposals, such as
        - conformance to the prescribed format and proposal template
        - submission before the set deadline, using the means prescribed
        - conformance to the prescribed language of the proposal

In addition, more generic applications and data, i.e. applications and data that can serve multiple different research questions will be preferred over less generic applications and data. Projects that are part of or fit in with international cooperation with partners from CLARIN in other countries will be preferred (of course, foreign organizations will have to find their own funding). Finally, a project that does not meet the IPR-requirements stated or is insufficiently clear about it will be considered formally noncompliant.

9 Duration

The duration of the project must be justified. The default maximum duration is 12 months. Any duration longer than 12 months requires thorough justification.

10 Budget

The project budget must be in accordance with the tasks to be carried out, and this must be justified in the project proposal. The maximum budget is 80k€.

11 Intellectual Property Rights (IPR)

Ownership of all original data and software remains with the original owners.

An agreement must be in place between the owners of the original data and software and the project participants on the IPR of the adapted data and software before the submission date of a proposal if the owners of the original data and software are not identical to the project participants. If applicable, a copy of this agreement must be uploaded together with the project proposal. Otherwise ownership of the created adaptations and extensions will be with the creator(s).

The project participants have the obligation and therefore must have the rights to make the research data, the application, its core component, and any runtime auxiliary data or software available on a CLARIN server for use by researchers having access to the CLARIN infrastructure. This is a sine qua non. Any proposal not satisfying this requirement or being insufficiently clear about this matter will be considered to be formally noncompliant and will be rejected on these grounds.

The project proposal should describe all issues related to IPR and present solutions for them. The relations between the partners in a project must be agreed upon in a consortium agreement before the start of the project.

Practical details

The CLARIN-NL Third Call Closed Call is open from Friday July 3, 2011. The total budget for this call is limited to a maximum of € 400,000. This call is specifically open for proposals targeting resource curation projects or demonstrator projects. Only those proposals that specifically target this priority will be eligible.

Full proposals must be submitted in English and in PDF format to the CLARIN-NL electronic proposal submission system using the prescribed template (which can be found on the CLARIN-NL website). The deadline for submitting full proposals in this call has been set for Friday September 30, 2011 13:00 hours CET.

The CLARIN-NL electronic proposal submission system can be accessed as of July 3, 2011 via the CLARIN-NL website. If the electronic proposal submission system would not work, contact the CLARIN-NL Office clarinnl@uu.nl.

Who can apply?

Applications can be submitted only by researchers who have been explicitly invited to submit a project proposal for the closed call by the CLARIN-NL Board.

Eligible costs

- Personnel costs directly related to the project, in accordance with the Akkoord NWO-VSNU 2008 (and any additions to it).
- A fee of maximally 3.000 € per FTE per year (or a pro rata part for less than 1 FTE per year) for covering travel and subsistence costs
- The requested funding cannot exceed 80,000€

Evaluation procedure full proposals

All eligible full proposals submitted in this call will be presented to a panel of international experts in the humanities, language and speech technology and infrastructures (International Advisory Panel, IAP). The composition of the IAP can be found on the CLARIN-NL website. If the proposals require this, the CLARIN-NL Executive Board can decide to involve additional experts in the evaluation.

This international panel will assess each eligible application based on the assessment criteria relevant to this call and may formulate a set of recommendations for improving the individual proposals. The panel's assessment and recommendations will be presented to the National Advisory Panel (NAP). The members of the NAP who are not directly involved in the submitted proposals will also assess each application and the IAP's assessment and recommendations. On the basis of the IAP's advice and the NAP's advice, the CLARIN-NL Board will finally determine whether the project will be funded.

Projects should start within three months after the applicant has received the formal notification of funding.

CLARIN-NL consortium agreement

More information as to which legal rules apply for this specific CLARIN-NL granting scheme are laid down in the CLARIN-NL consortium agreement, which can be found on the CLARIN-NL website (http://www.clarin.nl/node/72).

Timetable

Activity	Date
CLARIN-NL Third Call Open	Sunday July 3, 2011
CLARIN-NL Third Call Information Session	Thursday August 25, 2011 (afternoon)
Deadline Proposal Submission	Friday September 30, 2011 13:00hrs CET

The exact deadlines of the assessments by the IAP and NAP, and the final decision by the board will be communicated later via the CLARIN-NL website. The target is to make the final decisions before the end of 2011.

CLARIN-NL Organization

The CLARIN-NL project is funded by NWO.

The CLARIN-NL project is coordinated by the Programme Director, prof.dr. J.E.J.M. Odijk, who is a member of the CLARIN-NL Executive Board. Drs. Jolien Scholten is the CLARIN-NL project secretary.

The International Advisory Panel (IAP) is a group of international experts in the areas of humanities, in particular linguistics, language and speech technology, and infrastructures for scientific research.

The National Advisory Panel is a group of national researchers representative for the fields of linguistics and humanities, language and speech technologies and infrastructures for scientific research.

The CLARIN-NL Board consists of national senior researchers with great expertise in governance and/or relevant technical expertise

The composition of these CLARIN-NL governance bodies can be found on the CLARIN-NL website.

CLARIN-NL Helpdesk

Contact the CLARIN-NL Helpdesk for any technical questions related to this call, e.g.

- Questions on metadata and CMDI
- Questions on data categories en ISOCAT
- Questions on standards supported in CLARIN
- Questions on web services and CLAM.
- Etc.

e-mail: helpdesk@clarin.nl

CLARIN-NL Centres

A list of the current candidate CLARIN centres and their contact persons can be found on the CLARIN-NL website: http://www.clarin.nl/node/130

CLARIN-NL Office

Contact the CLARIN-NL office for any organizational or practical questions related to this call, e.g.

- Additional clarification
- Advice on eligibility of your plans
- Assistance with finding experts, data or technology required
- Assistance with finding project partners
- Assistance with selecting a CLARIN Centre.
- Etc.

Jan Odijk/Jolien Scholten
Tel: +31 30 253 6279
Fax: +31 30 253 6000
Address: Trans 10, 3512 JK Utrecht
E-mail: clarinnl@uu.nl
CLARIN-NL Website: http://www.clarin.nl

Call 4 (2012)
Open Call

CLARIN-NL launches the fourth open call for project proposals. This call is open from Monday July 2, 2012. The total budget for this call is limited to a maximum of € 600,000. This call is specifically open for proposals targeting resource curation projects or demonstrator projects. Such projects take existing data or tools as a basis and attempt to apply CLARIN-supported standards and best practices to make the data and tools CLARIN-compliant. Examples of such projects awarded in 2009 through 2011 can be found on the CLARIN-NL website. Only those proposals that specifically target this priority will be eligible. The maximum budget per project is 120,000€.

Important Note It is an essential requirement that a CLARIN centre is involved in each resource curation or demonstrator project. It is therefore recommended to contact a CLARIN-centre as early as possible while preparing a project proposal. The list of CLARIN centres, the kinds of resources they focus on, and their contact persons can be found on the CLARIN web site, more specifically here. Some aspects of this call text are technical in nature: the CLARIN centre will, inter alia, be able to explain and discuss these technical matters with you. At the end of this document, one will find a list of acronyms and terms for technical notions with some explanation and references.

1 Introduction

The first phase of the CLARIN-NL project focuses on the specification and design of the infrastructure. Obviously, the CLARIN infrastructure should be designed in such a way that it can incorporate the data and tools currently used by humanities researchers to address their research questions. On the other hand, the CLARIN infrastructure can only be successful if these data and tools meet certain requirements with regard to standards and interoperability, not only with regard to the data and tools themselves but also with regard to their visibility and accessibility. Many of the data and tools currently in use do not meet these requirements, or meet them only partially. One of the aims of the fourth call is to make existing data and tools compliant with the requirements on standards and interoperability currently advocated in CLARIN.

However, the list of standards, best practices and interoperability requirements currently advocated by CLARIN has not been completely fixed yet. There are good reasons for this: (1) there may be crucial data or tools for which none of the currently advocated standards or best practices is suited; (2) the currently advocated standards and best practices may be incomplete, insufficiently specific, inconvenient or even incompatible with crucial data and tools. Therefore, a second aim of the fourth call is to test whether and to which extent the currently existing data and tools can be made compatible with the standards and best practices currently advocated in CLARIN, and to get a good overview of any incompatibilities as well as suggestions for adaptations of these standards.

These matters can best be investigated by actually attempting the necessary conversions and adaptations. When applied to tools and other software this will result in CLARIN-compliant web applications that can also serve as showcases of the kind of functionality CLARIN aims to offer. A third aim of the fourth call is therefore to make sure that the applications developed become available in demonstrators which can help promote the particular applications but also the CLARIN infrastructure as a whole.

Finally, by investigating these matters using specific data and tools that are currently in use a lot can be learnt about requirements the CLARIN infrastructure should meet or desiderata that it should offer. The fourth aim of this call is therefore obtaining a detailed list of such requirements and desiderata.

2 Project Types

In the fourth call, CLARIN-NL therefore solicits (1) projects that carry out resource (data or tools) curation, and (2) demonstrator projects.

Curation projects

CLARIN-NL aims to support the curation of digital language resources (data and tools) so that these resources can participate in the CLARIN infrastructure, more easily be accessed by interested researchers via online methods and become part of appealing new applications. Many language resources are neither visible nor accessible. Visibility is mainly achieved by standardized metadata that are being harvested by service providers. Accessibility has many different aspects:

1. The resource needs to be stored at computers that are accessible via the internet.
2. The resource needs to be identified in a persistent manner.
3. The resource needs to be interpretable, which requires a format that adheres to best practices and it requires references to registries were the used concepts are defined.

Demonstrator projects

CLARIN-NL aims to support projects that create appealing showcases of functionality that the CLARIN infrastructure should support. Such projects should make available web applications that can be used as demonstrators of functionality that supports addressing research questions of the CLARIN-NL intended user group. The development of these demonstrators will also be used to inventory a list of requirements the CLARIN infrastructure should meet and desiderata it preferably should offer.

3 Goals

The goal of a curation project is:

- Adapting specific resources so that they are visible, uniquely referable and accessible via the web, and properly documented.

The goal of a demonstrator project is:

- Creating a documented web application starting from an existing tool or application that can be used as a demonstrator and function as a showcase of the type of functionality CLARIN will incorporate and support. Within the web application there must be a clear separation between the web-based user interface and the core component. For the latter an API must be defined and documented.

Important goals common to both types of projects are:

- Applying standards and best practices and making use of the suggested CLARIN architecture and agreements to understand their limitations and the requirements for extensions.
- Establishing requirements and desiderata for the CLARIN infrastructure

In both project types the use of CLARIN-supported standards and best practices is essential.

The selection of CLARIN-supported standards and best practices is currently ongoing and it is the intention that the projects supported in this call contribute to this. A preliminary set of candidates for CLARIN-supported standards and best practices is available, and though it is by no means final we will refer to this list in this document by the term “CLARIN-standards”.

4 Roles

Four roles of persons involved in the projects can be distinguished: the *user*, the *data provider* (DP), the *technology provider* (TP), and the *infrastructure specialist* (IS). The *user* is a researcher from a linguistics or humanities institute who aims to investigate one or more specific research questions. The *DP* has a certain set of digital language-related data at his/her disposal (the ‘research data’) that can be used for addressing the research questions of the user. The *TP* has a certain technology (e.g. language or speech technology) at its disposal and a thorough understanding of this technology (e.g. because it was developed by the TP). This technology offers functionality that makes it possible to (better) address the user’s research question by applying this functionality to the research data. The *infrastructure specialist* (IS) is a specialist who has a deep understanding of the CLARIN service-oriented architecture and its requirements, and/or a specialist in data and tool format standards supported in CLARIN. IS specialists usually get involved in the project via the CLARIN-NL Helpdesk. In many cases the different roles of user, DP, TP and IS will be played by persons from different organizations, but they may originate from a single organization, and occasionally even be played by a single person. CLARIN-NL can offer assistance in bringing the right experts together, if desired.

4.1 User

The user is a researcher from a linguistics or humanities institute who aims to investigate one or more specific research questions. The project proposal should clearly describe the research question(s) of the user, and the research question(s) must be in the domain of research in the humanities in general and the study of language in particular. The project must be led by the user.

4.2 Data Provider

The DP has a certain set of digital language-related data at his/her disposal (the 'research data') that can be used for addressing the research questions of the user. The project proposal should clearly describe the research data the DP has at his/her disposal that can be used to address the research question(s), and how they can be used for this purpose. The research data must be existing digital language or language-related data. No new research data should be created in the project. The DP must have the right to make the research data available on a CLARIN server running at a dedicated centre. If the data are in a format that is not currently on the list of CLARIN standards, a resource curation project is in order. Otherwise, the data can be used in a demonstrator project. The project proposal should contain a detailed description of the research data, its current state and format, the plans to convert it if needed, justification for using different formats if applicable, and a detailed plan for dealing with the data and its metadata (see below). Any restrictions on the use of the data as well as any ethical issues that apply or may arise must be properly documented in the proposal.

4.3 The Technology Provider

The TP has a certain technology at his/her disposal that can be used as a basis for the development of a web-based application (possibly web-services based) and concomitant demonstrator, or that can be used for resource curation. Since the research data are language data, the technology will in most cases be language or speech technology.

The proposal should contain a detailed description of the available technology and its current status. It should make clear that the TP has a thorough understanding of this technology and describe how the TP obtained this understanding (e.g. because the TP developed the technology).

The intended use of the technology in the project should be described, as well as any extensions or modifications that have to be made to the technology in the project, and a plan to achieve this.

The TP must have the right to use this technology and indicate how it will be used in the project.

4.4 CLARIN-NL Helpdesk and Infrastructure Specialist

CLARIN-NL has set up a Helpdesk that project participants can turn to for all kinds of technical questions related to their project. If the relevant questions cannot be addressed by the Helpdesk functionality (FAQ section, etc.) or staff, the Helpdesk will involve infrastructure specialists. The infrastructure specialist (IS) is a specialist who has a deep understanding of the CLARIN service-oriented architecture and its requirements, and/or a specialist in data, metadata and tool format standards and best practices supported in CLARIN. The Helpdesk, and where needed IS specialists, will advise and assist the project partners.

5 Project Types

5.1 Resource Curation

Resource curation involves a number of different aspects:

1. The resource should be brought into a format that adheres to widely accepted standards and best practices currently considered as likely candidates by CLARIN.

2. Proper metadata descriptions need to be created and made available. They must be compliant with the CLARIN component metadata infrastructure (CMDI) and it should be possible to harvest and access them.
3. Metadata descriptions should include persistent identifiers that can be resolved and the CLARIN requirements should hold for the PID system.
4. All data categories used in the metadata and in the actual data (e.g. linguistic annotations) must be related to a CLARIN-recognized data category registry (currently only ISOCAT), i.e. data categories used must be mapped to corresponding ISOCAT data categories where they exist in a formal way (e.g. via an XML Schema) and new data categories must be added to ISOCAT if they do not exist there yet.
5. Provide proper documentation of the resource, at least in English.

The plan for a curation project should describe in detail how these different aspects are going to be addressed in the project.

The results of these aspects should be tested by the project participants. Setting up tests for this should be included in the project plan and the results of these tests will be included in the project's success criteria. Example tests are e.g. a metadata harvesting test and formal procedures such as testing against an XML Schema.

The resulting resource and its metadata must be made available on a server of a recognized CLARIN centre. The project proposal must specify which (candidate) CLARIN centre this will be and concrete arrangements must have been made with this centre.

5.2 Demonstrator

In a demonstrator project a demonstrator is developed using a documented web-based application based on a technology that the TP currently has at his/her disposal. The development is carried out in close cooperation with the user

The project proposal must contain a detailed description of the targeted functionality, including input and output specifications, and how it can contribute to addressing the user's research questions. More generic functionality, i.e. functionality that can serve multiple different research questions from linguistics and humanities research will be preferred over less generic or completely idiosyncratic functionality. See below for additional criteria related to the functionality that will be used to rank proposals. Since a demonstrator project is short in duration and is relatively small, this functionality must already be available to the TP, though perhaps not in the form of a web-based application, and it perhaps only operates on data formats other than the ones listed in the CLARIN standards. The project proposal should contain a detailed description of the functionality in its current state, the targeted web-based application and its components, and a plan to achieve this. The application includes a web-based user interface that takes care of user interactions and method invocations to the core component. An Application Programming Interface (API) to the core component must be provided and documented. The TP must have the rights to make the targeted core component as well as the web-application available on a CLARIN server running at a dedicated centre.

The core component of the web-application must at least be able to operate on the research data and yield output in the formats agreed upon between user and TP. It is a pre if it can apply to other

formats from the CLARIN standard list and yield additional output formats. The web-application and its core component will be used to obtain requirements and specifications of the architectural framework that is being worked out in CLARIN and may be used to test it.

A research data resource often consists of information of various kinds contained in multiple folders and multiple files of varying types. The information contained in such a resource can include documentation, source data, annotations of the source data, aggregate statistics tables on the source data and/or annotations, etc. Any tool (which eventually will very likely be integrated in the infrastructure as web service) should find out in a fully automated manner whether the research data selected by the user are appropriate input for it, and, if so, that it is applied to the right information (e.g. to the source data but not to the documentation). To achieve this it will integrate wrappers that read and write metadata and provenance information provided by the CLARIN infrastructure. A popular wrapper, used a lot and supported in CLARIN-NL is CLAM, developed by Tilburg University and currently maintained by Radboud University Nijmegen. Any requirements or desiderata that follow from this for metadata and data contents and formats should be properly documented in the documents with the requirements and desiderata for the CLARIN infrastructure.

The demonstrator consists, as a minimum, of a web application, the research data, and a demonstration scenario. A demonstration scenario is a detailed description of (sequences of) actions a user can take to have the application applied to the research data and the corresponding system responses in order to get a representative picture of the functionality offered. A movie or sequence of screen captures to illustrate the functionality is nice to have. The application will have to be installed on a CLARIN server, and the project proposal must contain a plan for doing this. It is the intention to have the demonstrator applications available for the lifetime of the CLARIN-NL project (2009-2015), so occasional support may be needed from the original developers even after the demonstrator project has finished.

The application must be tested with at least one of the common web browsers on the client side (MS IE, Firefox). Agreements about additional technical details (operating system, programming language, workspace requirements, etc) need to be made with the dedicated centre where the services should be executed.

Any vendor, platform or operating system dependent aspects of the application must be made explicit in the proposal and properly documented in the project.

The web-application and its core component should be properly documented, for users (user documentation), for application developers who want to use the core component (documentation of the API), and for technology developers who want to modify or extend the basic functionality of the application. The documentation, as well as the software user interface, must be provided at least in English.

Auxiliary Resources

The web-application may require data and other software (auxiliary resources) while running.

It must be documented which auxiliary data (e.g. a lexicon) and software (e.g. a library, converters) are needed during runtime for the application. The TP must have the right to make these auxiliary data and software available on a CLARIN server. Any restrictions on their usage (including costs)

should be properly documented in the project proposal and in the documentation of the resulting application.

The application and the core component must be able to run on a dedicated CLARIN-server. An application or core component that can run only on a specific (non CLARIN) server (e.g. because it contains auxiliary resources that cannot be made available otherwise) is not acceptable.

6 Metadata

For the web application and its core component, the research data and all runtime auxiliary data used in the application, metadata descriptions must be made in accordance with the CLARIN metadata standard (CMDI). CMDI provides a flexible component-based framework for dealing with metadata; the data and tools of the projects of this call may require the development of new CMDI components or the adaptation of existing components and thus can contribute to the further development of the CMDI framework. Any required or desirable extension or modifications of the CMDI framework must be properly documented and be included in the CLARIN Requirements and Desiderata document.

7 Requirements and Desiderata for CLARIN infrastructure

One important result of both demonstrator and curation projects is a document or series of documents describing requirements and desiderata for the CLARIN infrastructure resulting from the experiences gained with the curation of the research data and/or tools, and with the development of the application, its core component and web-services derived from it. These requirements and desiderata can concern many aspects. The following is a non-exhaustive list of aspects that should be considered:

- Requirements for data formats and encoding standards
- Web-service wrappers
- Metadata elements, components and profiles
- Processing requirements
- Memory requirements
- Network Bandwidth requirements
- User workspace requirements
- API requirements (e.g. Calling conventions)
- IPR / restricted use / ethical issues requirements
- Documentation requirements
- Repository Requirements
- Requirements for registering and resolving PIDs
- Requirements related to semantic interoperability

8 Evaluation Criteria

Proposals are elicited from humanities topics and disciplines, with a priority for the following (sub)disciplines:

1. Philosophy

2. Literary sciences
3. Cultural sciences
4. Communication & Media Studies, including Theatre sciences
5. Linguistics
6. History and political sciences

The order reflects the level of priority: the higher on the list, the higher the priority.

Projects involving multiple priority disciplines are especially encouraged.

Proposals on topics other that the listed priorities will be considered but must show their special relevance for the CLARIN infrastructure.

A proposal must describe a project that is compatible with the requirements mentioned in this call; in particular it must be a resource curation and/or a demonstrator project.

Proposals for projects will furthermore be evaluated and ranked according to the following more general criteria:

- Quality
  - Clarity and added value of the project proposal, in particular of the problem and the proposed approach
  - Suitability of the method and plan for the problem at hand
  - Feasibility of the project targets: can they be realized within the specified amount of time and with the instruments proposed?
  - Adequate balance between requested instruments and funds and proposed targets
  - Clearly specified and realistic work plan
  - Conformance to established standards and protocols as supported within CLARIN, or contribute to the development such standards and protocols.
- Project Participants
  - Competence of the participating partners (including their past performance);
  - Balanced cooperation and task assignments within the project. Justification of the composition of the team.
  - Availability of the infrastructure required for the project to be successful
  - Embedding of the work in other research programmes or projects, and/or additional funding from other funding sources is an advantage
- User-orientation of the project
  - Does the project address needs of the targeted infrastructure users (linguists and humanities researchers)?
  - Projects on tools or data that are widely in use in the targeted user community will score higher than projects focusing on lesser used tools and data.
  - More generic data or functionality, i.e. data or functionality that can serve multiple different research questions from linguistics and humanities research will be preferred over less generic or completely idiosyncratic functionality
  - Is there cooperation with or support from the targeted (future) infrastructure users?
  - Is the resulting tool / service user-friendly, i.e. will non-technical linguistic and humanities researchers be able to use it?

- Is dissemination of the results to the targeted users and (where appropriate) training of them planned?

- Contribution to CLARIN-NL as a whole
  - Conformance to the goals of CLARIN-NL in particular and CLARIN in general and the priorities set within them
  - Contribution to knowledge transfer and network creation. In particular, cooperation between the intended users (linguists and humanities researchers) and technology and service providers (researchers in language and speech technology, computer science, etc.) is an advantage.
- Intellectual Property Rights and Synergy
  - Each proposal must contain clear statements about the situation of the IPR of the data and tools/technologies used, and a detailed plan to resolve any open issues.
  - The project participants have the obligation and must therefore have the rights to incorporate the core data and tools used in a project into the CLARIN infrastructure (this is a sine qua non). There has to be a clear specification and justification of the use of any data or tools needed in the project that cannot be incorporated into the CLARIN infrastructure.
  - Each proposal must show that the submitters have adequate and up-to-date knowledge of data, tools and services that are already available, so that any duplication of effort can be avoided.
- Formal compliance
  - A proposal must meet the formal requirements imposed by the CLARIN-NL organization for proposals, such as
    - conformance to the prescribed format and proposal template
    - submission before the set deadline, using the means prescribed
    - conformance to the prescribed language of the proposal

In addition, projects that are part of or fit in with international cooperation with partners from CLARIN in other countries will be preferred (of course, foreign organizations will have to find their own funding). Finally, a project that does not meet the IPR-requirements stated or is insufficiently clear about it will be considered formally noncompliant.

9 Duration

The duration of the project must be justified. The default maximum duration is 12 months. Any duration longer than 12 months requires thorough justification.

10 Budget

The project budget must be in accordance with the tasks to be carried out, and this must be justified in the project proposal. The maximum budget is 120k€.

11 Intellectual Property Rights (IPR)

Ownership of all original data and software remains with the original owners.

An agreement must be in place between the owners of the original data and software and the project participants on the IPR of the adapted data and software before the submission date of a proposal if the owners of the original data and software are not identical to the project participants. If applicable, a copy of this agreement must be uploaded together with the project proposal. Otherwise ownership of the created adaptations and extensions will be with the creator(s).

The project participants have the obligation and therefore must have the rights to make the research data, the application, its core component, and any runtime auxiliary data or software available on a CLARIN server for use by researchers having access to the CLARIN infrastructure. This is a sine qua non. Any proposal not satisfying this requirement or being insufficiently clear about this matter will be considered to be formally noncompliant and will be rejected on these grounds.

The project proposal should describe all issues related to IPR and present solutions for them. The relations between the partners in a project must be agreed upon in a consortium agreement before the start of the project.

Practical details

The CLARIN-NL Fourth Open Call is open from Monday July 2, 2012. The total budget for this call is limited to a maximum of € 600,000. This call is specifically open for proposals targeting resource curation projects or demonstrator projects. Only those proposals that specifically target this priority will be eligible.

Full proposals must be submitted in English and in PDF format to the CLARIN-NL electronic proposal submission system using the prescribed template (which can be found on the CLARIN-NL website). The deadline for submitting full proposals in this call has been set for Wednesday September 26, 2012 13:00 hours CET.

The CLARIN-NL electronic proposal submission system can be accessed as of Monday July 2, 2012 via the CLARIN-NL website. If the electronic proposal submission system would not work, contact the CLARIN-NL Office clarinnl@uu.nl.

Who can apply?

Applications can be submitted only by researchers affiliated to CLARIN-NL participants that have signed the CLARIN-NL consortium agreement. The list of CLARIN-NL participants is available on the CLARIN-NL website. The main applicant and coordinator of the project must be a researcher from linguistics or humanities more broadly.

CLARIN-NL is in principle open to new participants. If your institute is not currently a CLARIN-NL participant, contact the CLARIN-NL office if you want your organization to become a participant in CLARIN-NL well in advance of the submission deadline.

Applicants who are planning to submit a proposal are strongly advised to contact the CLARIN-NL Office for an eligibility check of their plans, or for assistance in finding suitable partners.

Eligible costs

- Personnel costs directly related to the project, in accordance with the Akkoord NWO-VSNU 2008 (and any additions to it).

- A fee of maximally 3.000 € per FTE per year (or a pro rata part for less than 1 FTE per year) for covering travel and subsistence costs
- The requested funding cannot exceed 120,000€

Evaluation procedure full proposals
All eligible full proposals submitted in this call will be presented to a panel of international experts in the humanities, language and speech technology and infrastructures (International Advisory Panel, IAP). The composition of the IAP can be found on the CLARIN-NL website. If the proposals require this, the CLARIN-NL Executive Board can decide to involve additional experts in the evaluation.

This international panel will assess and rank the eligible applications based on the assessment criteria relevant to this call and may formulate a set of recommendations for improving the individual proposals. The IAP's assessment and ranking will not in this stage be sent to the proposers, but the IAP will, if needed, formulate questions and remarks to the project proposers. The project proposers will get the opportunity to answer these questions and comment on the remarks. The response must have been received by the CLARIN-NL office before the deadline set for it (one will typically have about a week to make the response). The IAP's assessment and recommendations will be presented, together with the response of the project proposers to the National Advisory Panel (NAP). The members of the NAP who are not directly involved in the submitted proposals will also assess the applications, in particular with regard to national considerations and based on the project proposers' feedback, and it make its own recommendations. On the basis of the IAP's advice and the NAP's advice, the CLARIN-NL Board will finally determine which projects will be funded. The coordinator of each project will receive a message on the final decision together with the NAP's and IAP's assessment reports on the project.

Projects should start within three months after the applicant has received the formal notification of funding.

CLARIN-NL consortium agreement
More information as to which legal rules apply for this specific CLARIN-NL granting scheme are laid down in the CLARIN-NL consortium agreement, which can be found on the CLARIN-NL website (http://www.clarin.nl/node/72).

Timetable

Activity	**Date**
CLARIN-NL Fourth Call Open	Monday July 2, 2012
CLARIN-NL Fourth Call Information Session	Thursday August 30, 2012 (afternoon)
Deadline Proposal Submission	Wednesday September 26, 2012 13:00hrs CET
Feedback / Questions from IAP	Monday November 12, 2012
Response to the IAP feedback	one week after the feedback has been sent by e-mail (normally:

/questions	Monday November 19, 2012)
Decision by the Board	Mid December 2012

The exact deadlines of the assessments by the IAP and NAP, and the final decision by the board will be communicated later via the CLARIN-NL website. The target is to make the final decisions before the end of 2012.

CLARIN-NL Organization
The CLARIN-NL project is funded by NWO.

The CLARIN-NL project is coordinated by the Programme Director, prof.dr. J.E.J.M. Odijk, who is a member of the CLARIN-NL Executive Board. Drs. Jolien Scholten is the CLARIN-NL project secretary.

The International Advisory Panel (IAP) is a group of international experts in the areas of humanities, in particular linguistics, language and speech technology, and infrastructures for scientific research.

The National Advisory Panel is a group of national researchers representative for the fields of linguistics and humanities, language and speech technologies and infrastructures for scientific research.

The CLARIN-NL Board consists of national senior researchers with great expertise in governance and/or relevant technical expertise

The composition of these CLARIN-NL governance bodies can be found on the CLARIN-NL website.

CLARIN-NL Helpdesk
Contact the CLARIN-NL Helpdesk for any technical questions related to this call, e.g.

- Questions on metadata and CMDI
- Questions on data categories en ISOCAT
- Questions on standards supported in CLARIN
- Questions on web services and CLAM.
- Etc.

At the end of this document, one will find a list of acronyms and terms for technical notions with some explanation and references.

Answers to Frequently Asked Questions are provided on the Helpdesk FAQ section

Website: http://trac.clarin.nl/trac
e-mail: helpdesk@clarin.nl

CLARIN-NL Centres
A list of the current candidate CLARIN centres and their contact persons can be found on the CLARIN-NL website: http://www.clarin.nl/node/130

CLARIN-NL Office

Contact the CLARIN-NL office for any organizational or practical questions related to this call, e.g.

- Additional clarification
- Advice on eligibility of your plans
- Assistance with finding experts, data or technology required
- Assistance with finding project partners
- Assistance with selecting a CLARIN Centre.
- Etc.

The CLARIN-NL office also offers other forms of help with writing a project proposal. In particular, it is wise to have the CLARIN-NL office carry out a sanity check on a preliminary version of the proposal, so that evidently ineligible proposals and trivial mistakes in the proposals can be avoided.

Jan Odijk/Jolien Scholten
Tel: +31 30 253 6279
Fax: +31 30 253 6000
Address: Trans 10, 3512 JK Utrecht
E-mail: clarinnl@uu.nl
CLARIN-NL Website: http://www.clarin.nl.

Term	Expansion	Explanation	URL
API	Application Programmers Interface	see there	Wikipedia
Application Application software		a piece of software developed to for a specific task and a specific user group with a user interface specifically designed for the targeted user group	Wikipedia
Application Programmer		software engineer that develops an application and usually accomplishes this by making (partly) use of other software	
Application Programmers Interface		an interface that allows a piece of software to use data, procedures and methods from another piece of software	Wikipedia
Application Programmers Interface documentation		documentation of the Application Programmers Interface aimed at application programmers	
CLARIN centre		organisation that has the ambition to become a recognized CLARIN centre. See http://www.clarin.nl/node/130 for more details about CLARIN centres in the Netherlands	
CLARIN server		server hosted by or on behalf of a CLARIN centre that runs services accessible for the (CLARIN) research community via the CLARIN infrastructure	
CLARIN-compliant		For a detailed description see http://trac.clarin.nl/trac/wiki/WikiStart#CLARIN-compatible. By curating a resource as described in this call, you make it CLARIN-compliant	
CMDI	Component-based MetaData	the approach to metadata advocated in CLARIN. For details, see	

	Infrastructure	http://www.clarin.eu/cmdi	
CMDI component		A metadata component defined in accordance with CMDI.	
CMDI profile		A metadata profile defined in accordance with CMDI	
Component-based MetaData Infrastructure		the approach to metadata advocated in CLARIN . See http://www.clarin.eu/cmdi	
core component		(of a software application) the whole application minus the user interface	
curation		explained in section **Fout! Verwijzingsbron niet gevonden.**	
data category		a name and associated information to express a concept	
data category registry		a data-base or directory where data categories can be stored and viewed	
demonstration scenario		a detailed description of (sequences of) actions a user can take to have the application applied to the research data and the corresponding system responses in order to get a representative picture of the functionality offered (see p. 145). The demonstration scenario should serve as an example of how to use the application and/or solve a particular type of problem.	
demonstrator		explained in section **Fout! Verwijzingsbron niet gevonden.**, p. 145, second paragraph	
Ethical Issues		Issues related to ethics that may arise by making data available to a wider audience, e.g. privacy violations etc.	

Formal description		a formal description is a description in a form defined by well-specified rules that make it possible to completely and unambiguously determine the meaning of the description from the form	
infrastructure		(in the CLARIN context) a combination of hardware and software that allows a researcher or research team to find and use language resources, apply tools to language resources, and store own data and metadata to make them available to the research community.	
Intellectual Property		a variety of intangible assets produced by intellectual labor, such as musical, literary, and artistic works; discoveries and inventions; and words, phrases, symbols, and designs	Wikipedia
Intellectual Property Rights		All the rights and restrictions that are associated with intellectual property: who is the owner, is the resource licensed, what is one allowed to do with a resource, etc etc. Common types of intellectual property rights include copyrights, trademarks, patents, etc.	Wikipedia
interoperability		resources (data and software tools) are interoperable when they can work together in a fully automated manner with minimal human intervention	
IPR	Intellectual Property Rights	see there	
language resource		data containing language or a software tool operating on or yielding language data	
metadata		Literally: data about data; in CLARIN usually restricted to descriptions of language resources (in CMDI format)	

metadata component		A metadata component is an XML structure containing metadata elements and (recursively) other metadata components; it is used to describe a related set of properties of a language resource.	
metadata harvesting		the process of obtaining metadata from a different server (metadata provider) over a network (e.g. to get these metadata available in central data catalog)	
metadata profile		A metadata profile is an XML document definition consisting of metadata components and is used to describe the properties of a language resource	
method		(in programming languages) a special type of procedure or function	
method invocation		(in programming languages) calling a procedure or function	
registry		(digital) data-base or directory to store and access information	
resource		used in this document mostly as equivalent to language resource	
resource curation		see: curation	
Service		(here used in a software context) a software program that provides certain functionality and communicates its input and output with other software programs	
Service Oriented Architecture		Software architecture that is designed around (distributed) services interacting with each other to provide certain functionality	
web application		an application with a web-based user interface	

Web service		service that communicates with other programs over a network (in particular, the internet)	
wrapper (function)		is a software component whose main purpose is to make the functions of a second software component available and does little or no additional computation	Wikipedia
XML	eXtensible Markup Language	is a markup language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable.	Wikipedia
XML element		a particular constituent in a document with XML markup	W3CSchools
XML Schema		A specific way of defining the structure, content and semantics of XML documents	W3C

Closed Call

CLARIN-NL launches in its fourth call a Closed Call for project proposals. This called is only open for researchers who have been explicitly invited to submit a project proposal for the closed call by the CLARIN-NL Board. Information on the call will be available as of Monday July 2, 2012 and proposals can be submitted as of that data. The total budget for this closed call is limited to a maximum of € 400,000. The call is specifically open for proposals targeting resource curation projects or demonstrator projects. Such projects take existing data or tools as a basis and attempt to apply CLARIN-supported standards and best practices to make the data and tools CLARIN-compliant. Examples of such projects awarded in 2009 through 2011 can be found on the CLARIN-NL website. Only those proposals that specifically target this priority will be eligible. The maximum budget per project is 80,000€.

Important Note It is an essential requirement that a CLARIN centre is involved in each resource curation or demonstrator project. It is therefore recommended to contact a CLARIN-centre as early as possible while preparing a project proposal. The list of CLARIN centres, the kinds of resources they focus on, and their contact persons can be found on the CLARIN web site, more specifically here. Some aspects of this call text are technical in nature: the CLARIN centre will, inter alia, be able to explain and discuss these technical matters with you. At the end of this document, one will find a list of acronyms and terms for technical notions with some explanation and references.

1 Introduction

The first phase of the CLARIN-NL project focuses on the specification and design of the infrastructure. Obviously, the CLARIN infrastructure should be designed in such a way that it can incorporate the data and tools currently used by humanities researchers to address their research questions. On the other hand, the CLARIN infrastructure can only be successful if these data and tools meet certain requirements with regard to standards and interoperability, not only with regard to the data and tools themselves but also with regard to their visibility and accessibility. Many of the data and tools currently in use do not meet these requirements, or meet them only partially. One of the aims of the fourth call is to make existing data and tools compliant with the requirements on standards and interoperability currently advocated in CLARIN.

However, the list of standards, best practices and interoperability requirements currently advocated by CLARIN has not been completely fixed yet. There are good reasons for this: (1) there may be crucial data or tools for which none of the currently advocated standards or best practices is suited; (2) the currently advocated standards and best practices may be incomplete, insufficiently specific, inconvenient or even incompatible with crucial data and tools. Therefore, a second aim of the fourth call is to test whether and to which extent the currently existing data and tools can be made compatible with the standards and best practices currently advocated in CLARIN, and to get a good overview of any incompatibilities as well as suggestions for adaptations of these standards.

These matters can best be investigated by actually attempting the necessary conversions and adaptations. When applied to tools and other software this will result in CLARIN-compliant web applications that can also serve as showcases of the kind of functionality CLARIN aims to offer. A

third aim of the fourth call is therefore to make sure that the applications developed become available in demonstrators which can help promote the particular applications but also the CLARIN infrastructure as a whole.

Finally, by investigating these matters using specific data and tools that are currently in use a lot can be learnt about requirements the CLARIN infrastructure should meet or desiderata that it should offer. The fourth aim of this call is therefore obtaining a detailed list of such requirements and desiderata.

2 Project Types

In the fourth call, CLARIN-NL therefore solicits (1) projects that carry out resource (data or tools) curation, and (2) demonstrator projects.

Curation projects

CLARIN-NL aims to support the curation of digital language resources (data and tools) so that these resources can participate in the CLARIN infrastructure, more easily be accessed by interested researchers via online methods and become part of appealing new applications. Many language resources are neither visible nor accessible. Visibility is mainly achieved by standardized metadata that are being harvested by service providers. Accessibility has many different aspects:

1. The resource needs to be stored at computers that are accessible via the internet.
2. The resource needs to be identified in a persistent manner.
3. The resource needs to be interpretable, which requires a format that adheres to best practices and it requires references to registries were the used concepts are defined.

Demonstrator projects

CLARIN-NL aims to support projects that create appealing showcases of functionality that the CLARIN infrastructure should support. Such projects should make available web applications that can be used as demonstrators of functionality that supports addressing research questions of the CLARIN-NL intended user group. The development of these demonstrators will also be used to inventory a list of requirements the CLARIN infrastructure should meet and desiderata it preferably should offer.

3 Goals

The goal of a curation project is:

- Adapting specific resources so that they are visible, uniquely referable and accessible via the web, and properly documented.

The goal of a demonstrator project is:

- Creating a documented web application starting from an existing tool or application that can be used as a demonstrator and function as a showcase of the type of functionality CLARIN will incorporate and support. Within the web application there must be a clear separation between the web-based user interface and the core component. For the latter an programming language API must be defined and documented.

Important goals common to both types of projects are:

- Applying standards and best practices and makes use of the suggested CLARIN architecture and agreements to understand their limitations and the requirements for extensions.
- Establishing requirements and desiderata for the CLARIN infrastructure

In both project types the use of CLARIN-supported standards and best practices is essential.

The selection of CLARIN-supported standards and best practices is currently ongoing and it is the intention that the projects supported in this call contribute to this. A preliminary set of candidates for CLARIN-supported standards and best practices is available, and though it is by no means final we will refer to this list in this document by the term "CLARIN-standards".

4 Roles

Four roles of persons involved in the projects can be distinguished: the *user*, the *data provider* (DP), the *technology provider* (TP), and the *infrastructure specialist* (IS). The *user* is a researcher from a linguistics or humanities institute who aims to investigate one or more specific research questions. The *DP* has a certain set of digital language-related data at his/her disposal (the 'research data') that can be used for addressing the research questions of the user. The *TP* has a certain technology (e.g. language or speech technology) at its disposal and a thorough understanding of this technology (e.g. because it was developed by the TP). This technology offers functionality that makes it possible to (better) address the user's research question by applying this functionality to the research data. The *infrastructure specialist* (IS) is a specialist who has a deep understanding of the CLARIN service-oriented architecture and its requirements, and/or a specialist in data and tool format standards supported in CLARIN. IS specialists usually get involved in the project via the CLARIN-NL Helpdesk. In many cases the different roles of user, DP, TP and IS will be played by persons from different organizations, but they may originate from a single organization, and occasionally even be played by a single person. CLARIN-NL can offer assistance in bringing the right experts together, if desired.

4.1 User

The user is a researcher from a linguistics or humanities institute who aims to investigate one or more specific research questions. The project proposal should clearly describe the research question(s) of the user, and the research question(s) must be in the domain of research in the humanities in general and the study of language in particular. The project must be led by the user.

4.2 Data Provider

The DP has a certain set of digital language-related data at his/her disposal (the 'research data') that can be used for addressing the research questions of the user. The project proposal should clearly describe the research data the DP has at his/her disposal that can be used to address the research question(s), and how they can be used for this purpose. The research data must be existing digital language or language-related data. No new research data should be created in the project. The DP must have the right to make the research data available on a CLARIN server running at a dedicated centre. If the data are in a format that is not currently on the list of CLARIN standards, a resource curation project is in order. Otherwise, the data can be used in a demonstrator project. The project proposal should contain a detailed description of the research data, its current state and format, the plans to convert it if needed, justification for using different formats if applicable, and a detailed plan for dealing with the data and its metadata (see below). Any restrictions on the use of the data as well as any ethical issues that apply or may arise must be properly documented in the proposal.

4.3 The Technology Provider

The TP has a certain technology at his/her disposal that can be used as a basis for the development of a web-based application (possibly web-services based) and concomitant demonstrator, or that can be used for resource curation. Since the research data are language data, the technology will in most cases be language or speech technology.

The proposal should contain a detailed description of the available technology and its current status. It should make clear that the TP has a thorough understanding of this technology and describe how the TP obtained this understanding (e.g. because the TP developed the technology).

The intended use of the technology in the project should be described, as well as any extensions or modifications that have to be made to the technology in the project, and a plan to achieve this.

The TP must have the right to use this technology and indicate how it will be used in the project.

4.4 CLARIN-NL Helpdesk and Infrastructure Specialist

CLARIN-NL has set up a Helpdesk that project participants can turn to for all kinds of technical questions related to their project. If the relevant questions cannot be addressed by the Helpdesk functionality (FAQ section, etc.) or staff, the Helpdesk will involve infrastructure specialists. The infrastructure specialist (IS) is a specialist who has a deep understanding of the CLARIN service-oriented architecture and its requirements, and/or a specialist in data, metadata and tool format standards and best practices supported in CLARIN. The Helpdesk, and where needed IS specialists, will advise and assist the project partners.

5 Project Types

5.1 Resource Curation

Resource curation involves a number of different aspects:

1. The resource should be brought into a format that adheres to widely accepted standards and best practices currently considered as likely candidates by CLARIN.
2. Proper metadata descriptions need to be created and made available. They must be compliant with the CLARIN component metadata infrastructure (CMDI) and it should be possible to harvest and access them.
3. Metadata descriptions should include persistent identifiers that can be resolved and the CLARIN requirements should hold for the PID system.
4. All data categories used in the metadata and in the actual data (e.g. linguistic annotations) must be related to a CLARIN-recognized data category registry (currently only ISOCAT), i.e. data categories used must be mapped to corresponding ISOCAT data categories where they exist in a formal way (e.g. via an XML Schema) and new data categories must be added to ISOCAT if they do not exist there yet.
5. Provide proper documentation of the resource, at least in English.

The plan for a curation project should describe in detail how these different aspects are going to be addressed in the project.

The results of these aspects should be tested by the project participants. Setting up tests for this should be included in the project plan and the results of these tests will be included in the project's

success criteria. Example tests are e.g. a metadata harvesting test and formal procedures such as testing against an XML Schema.

The resulting resource and its metadata must be made available on a server of a recognized CLARIN centre. The project proposal must specify which (candidate) CLARIN centre this will be and concrete arrangements must have been made with this centre.

5.2 Demonstrator

In a demonstrator project a demonstrator is developed using a documented web-based application based on a technology that the TP currently has at his/her disposal. The development is carried out in close cooperation with the user

The project proposal must contain a detailed description of the targeted functionality, including input and output specifications, and how it can contribute to addressing the user's research questions. More generic functionality, i.e. functionality that can serve multiple different research questions from linguistics and humanities research will be preferred over less generic or completely idiosyncratic functionality. See below for additional criteria related to the functionality that will be used to rank proposals. Since a demonstrator project is short in duration and is relatively small, this functionality must already be available to the TP, though perhaps not in the form of a web-based application, and it perhaps only operates on data formats other than the ones listed in the CLARIN standards. The project proposal should contain a detailed description of the functionality in its current state, the targeted web-based application and its components, and a plan to achieve this. The application includes a web-based user interface that takes care of user interactions and method invocations to the core component. An Application Programming Interface (API) to the core component must be provided and documented. The TP must have the rights to make the targeted core component as well as the web-application available on a CLARIN server running at a dedicated centre.

The core component of the web-application must at least be able to operate on the research data and yield output in the formats agreed upon between user and TP. It is a pre if it can apply to other formats from the CLARIN standard list and yield additional output formats. The web-application and its core component will be used to obtain requirements and specifications of the architectural framework that is being worked out in CLARIN and may be used to test it.

A research data resource often consists of information of various kinds contained in multiple folders and multiple files of varying types. The information contained in such a resource can include documentation, source data, annotations of the source data, aggregate statistics tables on the source data and/or annotations, etc. Any tool (which eventually will very likely be integrated in the infrastructure as web service) should find out in a fully automated manner whether the research data selected by the user are appropriate input for it, and, if so, that it is applied to the right information (e.g. to the source data but not to the documentation). To achieve this it will integrate wrappers that read and write metadata and provenance information provided by the CLARIN infrastructure. A popular wrapper, used a lot and supported in CLARIN-NL is CLAM, developed by Tilburg University and currently maintained by Radboud University Nijmegen. Any requirements or desiderata that follow from this for metadata and data contents and formats should be properly documented in the documents with the requirements and desiderata for the CLARIN infrastructure.

The demonstrator consists, as a minimum, of a web application, the research data, and a demonstration scenario. A demonstration scenario is a detailed description of (sequences of) actions a user can take to have the application applied to the research data and the corresponding system responses in order to get a representative picture of the functionality offered. A movie or sequence of screen captures to illustrate the functionality is nice to have. The application will have to be installed on a CLARIN server, and the project proposal must contain a plan for doing this. It is the intention to have the demonstrator applications available for the lifetime of the CLARIN-NL project (2009-2014), so occasional support may be needed from the original developers even after the demonstrator project has finished.

The application must be tested with at least one of the common web browsers on the client side (MS IE, Firefox). Agreements about additional technical details (operating system, programming language, workspace requirements, etc) need to be made with the dedicated centre where the services should be executed.

Any vendor, platform or operating system dependent aspects of the application must be made explicit in the proposal and properly documented in the project.

The web-application and its core component should be properly documented, for users (user documentation), for application developers who want to use the core component (documentation of the API), and for technology developers who want to modify or extend the basic functionality of the application. The documentation, as well as the software user interface, must be provided at least in English.

Auxiliary Resources
The web-application may require data and other software (auxiliary resources) while running.

It must be documented which auxiliary data (e.g. a lexicon) and software (e.g. a library, converters) are needed during runtime for the application. The TP must have the right to make these auxiliary data and software available on a CLARIN server. Any restrictions on their usage (including costs) should be properly documented in the project proposal and in the documentation of the resulting application.

The application and the core component must be able to run on a dedicated CLARIN-server. An application or core component that can run only on a specific (non CLARIN) server (e.g. because it contains auxiliary resources that cannot be made available otherwise) is not acceptable.

6 Metadata

For the web application and its core component, the research data and all runtime auxiliary data used in the application, metadata descriptions must be made in accordance with the CLARIN metadata standard (CMDI). CMDI provides a flexible component-based framework for dealing with metadata; the data and tools of the projects of this call may require the development of new CMDI components or the adaptation of existing components and thus can contribute to the further development of the CMDI framework. Any required or desirable extension or modifications of the CMDI framework must be properly documented and be included in the CLARIN Requirements and Desiderata document.

7 Requirements and Desiderata for CLARIN infrastructure

One important result of both demonstrator and curation projects is a document or series of documents describing requirements and desiderata for the CLARIN infrastructure resulting from the experiences gained with the curation of the research data and/or tools, and with the development of the application, its core component and web-services derived from it. These requirements and desiderata can concern many aspects. The following is a non-exhaustive list of aspects that should be considered:

- Requirements for data formats and encoding standards
- Web-service wrappers
- Metadata elements, components and profiles
- Processing requirements
- Memory requirements
- Network Bandwidth requirements
- User workspace requirements
- API requirements (e.g. Calling conventions)
- IPR / restricted use / ethical issues requirements
- Documentation requirements
- Repository Requirements
- Requirements for registering and resolving PIDs
- Requirements related to semantic interoperability

8 Evaluation Criteria

A proposal submitted in the Closed Call will not compete with other proposals. Each proposal is evaluated in accordance with the criteria described here, and if this evaluation is positive, the project will be funded.

A proposal must describe a project that is compatible with the requirements mentioned in this call; in particular it must be a resource curation and/or a demonstrator project.

Proposals for projects will furthermore be evaluated according to the following more general criteria:

- Quality
    - Clarity and added value of the project proposal, in particular of the problem and the proposed approach
    - Suitability of the method and plan for the problem at hand
    - Feasibility of the project targets: can they be realized within the specified amount of time and with the instruments proposed?
    - Adequate balance between requested instruments and funds and proposed targets
    - Clearly specified and realistic work plan
    - Conformance to established standards and protocols as supported within CLARIN, or contribute to the development such standards and protocols.
- Project Participants
    - Competence of the participating partners (including their past performance);

    - Balanced cooperation and task assignments within the project. Justification of the composition of the team.
    - Availability of the infrastructure required for the project to be successful
    - Embedding of the work in other research programmes or projects, and/or additional funding from other funding sources is an advantage
- User-orientation of the project
    - Does the project address needs of the targeted infrastructure users (linguists and humanities researchers)?
    - Projects on tools or data that are widely in use in the targeted user community will score higher than projects focusing on lesser used tools and data.
    - More generic data or functionality, i.e. data or functionality that can serve multiple different research questions from linguistics and humanities research will be preferred over less generic or completely idiosyncratic functionality.
    - Is there cooperation with or support from the targeted (future) infrastructure users?
    - Is the resulting tool / service user-friendly, i.e. will non-technical linguistic and humanities researchers be able to use it?
    - Is dissemination of the results to the targeted users and (where appropriate) training of them planned?
- Contribution to CLARIN-NL as a whole
    - Conformance to the goals of CLARIN-NL in particular and CLARIN in general and the priorities set within them
    - Contribution to knowledge transfer and network creation. In particular, cooperation between the intended users (linguists and humanities researchers) and technology and service providers (researchers in language and speech technology, computer science, etc.) is an advantage.
    - Intellectual Property Rights and Synergy
    - Each proposal must contain clear statements about the situation of the IPR of the data and tools/technologies used, and a detailed plan to resolve any open issues.
    - The project participants have the obligation and must therefore have the rights to incorporate the core data and tools used in a project into the CLARIN infrastructure (this is a sine qua non). There has to be a clear specification and justification of the use of any data or tools needed in the project that cannot be incorporated into the CLARIN infrastructure.
    - Each proposal must show that the submitters have adequate and up-to-date knowledge of data, tools and services that are already available, so that any duplication of effort can be avoided.
- Formal compliance
    - A proposal must meet the formal requirements imposed by the CLARIN-NL organization for proposals, such as
        - conformance to the prescribed format and proposal template
        - submission before the set deadline, using the means prescribed
        - conformance to the prescribed language of the proposal

In addition, projects that are part of or fit in with international cooperation with partners from CLARIN in other countries will be preferred (of course, foreign organizations will have to find their

own funding). Finally, a project that does not meet the IPR-requirements stated or is insufficiently clear about it will be considered formally noncompliant.

9 Duration

The duration of the project must be justified. The default maximum duration is 12 months. Any duration longer than 12 months requires thorough justification.

10 Budget

The project budget must be in accordance with the tasks to be carried out, and this must be justified in the project proposal. The maximum budget is 80k€.

11 Intellectual Property Rights (IPR)

Ownership of all original data and software remains with the original owners.

An agreement must be in place between the owners of the original data and software and the project participants on the IPR of the adapted data and software before the submission date of a proposal if the owners of the original data and software are not identical to the project participants. If applicable, a copy of this agreement must be uploaded together with the project proposal. Otherwise ownership of the created adaptations and extensions will be with the creator(s).

The project participants have the obligation and therefore must have the rights to make the research data, the application, its core component, and any runtime auxiliary data or software available on a CLARIN server for use by researchers having access to the CLARIN infrastructure. This is a sine qua non. Any proposal not satisfying this requirement or being insufficiently clear about this matter will be considered to be formally noncompliant and will be rejected on these grounds.

The project proposal should describe all issues related to IPR and present solutions for them. The relations between the partners in a project must be agreed upon in a consortium agreement before the start of the project.

Practical details

Information on the CLARIN-NL Fourth Call Closed Call is available as of Monday July 2, 2012, and proposals can be submitted as of that date. The total budget for this call is limited to a maximum of € 400,000. This call is specifically open for proposals targeting resource curation projects or demonstrator projects. Only those proposals that specifically target this priority will be eligible.

Full proposals must be submitted in English and in PDF format to the CLARIN-NL electronic proposal submission system using the prescribed template (which can be found on the CLARIN-NL website). The deadline for submitting full proposals in this call has been set for Wednesday Sep 26, 2012 13:00 hours CET.

The CLARIN-NL electronic proposal submission system can be accessed as of Monday July 2, 2012 via the CLARIN-NL website. If the electronic proposal submission system would not work, contact the CLARIN-NL Office clarinnl@uu.nl.

Who can apply?

Applications can be submitted only by researchers who have been explicitly invited to submit a project proposal for the closed call by the CLARIN-NL Board.

Eligible costs

- Personnel costs directly related to the project, in accordance with the Akkoord NWO-VSNU 2008 (and any additions to it).
- A fee of maximally 3.000 € per FTE per year (or a pro rata part for less than 1 FTE per year) for covering travel and subsistence costs
- The requested funding cannot exceed 80,000€

Evaluation procedure full proposals

All eligible full proposals submitted in this call will be presented to a panel of international experts in the humanities, language and speech technology and infrastructures (International Advisory Panel, IAP). The composition of the IAP can be found on the CLARIN-NL website. If the proposals require this, the CLARIN-NL Executive Board can decide to involve additional experts in the evaluation.

This international panel will assess each eligible application based on the assessment criteria relevant to this call and may formulate a set of recommendations for improving the individual proposals. The IAP's assessment will not in this stage be sent to the proposers, but the IAP will, if needed, formulate questions and remarks for the project proposers. The project proposers will get the opportunity to answer these questions and to comment on the remarks. The response must have been received by the CLARIN-NL office before the deadline set for it (one will typically have about a week to make the response). The IAP's assessment and recommendations will be presented, together with the response of the project proposers to the National Advisory Panel (NAP). The members of the NAP who are not directly involved in the submitted proposals will also assess each application and the IAP's assessment and recommendations. On the basis of the IAP's advice and the NAP's advice, the CLARIN-NL Board will finally determine whether the project will be funded. The coordinator of each project will receive a message on the final decision together with the NAP's and IAP's assessment reports on the project.

Projects should start within three months after the applicant has received the formal notification of funding.

CLARIN-NL consortium agreement

More information as to which legal rules apply for this specific CLARIN-NL granting scheme are laid down in the CLARIN-NL consortium agreement, which can be found on the CLARIN-NL website (http://www.clarin.nl/node/72).

Timetable

Activity	**Date**
CLARIN-NL Fourth Call Open	Monday July 2, 2012
CLARIN-NL Fourth Call	Thursday August 30, 2012 (afternoon)

Information Session	
Deadline Proposal Submission	Wednesday September 26, 2012 13:00hrs CET
Feedback / Questions from IAP	Monday November 12, 2012
Response to the IAP feedback /questions	one week after the feedback has been sent by e-mail (normally: Monday November 19, 2012)
Decision by the Board	Mid December 2012

The exact date of the final decision by the board will be communicated later via the CLARIN-NL website.

CLARIN-NL Organization
The CLARIN-NL project is funded by NWO.

The CLARIN-NL project is coordinated by the Programme Director, prof.dr. J.E.J.M. Odijk, who is a member of the CLARIN-NL Executive Board. Drs. Jolien Scholten is the CLARIN-NL project secretary.

The International Advisory Panel (IAP) is a group of international experts in the areas of humanities, in particular linguistics, language and speech technology, and infrastructures for scientific research.

The National Advisory Panel is a group of national researchers representative for the fields of linguistics and humanities, language and speech technologies and infrastructures for scientific research.

The CLARIN-NL Board consists of national senior researchers with great expertise in governance and/or relevant technical expertise

The composition of these CLARIN-NL governance bodies can be found on the CLARIN-NL website.

CLARIN-NL Helpdesk
Contact the CLARIN-NL Helpdesk for any technical questions related to this call, e.g.

- Questions on metadata and CMDI
- Questions on data categories en ISOCAT
- Questions on standards supported in CLARIN
- Questions on web services and CLAM.
- Etc.

At the end of this document, one will find a list of acronyms and terms for technical notions with some explanation and references.

Answers to Frequently Asked Questions are provided on the Helpdesk FAQ section

Website: http://trac.clarin.nl/trac
e-mail: helpdesk@clarin.nl

CLARIN-NL Centres

A list of the current candidate CLARIN centres and their contact persons can be found on the CLARIN-NL website: http://www.clarin.nl/node/130

CLARIN-NL Office

Contact the CLARIN-NL office for any organizational or practical questions related to this call, e.g.

- Additional clarification
- Advice on eligibility of your plans
- Assistance with finding experts, data or technology required
- \Assistance with finding project partners
- Assistance with selecting a CLARIN Centre.
- Etc.

The CLARIN-NL office also offers other forms of help with writing a project proposal. In particular, it is wise to have the CLARIN-NL office carry out a sanity check on a preliminary version of the proposal, so that evidently ineligible proposals and trivial mistakes in the proposals can be avoided.

Jan Odijk/Jolien Scholten
Tel: +31 30 253 6279
Fax: +31 30 253 6000
Address: Trans 10, 3512 JK Utrecht
E-mail: clarinnl@uu.nl
CLARIN-NL Website: http://www.clarin.nl

Term	Expansion	Explanation
API	Application Programmers Interface	see there
Application Application software		a piece of software developed to for a specific task and a specif user interface specifically designed for the targeted user group
Application Programmer		software engineer that develops an application and usually acc making (partly) use of other software
Application Programmers Interface		an interface that allows a piece of software to use data, proced from another piece of software
Application Programmers Interface documentation		documentation of the Application Programmers Interface aime programmers
CLARIN centre		organisation that has the ambition to become a recognized CLA http://www.clarin.nl/node/130 for more details about CLARIN Netherlands
CLARIN server		server hosted by or on behalf of a CLARIN centre that runs serv the (CLARIN) research community via the CLARIN infrastructure
CLARIN-compliant		For a detailed description see http://trac.clarin.nl/trac/wiki/Wi compatible. By curating a resource as described in this call, you compliant
CMDI	Component-based MetaData Infrastructure	the approach to metadata advocated in CLARIN. For details, se http://www.clarin.eu/cmdi
CMDI component		A metadata component defined in accordance with CMDI.
CMDI profile		A metadata profile defined in accordance with CMDI
Component-based MetaData Infrastructure		the approach to metadata advocated in CLARIN . See http://ww
core component		(of a software application) the whole application minus the use
curation		explained in section **Fout! Verwijzingsbron niet gevonden.**
data category		a name and associated information to express a concept
data category registry		a data-base or directory where data categories can be stored a

demonstration scenario		a detailed description of (sequences of) actions a user can take application applied to the research data and the corresponding order to get a representative picture of the functionality offere demonstration scenario should serve as an example of how to and/or solve a particular type of problem.
demonstrator		explained in section **Fout! Verwijzingsbron niet gevonden.**, p. paragraph
Ethical Issues		Issues related to ethics that may arise by making data available audience, e.g. privacy violations etc.
Formal description		a formal description is a description in a form defined by well make it possible to completely and unambiguously determine description from the form
infrastructure		(in the CLARIN context) a combination of hardware and softwa researcher or research team to find and use language resource language resources, and store own data and metadata to make the research community.
Intellectual Property		a variety of intangible assets produced by intellectual labor, su literary, and artistic works; discoveries and inventions; and wor symbols, and designs
Intellectual Property Rights		All the rights and restrictions that are associated with intellectu the owner, is the resource licensed, what is one allowed to do etc. Common types of intellectual property rights include copyr patents, etc.
interoperability		resources (data and software tools) are interoperable when the in a fully automated manner with minimal human intervention
IPR	Intellectual Property Rights	see there
language resource		data containing language or a software tool operating on or yie
metadata		Literally: data about data; in CLARIN usually restricted to descr resources (in CMDI format)
metadata component		A metadata component is an XML structure containing metada (recursively) other metadata components; it is used to describe properties of a language resource.

metadata harvesting		the process of obtaining metadata from a different server (met a network (e.g. to get these metadata available in central data
metadata profile		A metadata profile is an XML document definition consisting of components and is used to describe the properties of a langua
method		(in programming languages) a special type of procedure or func
method invocation		(in programming languages) calling a procedure or function
registry		(digital) data-base or directory to store and access information
resource		used in this document mostly as equivalent to language resour
resource curation		see: curation
Service		(here used in a software context) a software program that prov functionality and communicates its input and output with othe
Service Oriented Architecture		Software architecture that is designed around (distributed) ser each other to provide certain functionality
web application		an application with a web-based user interface
Web service		service that communicates with other programs over a networ internet)
wrapper (function)		is a software component whose main purpose is to make the fu software component available and does little or no additional c
XML	eXtensible Markup Language	is a markup language that defines a set of rules for encoding do that is both human-readable and machine-readable.
XML element		a particular constituent in a document with XML markup
XML Schema		A specific way of defining the structure, content and semantics

Call3 (Open Call)

CLARIN-NL Third Open Call

CLARIN-NL launches the third open call for project proposals. This call is open from Friday July 3, 2011. The total budget for this call is limited to a maximum of € 600,000. This call is specifically open for proposals targeting resource curation projects or demonstrator projects. Such projects take existing data or tools as a basis and attempt to apply CLARIN-supported standards and best practices to make the data and tools CLARIN-compliant. Examples of such projects awarded in 2009 and 2010 can be found on the CLARIN-NL website.[1] Only those proposals that specifically target this priority will be eligible. The maximum budget per project is 120,000€.

Introduction

The first phase of the CLARIN-NL project focuses on the specification and design of the infrastructure. Obviously, the CLARIN infrastructure should be designed in such a way that it can incorporate the data and tools currently used by humanities researchers to address their research questions. On the other hand, the CLARIN infrastructure can only be successful if these data and tools meet certain requirements with regard to standards and interoperability, not only with regard to the data and tools themselves but also with regard to their visibility and accessibility. Many of the data and tools currently in use do not meet these requirements, or meet them only partially. One of the aims of the third call is to make existing data and tools compliant with the requirements on standards and interoperability currently advocated in CLARIN.

However, the list of standards, best practices and interoperability requirements currently advocated by CLARIN has not been completely fixed yet. There are good reasons for this: (1) there may be crucial data or tools for which none of the currently advocated standards or best practices is suited; (2) the currently advocated standards and best practices may be incomplete, insufficiently specific, inconvenient or even incompatible with crucial data and tools. Therefore, a second aim of the third call is to test whether and to which extent the currently existing data and tools can be made compatible with the standards and best practices currently advocated in CLARIN, and to get a good overview of any incompatibilities as well as suggestions for adaptations of these standards.

These matters can best be investigated by actually attempting the necessary conversions and adaptations. When applied to tools and other software this will result in CLARIN-

[1] http://www.clarin.nl/node/76

compliant web applications that can also serve as showcases of the kind of functionality CLARIN aims to offer. A third aim of the third call is therefore to make sure that the applications developed become available in demonstrators which can help promote the particular applications but also the CLARIN infrastructure as a whole.

Finally, by investigating these matters using specific data and tools that are currently in use a lot can be learnt about requirements the CLARIN infrastructure should meet or desiderata that it should offer. The fourth aim of this call is therefore obtaining a detailed list of such requirements and desiderata.

Project Types

In the third call, CLARIN-NL therefore solicits (1) projects that carry out resource (data or tools) curation, and (2) demonstrator projects.

Curation projects

CLARIN-NL aims to support the curation of digital language resources (data and tools) so that these resources can participate in the CLARIN infrastructure, more easily be accessed by interested researchers via online methods and become part of appealing new applications. Many language resources are neither visible nor accessible. Visibility is mainly achieved by standardized metadata that are being harvested by service providers. Accessibility has many different aspects:

1. The resource needs to be stored at computers that are accessible.
2. The resource needs to be identified in a persistent manner.
3. The resource needs to be interpretable, which requires a format that adheres to best practices and it requires references to registries were the used concepts are defined.

Demonstrator projects

CLARIN-NL aims to support projects that create appealing showcases of functionality that the CLARIN infrastructure should support. Such projects should make available web applications that can be used as demonstrators of functionality that supports addressing research questions of the CLARIN-NL intended user group. The development of these demonstrators will also be used to inventory a list of requirements the CLARIN infrastructure should meet and desiderata it preferably should offer.

Goals

The goal of a **curation project** is:

- Adapting specific resources so that they are visible, uniquely referable and accessible via the web, and properly documented.

The goal of a **demonstrator project** is:

- Creating a documented web application starting from an existing tool or application that can be used as a demonstrator and function as a showcase of the type of functionality CLARIN will incorporate and support. Within the web application there must be a clear separation between the web-based user interface and the core component. For the latter an API must be defined and documented.

Important goals **common** to both types of projects are:

- Applying standards and best practices and makes use of the suggested CLARIN architecture and agreements to understand their limitations and the requirements for extensions.
- Establishing requirements and desiderata for the CLARIN infrastructure

In both project types the use of CLARIN-supported standards and best practices is essential.

The selection of CLARIN-supported standards and best practices is currently ongoing and it is the intention that the projects supported in this call contribute to this. A preliminary set of candidates for CLARIN-supported standards and best practices is available,[2] and though it is by no means final we will refer to this list in this document by the term "CLARIN-standards".

Roles

Four roles of persons involved in the projects can be distinguished: the *user*, the *data provider* (DP), the *technology provider (TP)*, and the *infrastructure specialist (IS)*. The *user* is a researcher from a linguistics or humanities institute who aims to investigate one or more specific research questions. The DP has a certain set of digital language-related data at his/her disposal (the 'research data') that can be used for addressing the research questions of the user. The TP has a certain technology (e.g. language or speech technology) at its disposal and a thorough understanding of this technology (e.g. because it was developed by the TP). This technology offers functionality that makes it possible to (better) address the user's research question by applying this functionality to the research data. The *infrastructure specialist* (IS) is a specialist who has a deep understanding of the CLARIN service-oriented architecture and its requirements, and/or a specialist in data and tool format standards supported in CLARIN. IS specialists usually get involved in the project via the *CLARIN-NL Helpdesk*.[3] In many cases the different roles of user, DP, TP and IS will be played by persons from different organizations, but they may originate from a single organization, and occasionally even be played by a single person. CLARIN-NL can offer assistance in bringing the right experts together, if desired.

User

[2] http://www.clarin.eu/recommendations

[3] helpdesk@clarin.nl

The *user* is a researcher from a linguistics or humanities institute who aims to investigate one or more specific research questions. The project proposal should clearly describe the research question(s) of the user, and the research question(s) must be in the domain of research in the humanities in general and the study of language in particular. The project must be led by the user.

Data Provider

The DP has a certain set of digital language-related data at his/her disposal (the 'research data') that can be used for addressing the research questions of the user. The project proposal should clearly describe the research data the DP has at his/her disposal that can be used to address the research question(s), and how they can be used for this purpose. The research data must be existing digital language or language-related data. No new research data should be created in the project. The DP must have the right to make the research data available on a CLARIN server running at a dedicated centre. If the data are in a format that is not currently on the list of CLARIN standards, a resource curation project is in order. Otherwise, the data can be used in a demonstrator project. The project proposal should contain a detailed description of the research data, its current state and format, the plans to convert it if needed, justification for using different formats if applicable, and a detailed plan for dealing with the data and its metadata (see below). Any restrictions on the use of the data as well as any ethical issues that apply or may arise must be properly documented in the proposal.

The Technology Provider

The TP has a certain technology at his/her disposal that can be used as a basis for the development of a web-based application (possibly web-services based) and concomitant demonstrator, or that can be used for resource curation. Since the research data are language data, the technology will in most cases be language or speech technology.

The proposal should contain a detailed description of the available technology and its current status. It should make clear that the TP has a thorough understanding of this technology and describe how the TP obtained this understanding (e.g. because the TP developed the technology).

The intended use of the technology in the project should be described, as well as any extensions or modifications that have to be made to the technology in the project, and a plan to achieve this.

The TP must have the right to use this technology and indicate how it will be used in the project.

CLARIN-NL Helpdesk and Infrastructure Specialist

CLARIN-NL has set up a *Helpdesk* that project participants can turn to for all kinds of technical questions related to their project. If the relevant questions cannot be addressed by

the *Helpdesk* functionality (FAQ section, etc.) or staff, the Helpdesk will involve infrastructure specialists. The infrastructure specialist (IS) is a specialist who has a deep understanding of the CLARIN service-oriented architecture and its requirements, and/or a specialist in data, metadata and tool format standards and best practices supported in CLARIN. The *Helpdesk*, and where needed IS specialists, will advise and assist the project partners.

Project Types

Resource Curation

Resource curation involves a number of different aspects:

1. The resource should be brought into a format that adheres to widely accepted standards and best practices currently considered as likely candidates by CLARIN.
2. Proper metadata descriptions need to be created and made available. They must be compliant with the CLARIN component metadata infrastructure (CMDI) and it should be possible to harvest and access them.
3. Metadata descriptions should include persistent identifiers that can be resolved and the CLARIN requirements should hold for the PID system.
4. The linguistic encoding must be related to the data category registry, i.e. data categories used must be mapped to corresponding ISOCAT[4] data categories where they exist in a formal way (e.g. via an XML Schema) and new data categories must be added to ISOCAT if they do not exist there yet.
5. Provide proper documentation of the resource, at least in English.

The plan for a curation project should describe in detail how these different aspects are going to be addressed in the project.

The results of these aspects should be tested by the project participants. Setting up tests for this should be included in the project plan and the results of these tests will be included in the project's success criteria. Example tests are e.g. a metadata harvesting test and formal procedures such as testing against an XML Schema.

The resulting resource and its metadata must be made available on a server of a recognized CLARIN centre. The project proposal must specify which (candidate) CLARIN centre this will be and concrete arrangements must have been made with this centre.

Demonstrator

In a demonstrator project a demonstrator is developed using a documented web-based application based on a technology that the TP currently has at his/her disposal. The development is carried out in close cooperation with the user

The project proposal must contain a detailed description of the targeted functionality, including input and output specifications, and how it can contribute to addressing the user's research questions. More generic functionality, i.e. functionality that can serve multiple

[4] http://www.isocat.org/

different research questions from linguistics and humanities research will be preferred over less generic or completely idiosyncratic functionality. See below for additional criteria related to the functionality that will be used to rank proposals. Since a demonstrator project is short in duration and is relatively small, this functionality must already be available to the TP, though perhaps not in the form of a web-based application, and it perhaps only operates on data formats other than the ones listed in the CLARIN standards. The project proposal should contain a detailed description of the functionality in its current state, the targeted web-based application and its components, and a plan to achieve this. The application includes a web-based user interface that takes care of user interactions and method invocations to the core component. An Application Programming Interface (API) to the core component must be provided and documented. The TP must have the rights to make the targeted core component as well as the web-application available on a CLARIN server running at a dedicated centre.

The core component of the web-application must at least be able to operate on the research data and yield output in the formats agreed upon between user and TP. It is a pre if it can apply to other formats from the CLARIN standard list and yield additional output formats. The web-application and its core component will be used to obtain requirements and specifications of the architectural framework that is being worked out in CLARIN and may be used to test it.

A research data resource often consists of information of various kinds contained in multiple folders and multiple files of varying types. The information contained in such a resource can include documentation, source data, annotations of the source data, aggregate statistics tables on the source data and/or annotations, etc. Any tool (which eventually will very likely be integrated in the infrastructure as web service) should find out in a fully automated manner whether the research data selected by the user are appropriate input for it, and, if so, that it is applied to the right information (e.g. to the source data but not to the documentation). To achieve this it will integrate wrappers that read and write metadata and provenance information provided by the CLARIN infrastructure. A popular wrapper, used a lot and supported in CLARIN-NL is CLAM, developed by Tilburg University.[5] Any requirements or desiderata that follow from this for metadata and data contents and formats should be properly documented in the documents with the requirements and desiderata for the CLARIN infrastructure.

The demonstrator consists, as a minimum, of a web application, the research data, and a demonstration scenario. A demonstration scenario is a detailed description of example (sequences of) actions a user can take to have the application applied to the research data and the corresponding system responses in order to get a representative picture of the functionality offered. A movie or sequence of screen captures to illustrate the functionality is nice to have. The application will have to be installed on a CLARIN server, and the project proposal must contain a plan for doing this. It is the intention to have the demonstrator applications available for the lifetime of the CLARIN-NL project (2009-2014), so occasional

[5] http://ilk.uvt.nl/clam

support may be needed from the original developers even after the demonstrator project has finished.

The application must be tested with at least one of the common web browsers on the client side (MS IE, Firefox). Agreements about additional technical details (operating system, programming language, workspace requirements, etc) need to be made with the dedicated centre where the services should be executed.

Any vendor, platform or operating system dependent aspects of the application must be made explicit in the proposal and properly documented in the project.

The web-application and its core component should be properly documented, for users (user documentation), for application developers who want to use the core component (documentation of the API), and for technology developers who want to modify or extend the basic functionality of the application. The documentation, as well as the software user interface, must be provided at least in English.

Auxiliary Resources

The web-application may require data and other software (auxiliary resources) while running.

It must be documented which auxiliary data (e.g. a lexicon) and software (e.g. a library, converters) are needed during runtime for the application. The TP must have the right to make these auxiliary data and software available on a CLARIN server. Any restrictions on their usage (including costs) should be properly documented in the project proposal and in the documentation of the resulting application.

The application and the core component must be able to run on a dedicated CLARIN-server. An application or core component that can run only on a specific (non CLARIN) server (e.g. because it contains auxiliary resources that cannot be made available otherwise) is not acceptable.

Metadata

For the web application and its core component, the research data and all runtime auxiliary data used in the application, metadata descriptions must be made in accordance with the CLARIN metadata standard (CMDI).[6] CMDI provides a flexible component-based framework for dealing with metadata; the data and tools of the projects of this call may require the development of new CMDI components or the adaptation of existing components and thus can contribute to the further development of the CMDI framework.[7] Any required or

[6] http://www.clarin.eu/cmdi Login is required for editing.

[7] Existing components can be found in the Metadata Component Registry: http://catalog.clarin.eu/ds/ComponentRegistry/#

desirable extension or modifications of the CMDI framework must be properly documented and be included in the CLARIN Requirements and Desiderata document.

Requirements and Desiderata for CLARIN infrastructure

One important result of both demonstrator and curation projects is a document or series of documents describing requirements and desiderata for the CLARIN infrastructure resulting from the experiences gained with the curation of the research data and/or tools, and with the development of the application, its core component and web-services derived from it. These requirements and desiderata can concern many aspects. The following is a non-exhaustive list of aspects that should be considered:

- Requirements for data formats and encoding standards
- Web-service wrappers
- Metadata elements and formats
- Processing requirements
- Memory requirements
- Network Bandwidth requirements
- User workspace requirements
- API requirements (e.g. Calling conventions)
- IPR / restricted use / ethical issues requirements
- Documentation requirements
- Repository Requirements
- Requirements for registering and resolving PIDs
- Requirements related to semantic interoperability

Evaluation Criteria

Proposals are elicited from humanities topics and disciplines, with a priority for the following (sub)disciplines:

1. Literary studies
2. Communication and media studies (focused on the new media such as texting (sms), chat, blogs, twitter, etc.)
3. History
4. Linguistics, in particular
    a. Syntax
    b. Phonology
    c. Semantics
    d. Descriptive linguistics and language documentation
    e. Discourse

A proposal must describe a project that is compatible with the requirements mentioned in this call; in particular it must be a resource curation and/or a demonstrator project.

Proposals for projects will furthermore be evaluated and ranked according to the following more general criteria:

- **Quality**

  - Clarity and added value of the project proposal, in particular of the problem and the proposed approach
  - Suitability of the method and plan for the problem at hand
  - Feasibility of the project targets: can they be realized within the specified amount of time and with the instruments proposed?
  - Adequate balance between requested instruments and funds and proposed targets
  - Clearly specified and realistic work plan
  - Conformance to established standards and protocols as supported within CLARIN, or contribute to the development such standards and protocols.
- **Project Participants**
  - Competence of the participating partners (including their past performance);
  - Balanced cooperation and task assignments within the project. Justification of the composition of the team.
  - Availability of the infrastructure required for the project to be successful
  - Embedding of the work in other research programmes or projects, and/or additional funding from other funding sources is an advantage
- **User-orientation of the project**
  - Does the project address needs of the targeted infrastructure users (linguists and humanities researchers)?
  - Projects on tools or data that are widely in use in the targeted user community will score higher than projects focusing on lesser used tools and data.
  - Is there cooperation with or support from the targeted (future) infrastructure users?
  - Is the resulting tool / service user-friendly, i.e. will non-technical linguistic and humanities researchers be able to use it?
  - Is dissemination of the results to the targeted users and (where appropriate) training of them planned?
- **Contribution to CLARIN-NL as a whole**
  - Conformance to the goals of CLARIN-NL in particular and CLARIN in general and the priorities set within them
  - Contribution to knowledge transfer and network creation. In particular, cooperation between the intended users (linguists and humanities researchers) and technology and service providers (researchers in language and speech technology, computer science, etc.) is an advantage.
- **Intellectual Property Rights and Synergy**
  - Each proposal must contain clear statements about the situation of the IPR of the data and tools/technologies used, and a detailed plan to resolve any open issues.
  - The project participants have the obligation and must therefore have the rights to incorporate the core data and tools used in a project into the CLARIN infrastructure (this is a sine qua non). There has to be a clear specification and justification of the use of any data or tools needed in the project that cannot be incorporated into the CLARIN infrastructure.
  - Each proposal must show that the submitters have adequate and up-to-date knowledge of data, tools and services that are already available, so that any duplication of effort can be avoided.
- **Formal compliance**

- A proposal must meet the formal requirements imposed by the CLARIN-NL organization for proposals, such as
  - conformance to the prescribed format and proposal template
  - submission before the set deadline, using the means prescribed
  - conformance to the prescribed language of the proposal

In addition, more generic applications and data, i.e. applications and data that can serve multiple different research questions will be preferred over less generic applications and data. Projects that are part of or fit in with international cooperation with partners from CLARIN in other countries will be preferred (of course, foreign organizations will have to find their own funding). Finally, a project that does not meet the IPR-requirements stated or is insufficiently clear about it will be considered formally noncompliant.

Duration

The duration of the project must be justified. The default maximum duration is 12 months. Any duration longer than 12 months requires thorough justification.

Budget

The project budget must be in accordance with the tasks to be carried out, and this must be justified in the project proposal. The maximum budget is 120k€.

Intellectual Property Rights (IPR)

Ownership of all original data and software remains with the original owners.

An agreement must be in place between the owners of the original data and software and the project participants on the IPR of the adapted data and software before the submission date of a proposal if the owners of the original data and software are not identical to the project participants. If applicable, a copy of this agreement must be uploaded together with the project proposal. Otherwise ownership of the created adaptations and extensions will be with the creator(s).

The project participants have the obligation and therefore must have the rights to make the research data, the application, its core component, and any runtime auxiliary data or software available on a CLARIN server for use by researchers having access to the CLARIN infrastructure. This is a *sine qua non*. Any proposal not satisfying this requirement or being insufficiently clear about this matter will be considered to be formally noncompliant and will be rejected on these grounds.

The project proposal should describe all issues related to IPR and present solutions for them. The relations between the partners in a project must be agreed upon in a consortium agreement before the start of the project.

Practical details
On the submission and evaluation procedure regarding the
CLARIN-NL Third Open call

Call for Proposals

The CLARIN-NL Third Open Call is open from Friday July 3, 2011. The total budget for this call is limited to a maximum of € 600,000. This call is specifically open for proposals targeting resource curation projects or demonstrator projects. Only those proposals that specifically target this priority will be eligible.

Full proposals must be submitted in English and in PDF format to the CLARIN-NL electronic proposal submission system using the prescribed template (which can be found on the CLARIN-NL website). The deadline for submitting full proposals in this call has been set for Friday September 30, 2011 13:00 hours CET.

The CLARIN-NL electronic proposal submission system can be accessed as of July 3, 2011 via the CLARIN-NL website. If the electronic proposal submission system would not work, contact the CLARIN-NL Office clarinnl@uu.nl.

Who can apply?

Applications can be submitted only by researchers affiliated to CLARIN-NL participants that have signed the CLARIN-NL consortium agreement. The list of CLARIN-NL participants is available on the CLARIN-NL website. The main applicant and coordinator of the project must be a researcher from linguistics or humanities more broadly.

CLARIN-NL is in principle open to new participants. If your institute is not currently a CLARIN-NL participant, contact the CLARIN-NL office if you want your organization to become a participant in CLARIN-NL well in advance of the submission deadline.

Applicants who are planning to submit a proposal are strongly advised to contact the CLARIN-NL Office for an eligibility check of their plans, or for assistance in finding suitable partners.

Eligible costs

- Personnel costs directly related to the project, in accordance with the *Akkoord NWO-VSNU 2008* (and any additions to it).[8]
- A fee of maximally 3.000 € per FTE per year (or a pro rata part for less than 1 FTE per year) for covering travel and subsistence costs
- The requested funding cannot exceed 120,000€

[8] http://www.nwo.nl/nwohome.nsf/pages/NWOP_67QK4E more specifically http://www.nwo.nl/files.nsf/pages/NWOA_7LYGWY/$file/Ondertekende_Overeenkomst_NWO_VSNU.pdf . See also http://www.nwo.nl/files.nsf/pages/SPES_5VEDDR/$file/NWO-Regeling%20Subsidies%20per%201%20mei%202011.pdf for general guidelines that CLARIN-NL will follow where applicable.

Evaluation procedure full proposals

All eligible full proposals submitted in this call will be presented to a panel of international experts in the humanities, language and speech technology and infrastructures (International Advisory Panel, IAP). The composition of the IAP can be found on the CLARIN-NL website. If the proposals require this, the CLARIN-NL Executive Board can decide to involve additional experts in the evaluation.

This international panel will assess and rank the eligible applications based on the assessment criteria relevant to this call and may formulate a set of recommendations for improving the individual proposals. The panel's assessment and recommendations will be presented to the National Advisory Panel (NAP). The members of the NAP who are not directly involved in the submitted proposals will also assess the applications and the IAP's assessment and recommendations and make its own recommendations. On the basis of the IAP's advice and the NAP's advice, the CLARIN-NL Board will finally determine which projects will be funded.

Projects should start within three months after the applicant has received the formal notification of funding.

CLARIN-NL *consortium agreement*

More information as to which legal rules apply for this specific CLARIN-NL granting scheme are laid down in the *CLARIN-NL consortium agreement*, which can be found on the CLARIN-NL website (http://www.clarin.nl/node/72).

Timetable

Activity	**Date**
CLARIN-NL Third Call Open	Sunday July 3, 2011
CLARIN-NL Third Call Information Session	Thursday August 25, 2011 (afternoon)
Deadline Proposal Submission	Friday September 30, 2011 13:00hrs CET

The exact deadlines of the assessments by the IAP and NAP, and the final decision by the board will be communicated later via the CLARIN-NL website. The target is to make the final decisions before the end of 2011.

CLARIN-NL Organization

The CLARIN-NL project is funded by NWO.

The CLARIN-NL project is coordinated by the Programme Director, prof.dr. J.E.J.M. Odijk, who is a member of the CLARIN-NL Executive Board. Drs. Jolien Scholten is the CLARIN-NL project secretary.

The International Advisory Panel (IAP) is a group of international experts in the areas of humanities, in particular linguistics, language and speech technology, and infrastructures for scientific research.

The National Advisory Panel is a group of national researchers representative for the fields of linguistics and humanities, language and speech technologies and infrastructures for scientific research.

The CLARIN-NL Board consists of national senior researchers with great expertise in governance and/or relevant technical expertise

The composition of these CLARIN-NL governance bodies can be found on the CLARIN-NL website.

CLARIN-NL Helpdesk

Contact the CLARIN-NL Helpdesk for any technical questions related to this call, e.g.

- Questions on metadata and CMDI
- Questions on data categories en ISOCAT
- Questions on standards supported in CLARIN
- Questions on web services and CLAM.
- Etc.

CLARIN-NL Helpdesk
e-mail: helpdesk@clarin.nl

CLARIN-NL Centres
A list of the current candidate CLARIN centres and their contact persons can be found on the CLARIN-NL website: http://www.clarin.nl/node/130

CLARIN-NL Office

Contact the CLARIN-NL office for any organizational or practical questions related to this call, e.g.

- Additional clarification
- Advice on eligibility of your plans
- Assistance with finding experts, data or technology required
- Assistance with finding project partners
- Assistance with selecting a CLARIN Centre.
- Etc.

Jan Odijk/Jolien Scholten

Tel: +31 30 253 6279
Fax: +31 30 253 6000
Address: Trans 10, 3512 JK Utrecht
E-mail: clarinnl@uu.nl
CLARIN-NL Website: http://www,clarin,nl .

Call 3 (Closed Call)

CLARIN-NL Third Call: Closed Call

CLARIN-NL launches in its third call a Closed Call for project proposals. **This called is only open for researchers who have been explicitly invited to submit a project proposal for the closed call by the CLARIN-NL Board**. The call is open from Friday July 3, 2011. The total budget for this closed call is limited to a maximum of € 400,000. The call is specifically open for proposals targeting resource curation projects or demonstrator projects. Such projects take existing data or tools as a basis and attempt to apply CLARIN-supported standards and best practices to make the data and tools CLARIN-compliant. Examples of such projects awarded in 2009 and 2010 can be found on the CLARIN-NL website.[9] Only those proposals that specifically target this priority will be eligible. The maximum budget per project is 80,000€.

Introduction

The first phase of the CLARIN-NL project focuses on the specification and design of the infrastructure. Obviously, the CLARIN infrastructure should be designed in such a way that it can incorporate the data and tools currently used by humanities researchers to address their research questions. On the other hand, the CLARIN infrastructure can only be successful if these data and tools meet certain requirements with regard to standards and interoperability, not only with regard to the data and tools themselves but also with regard to their visibility and accessibility. Many of the data and tools currently in use do not meet these requirements, or meet them only partially. One of the aims of the third call is to make existing data and tools compliant with the requirements on standards and interoperability currently advocated in CLARIN.

However, the list of standards, best practices and interoperability requirements currently advocated by CLARIN has not been completely fixed yet. There are good reasons for this: (1) there may be crucial data or tools for which none of the currently advocated standards or best practices is suited; (2) the currently advocated standards and best practices may be incomplete, insufficiently specific, inconvenient or even incompatible with crucial data and tools. Therefore, a second aim of the third call is to test whether and to which extent the currently existing data and tools can be made compatible with the standards and best practices currently advocated in CLARIN, and to get a good overview of any incompatibilities as well as suggestions for adaptations of these standards.

These matters can best be investigated by actually attempting the necessary conversions and adaptations. When applied to tools and other software this will result in CLARIN-compliant web applications that can also serve as showcases of the kind of functionality CLARIN aims to offer. A third aim of the third call is therefore to make sure that the

[9] http://www.clarin.nl/node/76

applications developed become available in demonstrators which can help promote the particular applications but also the CLARIN infrastructure as a whole.

Finally, by investigating these matters using specific data and tools that are currently in use a lot can be learnt about requirements the CLARIN infrastructure should meet or desiderata that it should offer. The fourth aim of this call is therefore obtaining a detailed list of such requirements and desiderata.

Project Types

In the third call, CLARIN-NL therefore solicits (1) projects that carry out resource (data or tools) curation, and (2) demonstrator projects.

Curation projects

CLARIN-NL aims to support the curation of digital language resources (data and tools) so that these resources can participate in the CLARIN infrastructure, more easily be accessed by interested researchers via online methods and become part of appealing new applications. Many language resources are neither visible nor accessible. Visibility is mainly achieved by standardized metadata that are being harvested by service providers. Accessibility has many different aspects:

4. The resource needs to be stored at computers that are accessible.
5. The resource needs to be identified in a persistent manner.
6. The resource needs to be interpretable, which requires a format that adheres to best practices and it requires references to registries were the used concepts are defined.

Demonstrator projects

CLARIN-NL aims to support projects that create appealing showcases of functionality that the CLARIN infrastructure should support. Such projects should make available web applications that can be used as demonstrators of functionality that supports addressing research questions of the CLARIN-NL intended user group. The development of these demonstrators will also be used to inventory a list of requirements the CLARIN infrastructure should meet and desiderata it preferably should offer.

Goals

The goal of a **curation project** is:

- Adapting specific resources so that they are visible, uniquely referable and accessible via the web, and properly documented.

The goal of a **demonstrator project** is:

- Creating a documented web application starting from an existing tool or application that can be used as a demonstrator and function as a showcase of the type of functionality CLARIN will incorporate and support. Within the web application there must be a clear separation between the web-based user interface and the core component. For the latter an API must be defined and documented.

Important goals **common** to both types of projects are:

- Applying standards and best practices and makes use of the suggested CLARIN architecture and agreements to understand their limitations and the requirements for extensions.
- Establishing requirements and desiderata for the CLARIN infrastructure

In both project types the use of CLARIN-supported standards and best practices is essential.

The selection of CLARIN-supported standards and best practices is currently ongoing and it is the intention that the projects supported in this call contribute to this. A preliminary set of candidates for CLARIN-supported standards and best practices is available,[10] and though it is by no means final we will refer to this list in this document by the term "CLARIN-standards".

Roles

Four roles of persons involved in the projects can be distinguished: the *user*, the *data provider* (DP), the *technology provider (TP)*, and the *infrastructure specialist (IS)*. The *user* is a researcher from a linguistics or humanities institute who aims to investigate one or more specific research questions. The DP has a certain set of digital language-related data at his/her disposal (the 'research data') that can be used for addressing the research questions of the user. The TP has a certain technology (e.g. language or speech technology) at its disposal and a thorough understanding of this technology (e.g. because it was developed by the TP). This technology offers functionality that makes it possible to (better) address the user's research question by applying this functionality to the research data. The *infrastructure specialist* (IS) is a specialist who has a deep understanding of the CLARIN service-oriented architecture and its requirements, and/or a specialist in data and tool format standards supported in CLARIN. IS specialists usually get involved in the project via the *CLARIN-NL Helpdesk*.[11] In many cases the different roles of user, DP, TP and IS will be played by persons from different organizations, but they may originate from a single organization, and occasionally even be played by a single person. CLARIN-NL can offer assistance in bringing the right experts together, if desired.

User

[10] http://www.clarin.eu/recommendations

[11] helpdesk@clarin.nl

The *user* is a researcher from a linguistics or humanities institute who aims to investigate one or more specific research questions. The project proposal should clearly describe the research question(s) of the user, and the research question(s) must be in the domain of research in the humanities in general and the study of language in particular. The project must be led by the user.

Data Provider

The DP has a certain set of digital language-related data at his/her disposal (the 'research data') that can be used for addressing the research questions of the user. The project proposal should clearly describe the research data the DP has at his/her disposal that can be used to address the research question(s), and how they can be used for this purpose. The research data must be existing digital language or language-related data. No new research data should be created in the project. The DP must have the right to make the research data available on a CLARIN server running at a dedicated centre. If the data are in a format that is not currently on the list of CLARIN standards, a resource curation project is in order. Otherwise, the data can be used in a demonstrator project. The project proposal should contain a detailed description of the research data, its current state and format, the plans to convert it if needed, justification for using different formats if applicable, and a detailed plan for dealing with the data and its metadata (see below). Any restrictions on the use of the data as well as any ethical issues that apply or may arise must be properly documented in the proposal.

The Technology Provider

The TP has a certain technology at his/her disposal that can be used as a basis for the development of a web-based application (possibly web-services based) and concomitant demonstrator, or that can be used for resource curation. Since the research data are language data, the technology will in most cases be language or speech technology.

The proposal should contain a detailed description of the available technology and its current status. It should make clear that the TP has a thorough understanding of this technology and describe how the TP obtained this understanding (e.g. because the TP developed the technology).

The intended use of the technology in the project should be described, as well as any extensions or modifications that have to be made to the technology in the project, and a plan to achieve this.

The TP must have the right to use this technology and indicate how it will be used in the project.

CLARIN-NL Helpdesk and Infrastructure Specialist

CLARIN-NL has set up a *Helpdesk* that project participants can turn to for all kinds of technical questions related to their project. If the relevant questions cannot be addressed by

the *Helpdesk* functionality (FAQ section, etc.) or staff, the Helpdesk will involve infrastructure specialists. The infrastructure specialist (IS) is a specialist who has a deep understanding of the CLARIN service-oriented architecture and its requirements, and/or a specialist in data, metadata and tool format standards and best practices supported in CLARIN. The *Helpdesk*, and where needed IS specialists, will advise and assist the project partners.

Project Types

Resource Curation

Resource curation involves a number of different aspects:

6. The resource should be brought into a format that adheres to widely accepted standards and best practices currently considered as likely candidates by CLARIN.
7. Proper metadata descriptions need to be created and made available. They must be compliant with the CLARIN component metadata infrastructure (CMDI) and it should be possible to harvest and access them.
8. Metadata descriptions should include persistent identifiers that can be resolved and the CLARIN requirements should hold for the PID system.
9. The linguistic encoding must be related to the data category registry, i.e. data categories used must be mapped to corresponding ISOCAT[12] data categories where they exist in a formal way (e.g. via an XML Schema) and new data categories must be added to ISOCAT if they do not exist there yet.
10. Provide proper documentation of the resource, at least in English.

The plan for a curation project should describe in detail how these different aspects are going to be addressed in the project.

The results of these aspects should be tested by the project participants. Setting up tests for this should be included in the project plan and the results of these tests will be included in the project's success criteria. Example tests are e.g. a metadata harvesting test and formal procedures such as testing against an XML Schema.

The resulting resource and its metadata must be made available on a server of a recognized CLARIN centre. The project proposal must specify which (candidate) CLARIN centre this will be and concrete arrangements must have been made with this centre.

Demonstrator

In a demonstrator project a demonstrator is developed using a documented web-based application based on a technology that the TP currently has at his/her disposal. The development is carried out in close cooperation with the user

The project proposal must contain a detailed description of the targeted functionality, including input and output specifications, and how it can contribute to addressing the user's research questions. More generic functionality, i.e. functionality that can serve multiple

[12] http://www.isocat.org/

different research questions from linguistics and humanities research will be preferred over less generic or completely idiosyncratic functionality. See below for additional criteria related to the functionality that will be used to rank proposals. Since a demonstrator project is short in duration and is relatively small, this functionality must already be available to the TP, though perhaps not in the form of a web-based application, and it perhaps only operates on data formats other than the ones listed in the CLARIN standards. The project proposal should contain a detailed description of the functionality in its current state, the targeted web-based application and its components, and a plan to achieve this. The application includes a web-based user interface that takes care of user interactions and method invocations to the core component. An Application Programming Interface (API) to the core component must be provided and documented. The TP must have the rights to make the targeted core component as well as the web-application available on a CLARIN server running at a dedicated centre.

The core component of the web-application must at least be able to operate on the research data and yield output in the formats agreed upon between user and TP. It is a pre if it can apply to other formats from the CLARIN standard list and yield additional output formats. The web-application and its core component will be used to obtain requirements and specifications of the architectural framework that is being worked out in CLARIN and may be used to test it.

A research data resource often consists of information of various kinds contained in multiple folders and multiple files of varying types. The information contained in such a resource can include documentation, source data, annotations of the source data, aggregate statistics tables on the source data and/or annotations, etc. Any tool (which eventually will very likely be integrated in the infrastructure as web service) should find out in a fully automated manner whether the research data selected by the user are appropriate input for it, and, if so, that it is applied to the right information (e.g. to the source data but not to the documentation). To achieve this it will integrate wrappers that read and write metadata and provenance information provided by the CLARIN infrastructure. A popular wrapper, used a lot and supported in CLARIN-NL is CLAM, developed by Tilburg University.[13] Any requirements or desiderata that follow from this for metadata and data contents and formats should be properly documented in the documents with the requirements and desiderata for the CLARIN infrastructure.

The demonstrator consists, as a minimum, of a web application, the research data, and a demonstration scenario. A demonstration scenario is a detailed description of example (sequences of) actions a user can take to have the application applied to the research data and the corresponding system responses in order to get a representative picture of the functionality offered. A movie or sequence of screen captures to illustrate the functionality is nice to have. The application will have to be installed on a CLARIN server, and the project proposal must contain a plan for doing this. It is the intention to have the demonstrator applications available for the lifetime of the CLARIN-NL project (2009-2014), so occasional

[13] http://ilk.uvt.nl/clam

support may be needed from the original developers even after the demonstrator project has finished.

The application must be tested with at least one of the common web browsers on the client side (MS IE, Firefox). Agreements about additional technical details (operating system, programming language, workspace requirements, etc) need to be made with the dedicated centre where the services should be executed.

Any vendor, platform or operating system dependent aspects of the application must be made explicit in the proposal and properly documented in the project.

The web-application and its core component should be properly documented, for users (user documentation), for application developers who want to use the core component (documentation of the API), and for technology developers who want to modify or extend the basic functionality of the application. The documentation, as well as the software user interface, must be provided at least in English.

Auxiliary Resources

The web-application may require data and other software (auxiliary resources) while running.

It must be documented which auxiliary data (e.g. a lexicon) and software (e.g. a library, converters) are needed during runtime for the application. The TP must have the right to make these auxiliary data and software available on a CLARIN server. Any restrictions on their usage (including costs) should be properly documented in the project proposal and in the documentation of the resulting application.

The application and the core component must be able to run on a dedicated CLARIN-server. An application or core component that can run only on a specific (non CLARIN) server (e.g. because it contains auxiliary resources that cannot be made available otherwise) is not acceptable.

Metadata

For the web application and its core component, the research data and all runtime auxiliary data used in the application, metadata descriptions must be made in accordance with the CLARIN metadata standard (CMDI).[14] CMDI provides a flexible component-based framework for dealing with metadata; the data and tools of the projects of this call may require the development of new CMDI components or the adaptation of existing components and thus can contribute to the further development of the CMDI framework.[15] Any required or

[14] http://www.clarin.eu/cmdi Login is required for editing.

[15] Existing components can be found in the Metadata Component Registry: http://catalog.clarin.eu/ds/ComponentRegistry/#

desirable extension or modifications of the CMDI framework must be properly documented and be included in the CLARIN Requirements and Desiderata document.

Requirements and Desiderata for CLARIN infrastructure

One important result of both demonstrator and curation projects is a document or series of documents describing requirements and desiderata for the CLARIN infrastructure resulting from the experiences gained with the curation of the research data and/or tools, and with the development of the application, its core component and web-services derived from it. These requirements and desiderata can concern many aspects. The following is a non-exhaustive list of aspects that should be considered:

- Requirements for data formats and encoding standards
- Web-service wrappers
- Metadata elements and formats
- Processing requirements
- Memory requirements
- Network Bandwidth requirements
- User workspace requirements
- API requirements (e.g. Calling conventions)
- IPR / restricted use / ethical issues requirements
- Documentation requirements
- Repository Requirements
- Requirements for registering and resolving PIDs
- Requirements related to semantic interoperability

Evaluation Criteria

A proposal submitted in the Closed Call will not compete with other proposals. Each proposal is evaluated in accordance with the criteria described here, and if this evaluation is positive, the project will be funded.

A proposal must describe a project that is compatible with the requirements mentioned in this call; in particular it must be a resource curation and/or a demonstrator project.

Proposals for projects will furthermore be evaluated according to the following more general criteria:

- **Quality**
  - Clarity and added value of the project proposal, in particular of the problem and the proposed approach
  - Suitability of the method and plan for the problem at hand
  - Feasibility of the project targets: can they be realized within the specified amount of time and with the instruments proposed?
  - Adequate balance between requested instruments and funds and proposed targets
  - Clearly specified and realistic work plan
  - Conformance to established standards and protocols as supported within CLARIN, or contribute to the development such standards and protocols.

- **Project Participants**
  - Competence of the participating partners (including their past performance);
  - Balanced cooperation and task assignments within the project. Justification of the composition of the team.
  - Availability of the infrastructure required for the project to be successful
  - Embedding of the work in other research programmes or projects, and/or additional funding from other funding sources is an advantage
- **User-orientation of the project**
  - Does the project address needs of the targeted infrastructure users (linguists and humanities researchers)?
  - Projects on tools or data that are widely in use in the targeted user community will score higher than projects focusing on lesser used tools and data.
  - Is there cooperation with or support from the targeted (future) infrastructure users?
  - Is the resulting tool / service user-friendly, i.e. will non-technical linguistic and humanities researchers be able to use it?
  - Is dissemination of the results to the targeted users and (where appropriate) training of them planned?
- **Contribution to CLARIN-NL as a whole**
  - Conformance to the goals of CLARIN-NL in particular and CLARIN in general and the priorities set within them
  - Contribution to knowledge transfer and network creation. In particular, cooperation between the intended users (linguists and humanities researchers) and technology and service providers (researchers in language and speech technology, computer science, etc.) is an advantage.
- **Intellectual Property Rights and Synergy**
  - Each proposal must contain clear statements about the situation of the IPR of the data and tools/technologies used, and a detailed plan to resolve any open issues.
  - The project participants have the obligation and must therefore have the rights to incorporate the core data and tools used in a project into the CLARIN infrastructure (this is a sine qua non). There has to be a clear specification and justification of the use of any data or tools needed in the project that cannot be incorporated into the CLARIN infrastructure.
  - Each proposal must show that the submitters have adequate and up-to-date knowledge of data, tools and services that are already available, so that any duplication of effort can be avoided.
- **Formal compliance**
  - A proposal must meet the formal requirements imposed by the CLARIN-NL organization for proposals, such as
    - conformance to the prescribed format and proposal template
    - submission before the set deadline, using the means prescribed
    - conformance to the prescribed language of the proposal

In addition, more generic applications and data, i.e. applications and data that can serve multiple different research questions will be preferred over less generic applications and data. Projects that are part of or fit in with international cooperation with partners from CLARIN in other countries will be preferred (of course, foreign organizations will have to find

their own funding). Finally, a project that does not meet the IPR-requirements stated or is insufficiently clear about it will be considered formally noncompliant.

Duration

The duration of the project must be justified. The default maximum duration is 12 months. Any duration longer than 12 months requires thorough justification.

Budget

The project budget must be in accordance with the tasks to be carried out, and this must be justified in the project proposal. The maximum budget is 80k€.

Intellectual Property Rights (IPR)

Ownership of all original data and software remains with the original owners.

An agreement must be in place between the owners of the original data and software and the project participants on the IPR of the adapted data and software before the submission date of a proposal if the owners of the original data and software are not identical to the project participants. If applicable, a copy of this agreement must be uploaded together with the project proposal. Otherwise ownership of the created adaptations and extensions will be with the creator(s).

The project participants have the obligation and therefore must have the rights to make the research data, the application, its core component, and any runtime auxiliary data or software available on a CLARIN server for use by researchers having access to the CLARIN infrastructure. This is a *sine qua non*. Any proposal not satisfying this requirement or being insufficiently clear about this matter will be considered to be formally noncompliant and will be rejected on these grounds.

The project proposal should describe all issues related to IPR and present solutions for them. The relations between the partners in a project must be agreed upon in a consortium agreement before the start of the project.

Practical details
On the submission and evaluation procedure regarding the
CLARIN-NL Third Call: Closed Call

Call for Proposals

The CLARIN-NL Third Call Closed Call is open from Friday July 3, 2011. The total budget for this call is limited to a maximum of € 400,000. This call is specifically open for proposals targeting resource curation projects or demonstrator projects. Only those proposals that specifically target this priority will be eligible.

Full proposals must be submitted in English and in PDF format to the CLARIN-NL electronic proposal submission system using the prescribed template (which can be found on the CLARIN-NL website). The deadline for submitting full proposals in this call has been set for Friday September 30, 2011 13:00 hours CET.

The CLARIN-NL electronic proposal submission system can be accessed as of July 3, 2011 via the CLARIN-NL website. If the electronic proposal submission system would not work, contact the CLARIN-NL Office clarinnl@uu.nl.

Who can apply?

Applications can be submitted only by researchers who have been explicitly invited to submit a project proposal for the closed call by the CLARIN-NL Board.

Eligible costs

- Personnel costs directly related to the project, in accordance with the *Akkoord NWO-VSNU 2008* (and any additions to it).[16]
- A fee of maximally 3.000 € per FTE per year (or a pro rata part for less than 1 FTE per year) for covering travel and subsistence costs
- The requested funding cannot exceed 80,000€

Evaluation procedure full proposals

All eligible full proposals submitted in this call will be presented to a panel of international experts in the humanities, language and speech technology and infrastructures (International Advisory Panel, IAP). The composition of the IAP can be found on the CLARIN-NL website. If the proposals require this, the CLARIN-NL Executive Board can decide to involve additional experts in the evaluation.

[16] http://www.nwo.nl/nwohome.nsf/pages/NWOP_67QK4E more specifically http://www.nwo.nl/files.nsf/pages/NWOA_7LYGWY/$file/Ondertekende_Overeenkomst_NWO_VSNU.pdf . See also http://www.nwo.nl/files.nsf/pages/SPES_5VEDDR/$file/NWO-Regeling%20Subsidies%20per%201%20mei%202011.pdf for general guidelines that CLARIN-NL will follow where applicable.

This international panel will assess each eligible application based on the assessment criteria relevant to this call and may formulate a set of recommendations for improving the individual proposals. The panel's assessment and recommendations will be presented to the National Advisory Panel (NAP). The members of the NAP who are not directly involved in the submitted proposals will also assess each application and the IAP's assessment and recommendations. On the basis of the IAP's advice and the NAP's advice, the CLARIN-NL Board will finally determine whether the project will be funded.

Projects should start within three months after the applicant has received the formal notification of funding.

CLARIN-NL *consortium agreement*

More information as to which legal rules apply for this specific CLARIN-NL granting scheme are laid down in the *CLARIN-NL consortium agreement*, which can be found on the CLARIN-NL website (http://www.clarin.nl/node/72).

Timetable

Activity	**Date**
CLARIN-NL Third Call Open	Sunday July 3, 2011
CLARIN-NL Third Call Information Session	Thursday August 25, 2011 (afternoon)
Deadline Proposal Submission	Friday September 30, 2011 13:00hrs CET

The exact deadlines of the assessments by the IAP and NAP, and the final decision by the board will be communicated later via the CLARIN-NL website. The target is to make the final decisions before the end of 2011.

CLARIN-NL Organization

The CLARIN-NL project is funded by NWO.

The CLARIN-NL project is coordinated by the Programme Director, prof.dr. J.E.J.M. Odijk, who is a member of the CLARIN-NL Executive Board. Drs. Jolien Scholten is the CLARIN-NL project secretary.

The International Advisory Panel (IAP) is a group of international experts in the areas of humanities, in particular linguistics, language and speech technology, and infrastructures for scientific research.

The National Advisory Panel is a group of national researchers representative for the fields of linguistics and humanities, language and speech technologies and infrastructures for scientific research.

The CLARIN-NL Board consists of national senior researchers with great expertise in governance and/or relevant technical expertise

The composition of these CLARIN-NL governance bodies can be found on the CLARIN-NL website.

CLARIN-NL Helpdesk

Contact the CLARIN-NL Helpdesk for any technical questions related to this call, e.g.

- Questions on metadata and CMDI
- Questions on data categories en ISOCAT
- Questions on standards supported in CLARIN
- Questions on web services and CLAM.
- Etc.

CLARIN-NL Helpdesk
e-mail: helpdesk@clarin.nl

CLARIN-NL Centres
A list of the current candidate CLARIN centres and their contact persons can be found on the CLARIN-NL website: http://www.clarin.nl/node/130

CLARIN-NL Office

Contact the CLARIN-NL office for any organizational or practical questions related to this call, e.g.

- Additional clarification
- Advice on eligibility of your plans
- Assistance with finding experts, data or technology required
- Assistance with finding project partners
- Assistance with selecting a CLARIN Centre.
- Etc.

Jan Odijk/Jolien Scholten
Tel: +31 30 253 6279
Fax: +31 30 253 6000
Address: Trans 10, 3512 JK Utrecht
E-mail: clarinnl@uu.nl
CLARIN-NL Website: http://www,clarin,nl .

Call 4 (Open Call)

CLARIN-NL Fourth Open Call

CLARIN-NL launches the fourth open call for project proposals. This call is open from **Monday July 2, 2012**. The total budget for this call is limited to a maximum of € 600,000. This call is specifically open for proposals targeting resource curation projects or demonstrator projects. Such projects take existing data or tools as a basis and attempt to apply CLARIN-supported standards and best practices to make the data and tools CLARIN-compliant. Examples of such projects awarded in 2009 through 2011 can be found on the CLARIN-NL website.[17] Only those proposals that specifically target this priority will be eligible. The maximum budget per project is 120,000€.

Important Note It is an essential requirement that a CLARIN centre is involved in each resource curation or demonstrator project. It is therefore recommended to contact a CLARIN-centre as early as possible while preparing a project proposal. The list of CLARIN centres, the kinds of resources they focus on, and their contact persons can be found on the CLARIN web site, more specifically here.[18] Some aspects of this call text are technical in nature: the CLARIN centre will, inter alia, be able to explain and discuss these technical matters with you. At the end of this document, one will find a list of acronyms and terms for technical notions with some explanation and references.

Introduction

The first phase of the CLARIN-NL project focuses on the specification and design of the infrastructure. Obviously, the CLARIN infrastructure should be designed in such a way that it can incorporate the data and tools currently used by humanities researchers to address their research questions. On the other hand, the CLARIN infrastructure can only be successful if these data and tools meet certain requirements with regard to standards and interoperability, not only with regard to the data and tools themselves but also with regard to their visibility and accessibility. Many of the data and tools currently in use do not meet these requirements, or meet them only partially. One of the aims of the fourth call is to make existing data and tools compliant with the requirements on standards and interoperability currently advocated in CLARIN.

However, the list of standards, best practices and interoperability requirements currently advocated by CLARIN has not been completely fixed yet. There are good reasons for this: (1) there may be crucial data or tools for which none of the currently advocated standards or

[17] http://www.clarin.nl/node/76

[18] http://www.clarin.nl/node/130

best practices is suited; (2) the currently advocated standards and best practices may be incomplete, insufficiently specific, inconvenient or even incompatible with crucial data and tools. Therefore, a second aim of the fourth call is to test whether and to which extent the currently existing data and tools can be made compatible with the standards and best practices currently advocated in CLARIN, and to get a good overview of any incompatibilities as well as suggestions for adaptations of these standards.

These matters can best be investigated by actually attempting the necessary conversions and adaptations. When applied to tools and other software this will result in CLARIN-compliant web applications that can also serve as showcases of the kind of functionality CLARIN aims to offer. A third aim of the fourth call is therefore to make sure that the applications developed become available in demonstrators which can help promote the particular applications but also the CLARIN infrastructure as a whole.

Finally, by investigating these matters using specific data and tools that are currently in use a lot can be learnt about requirements the CLARIN infrastructure should meet or desiderata that it should offer. The fourth aim of this call is therefore obtaining a detailed list of such requirements and desiderata.

Project Types

In the fourth call, CLARIN-NL therefore solicits (1) projects that carry out resource (data or tools) curation, and (2) demonstrator projects.

Curation projects

CLARIN-NL aims to support the curation of digital language resources (data and tools) so that these resources can participate in the CLARIN infrastructure, more easily be accessed by interested researchers via online methods and become part of appealing new applications. Many language resources are neither visible nor accessible. Visibility is mainly achieved by standardized metadata that are being harvested by service providers. Accessibility has many different aspects:

7. The resource needs to be stored at computers that are accessible via the internet.
8. The resource needs to be identified in a persistent manner.
9. The resource needs to be interpretable, which requires a format that adheres to best practices and it requires references to registries were the used concepts are defined.

Demonstrator projects

CLARIN-NL aims to support projects that create appealing showcases of functionality that the CLARIN infrastructure should support. Such projects should make available web applications that can be used as demonstrators of functionality that supports addressing research questions of the CLARIN-NL intended user group. The development of these

demonstrators will also be used to inventory a list of requirements the CLARIN infrastructure should meet and desiderata it preferably should offer.

Goals

The goal of a **curation project** is:

- Adapting specific resources so that they are visible, uniquely referable and accessible via the web, and properly documented.

The goal of a **demonstrator project** is:

- Creating a documented web application starting from an existing tool or application that can be used as a demonstrator and function as a showcase of the type of functionality CLARIN will incorporate and support. Within the web application there must be a clear separation between the web-based user interface and the core component. For the latter an API must be defined and documented.

Important goals **common** to both types of projects are:

- Applying standards and best practices and making use of the suggested CLARIN architecture and agreements to understand their limitations and the requirements for extensions.
- Establishing requirements and desiderata for the CLARIN infrastructure

In both project types the use of CLARIN-supported standards and best practices is essential.

The selection of CLARIN-supported standards and best practices is currently ongoing and it is the intention that the projects supported in this call contribute to this. A preliminary set of candidates for CLARIN-supported standards and best practices is available,[19] and though it is by no means final we will refer to this list in this document by the term "CLARIN-standards".

Roles

Four roles of persons involved in the projects can be distinguished: the *user*, the *data provider* (DP), the *technology provider (TP)*, and the *infrastructure specialist (IS)*. The *user* is a researcher from a linguistics or humanities institute who aims to investigate one or more specific research questions. The DP has a certain set of digital language-related data at his/her disposal (the 'research data') that can be used for addressing the research questions of the user. The TP has a certain technology (e.g. language or speech technology) at its disposal and a thorough understanding of this technology (e.g. because it was developed by the TP). This technology offers functionality that makes it possible to (better) address the user's research question by applying this functionality to the research data. The *infrastructure specialist* (IS) is a specialist who has a deep understanding of the CLARIN

[19] http://www.clarin.eu/recommendations

service-oriented architecture and its requirements, and/or a specialist in data and tool format standards supported in CLARIN. IS specialists usually get involved in the project via the *CLARIN-NL Helpdesk*.[20] In many cases the different roles of user, DP, TP and IS will be played by persons from different organizations, but they may originate from a single organization, and occasionally even be played by a single person. CLARIN-NL can offer assistance in bringing the right experts together, if desired.

User

The *user* is a researcher from a linguistics or humanities institute who aims to investigate one or more specific research questions. The project proposal should clearly describe the research question(s) of the user, and the research question(s) must be in the domain of research in the humanities in general and the study of language in particular. The project must be led by the user.

Data Provider

The DP has a certain set of digital language-related data at his/her disposal (the 'research data') that can be used for addressing the research questions of the user. The project proposal should clearly describe the research data the DP has at his/her disposal that can be used to address the research question(s), and how they can be used for this purpose. The research data must be existing digital language or language-related data. No new research data should be created in the project. The DP must have the right to make the research data available on a CLARIN server running at a dedicated centre. If the data are in a format that is not currently on the list of CLARIN standards, a resource curation project is in order. Otherwise, the data can be used in a demonstrator project. The project proposal should contain a detailed description of the research data, its current state and format, the plans to convert it if needed, justification for using different formats if applicable, and a detailed plan for dealing with the data and its metadata (see below). Any restrictions on the use of the data as well as any ethical issues that apply or may arise must be properly documented in the proposal.

The Technology Provider

The TP has a certain technology at his/her disposal that can be used as a basis for the development of a web-based application (possibly web-services based) and concomitant demonstrator, or that can be used for resource curation. Since the research data are language data, the technology will in most cases be language or speech technology.

The proposal should contain a detailed description of the available technology and its current status. It should make clear that the TP has a thorough understanding of this technology and describe how the TP obtained this understanding (e.g. because the TP developed the technology).

[20] helpdesk@clarin.nl

The intended use of the technology in the project should be described, as well as any extensions or modifications that have to be made to the technology in the project, and a plan to achieve this.

The TP must have the right to use this technology and indicate how it will be used in the project.

CLARIN-NL Helpdesk and Infrastructure Specialist

CLARIN-NL has set up a *Helpdesk* that project participants can turn to for all kinds of technical questions related to their project. If the relevant questions cannot be addressed by the *Helpdesk* functionality (FAQ section, etc.) or staff, the Helpdesk will involve infrastructure specialists. The infrastructure specialist (IS) is a specialist who has a deep understanding of the CLARIN service-oriented architecture and its requirements, and/or a specialist in data, metadata and tool format standards and best practices supported in CLARIN. The *Helpdesk*, and where needed IS specialists, will advise and assist the project partners.

Project Types

Resource Curation

Resource curation involves a number of different aspects:

11. The resource should be brought into a format that adheres to widely accepted standards and best practices currently considered as likely candidates by CLARIN.
12. Proper metadata descriptions need to be created and made available. They must be compliant with the CLARIN component metadata infrastructure (CMDI) and it should be possible to harvest and access them.
13. Metadata descriptions should include persistent identifiers that can be resolved and the CLARIN requirements should hold for the PID system.
14. All data categories used in the metadata and in the actual data (e.g. linguistic annotations) must be related to a CLARIN-recognized data category registry (currently only **ISOCAT**[21]), i.e. data categories used must be mapped to corresponding ISOCAT data categories where they exist in a formal way (e.g. via an XML Schema) and new data categories must be added to ISOCAT if they do not exist there yet.
15. Provide proper documentation of the resource, at least in English.

The plan for a curation project should describe in detail how these different aspects are going to be addressed in the project.

The results of these aspects should be tested by the project participants. Setting up tests for this should be included in the project plan and the results of these tests will be included in the project's success criteria. Example tests are e.g. a metadata harvesting test and formal procedures such as testing against an XML Schema.

[21] http://www.isocat.org/

The resulting resource and its metadata must be made available on a server of a recognized CLARIN centre. The project proposal must specify which (candidate) CLARIN centre this will be and concrete arrangements must have been made with this centre.

Demonstrator

In a demonstrator project a demonstrator is developed using a documented web-based application based on a technology that the TP currently has at his/her disposal. The development is carried out in close cooperation with the user

The project proposal must contain a detailed description of the targeted functionality, including input and output specifications, and how it can contribute to addressing the user's research questions. More generic functionality, i.e. functionality that can serve multiple different research questions from linguistics and humanities research will be preferred over less generic or completely idiosyncratic functionality. See below for additional criteria related to the functionality that will be used to rank proposals. Since a demonstrator project is short in duration and is relatively small, this functionality must already be available to the TP, though perhaps not in the form of a web-based application, and it perhaps only operates on data formats other than the ones listed in the CLARIN standards. The project proposal should contain a detailed description of the functionality in its current state, the targeted web-based application and its components, and a plan to achieve this. The application includes a web-based user interface that takes care of user interactions and method invocations to the core component. An Application Programming Interface (API) to the core component must be provided and documented. The TP must have the rights to make the targeted core component as well as the web-application available on a CLARIN server running at a dedicated centre.

The core component of the web-application must at least be able to operate on the research data and yield output in the formats agreed upon between user and TP. It is a pre if it can apply to other formats from the CLARIN standard list and yield additional output formats. The web-application and its core component will be used to obtain requirements and specifications of the architectural framework that is being worked out in CLARIN and may be used to test it.

A research data resource often consists of information of various kinds contained in multiple folders and multiple files of varying types. The information contained in such a resource can include documentation, source data, annotations of the source data, aggregate statistics tables on the source data and/or annotations, etc. Any tool (which eventually will very likely be integrated in the infrastructure as web service) should find out in a fully automated manner whether the research data selected by the user are appropriate input for it, and, if so, that it is applied to the right information (e.g. to the source data but not to the documentation). To achieve this it will integrate wrappers that read and write metadata and provenance information provided by the CLARIN infrastructure. A popular wrapper, used a lot and supported in CLARIN-NL is CLAM, developed by Tilburg University and currently

maintained by Radboud University Nijmegen.[22] Any requirements or desiderata that follow from this for metadata and data contents and formats should be properly documented in the documents with the requirements and desiderata for the CLARIN infrastructure.

The demonstrator consists, as a minimum, of a web application, the research data, and a demonstration scenario. A demonstration scenario is a detailed description of (sequences of) actions a user can take to have the application applied to the research data and the corresponding system responses in order to get a representative picture of the functionality offered. A movie or sequence of screen captures to illustrate the functionality is nice to have. The application will have to be installed on a CLARIN server, and the project proposal must contain a plan for doing this. It is the intention to have the demonstrator applications available for the lifetime of the CLARIN-NL project (2009-2015), so occasional support may be needed from the original developers even after the demonstrator project has finished.

The application must be tested with at least one of the common web browsers on the client side (MS IE, Firefox). Agreements about additional technical details (operating system, programming language, workspace requirements, etc) need to be made with the dedicated centre where the services should be executed.

Any vendor, platform or operating system dependent aspects of the application must be made explicit in the proposal and properly documented in the project.

The web-application and its core component should be properly documented, for users (user documentation), for application developers who want to use the core component (documentation of the API), and for technology developers who want to modify or extend the basic functionality of the application. The documentation, as well as the software user interface, must be provided at least in English.

Auxiliary Resources

The web-application may require data and other software (auxiliary resources) while running.

It must be documented which auxiliary data (e.g. a lexicon) and software (e.g. a library, converters) are needed during runtime for the application. The TP must have the right to make these auxiliary data and software available on a CLARIN server. Any restrictions on their usage (including costs) should be properly documented in the project proposal and in the documentation of the resulting application.

The application and the core component must be able to run on a dedicated CLARIN-server. An application or core component that can run only on a specific (non CLARIN) server (e.g. because it contains auxiliary resources that cannot be made available otherwise) is not acceptable.

Metadata

[22] http://ilk.uvt.nl/clam

For the web application and its core component, the research data and all runtime auxiliary data used in the application, metadata descriptions must be made in accordance with the CLARIN metadata standard (CMDI).[23] CMDI provides a flexible component-based framework for dealing with metadata; the data and tools of the projects of this call may require the development of new CMDI components or the adaptation of existing components and thus can contribute to the further development of the CMDI framework.[24] Any required or desirable extension or modifications of the CMDI framework must be properly documented and be included in the CLARIN Requirements and Desiderata document.

Requirements and Desiderata for CLARIN infrastructure

One important result of both demonstrator and curation projects is a document or series of documents describing requirements and desiderata for the CLARIN infrastructure resulting from the experiences gained with the curation of the research data and/or tools, and with the development of the application, its core component and web-services derived from it. These requirements and desiderata can concern many aspects. The following is a non-exhaustive list of aspects that should be considered:

- Requirements for data formats and encoding standards
- Web-service wrappers
- Metadata elements, components and profiles
- Processing requirements
- Memory requirements
- Network Bandwidth requirements
- User workspace requirements
- API requirements (e.g. Calling conventions)
- IPR / restricted use / ethical issues requirements
- Documentation requirements
- Repository Requirements
- Requirements for registering and resolving PIDs
- Requirements related to semantic interoperability

Evaluation Criteria

Proposals are elicited from humanities topics and disciplines, with a priority for the following (sub)disciplines:

1. Philosophy
2. Literary sciences
3. Cultural sciences
4. Communication & Media Studies, including Theatre sciences
5. Linguistics
6. History and political sciences

[23] http://www.clarin.eu/cmdi Login is required for editing.

[24] Existing components can be found in the Metadata Component Registry: http://catalog.clarin.eu/ds/ComponentRegistry/#

The order reflects the level of priority: the higher on the list, the higher the priority.

Projects involving multiple priority disciplines are especially encouraged.

Proposals on topics other that the listed priorities will be considered but must show their special relevance for the CLARIN infrastructure.

A proposal must describe a project that is compatible with the requirements mentioned in this call; in particular it must be a resource curation and/or a demonstrator project.

Proposals for projects will furthermore be evaluated and ranked according to the following more general criteria:

- **Quality**
  - Clarity and added value of the project proposal, in particular of the problem and the proposed approach
  - Suitability of the method and plan for the problem at hand
  - Feasibility of the project targets: can they be realized within the specified amount of time and with the instruments proposed?
  - Adequate balance between requested instruments and funds and proposed targets
  - Clearly specified and realistic work plan
  - Conformance to established standards and protocols as supported within CLARIN, or contribute to the development such standards and protocols.
- **Project Participants**
  - Competence of the participating partners (including their past performance);
  - Balanced cooperation and task assignments within the project. Justification of the composition of the team.
  - Availability of the infrastructure required for the project to be successful
  - Embedding of the work in other research programmes or projects, and/or additional funding from other funding sources is an advantage
- **User-orientation of the project**
  - Does the project address needs of the targeted infrastructure users (linguists and humanities researchers)?
  - Projects on tools or data that are widely in use in the targeted user community will score higher than projects focusing on lesser used tools and data.
  - More generic data or functionality, i.e. data or functionality that can serve multiple different research questions from linguistics and humanities research will be preferred over less generic or completely idiosyncratic functionality
  - Is there cooperation with or support from the targeted (future) infrastructure users?
  - Is the resulting tool / service user-friendly, i.e. will non-technical linguistic and humanities researchers be able to use it?
  - Is dissemination of the results to the targeted users and (where appropriate) training of them planned?
- **Contribution to CLARIN-NL as a whole**

  - Conformance to the goals of CLARIN-NL in particular and CLARIN in general and the priorities set within them
  - Contribution to knowledge transfer and network creation. In particular, cooperation between the intended users (linguists and humanities researchers) and technology and service providers (researchers in language and speech technology, computer science, etc.) is an advantage.
- **Intellectual Property Rights and Synergy**
  - Each proposal must contain clear statements about the situation of the IPR of the data and tools/technologies used, and a detailed plan to resolve any open issues.
  - The project participants have the obligation and must therefore have the rights to incorporate the core data and tools used in a project into the CLARIN infrastructure (this is a sine qua non). There has to be a clear specification and justification of the use of any data or tools needed in the project that cannot be incorporated into the CLARIN infrastructure.
  - Each proposal must show that the submitters have adequate and up-to-date knowledge of data, tools and services that are already available, so that any duplication of effort can be avoided.
- **Formal compliance**
  - A proposal must meet the formal requirements imposed by the CLARIN-NL organization for proposals, such as
    - conformance to the prescribed format and proposal template
    - submission before the set deadline, using the means prescribed
    - conformance to the prescribed language of the proposal

In addition, projects that are part of or fit in with international cooperation with partners from CLARIN in other countries will be preferred (of course, foreign organizations will have to find their own funding). Finally, a project that does not meet the IPR-requirements stated or is insufficiently clear about it will be considered formally noncompliant.

Duration

The duration of the project must be justified. The default maximum duration is 12 months. Any duration longer than 12 months requires thorough justification.

Budget

The project budget must be in accordance with the tasks to be carried out, and this must be justified in the project proposal. The maximum budget is 120k€.

Intellectual Property Rights (IPR)

Ownership of all original data and software remains with the original owners.

An agreement must be in place between the owners of the original data and software and the project participants on the IPR of the adapted data and software before the submission date of a proposal if the owners of the original data and software are not identical to the project participants. If applicable, a copy of this agreement must be uploaded together with the project proposal. Otherwise ownership of the created adaptations and extensions will be with the creator(s).

The project participants have the obligation and therefore must have the rights to make the research data, the application, its core component, and any runtime auxiliary data or software available on a CLARIN server for use by researchers having access to the CLARIN infrastructure. This is a *sine qua non*. Any proposal not satisfying this requirement or being insufficiently clear about this matter will be considered to be formally noncompliant and will be rejected on these grounds.

The project proposal should describe all issues related to IPR and present solutions for them. The relations between the partners in a project must be agreed upon in a consortium agreement before the start of the project.

Practical details
On the submission and evaluation procedure regarding the
CLARIN-NL Fourth Open call

Call for Proposals

The CLARIN-NL Fourth Open Call is open from **Monday July 2, 2012**. The total budget for this call is limited to a maximum of € 600,000. This call is specifically open for proposals targeting resource curation projects or demonstrator projects. Only those proposals that specifically target this priority will be eligible.

Full proposals must be submitted in English and in PDF format to the CLARIN-NL electronic proposal submission system using the prescribed template (which can be found on the CLARIN-NL website). The deadline for submitting full proposals in this call has been set for **Wednesday September 26, 2012 13:00 hours CET**.

The CLARIN-NL electronic proposal submission system can be accessed as of Monday July 2, 2012 via the CLARIN-NL website. If the electronic proposal submission system would not work, contact the CLARIN-NL Office clarinnl@uu.nl.

Who can apply?

Applications can be submitted only by researchers affiliated to CLARIN-NL participants that have signed the CLARIN-NL consortium agreement. The list of CLARIN-NL participants is available on the CLARIN-NL website. The main applicant and coordinator of the project must be a researcher from linguistics or humanities more broadly.

CLARIN-NL is in principle open to new participants. If your institute is not currently a CLARIN-NL participant, contact the CLARIN-NL office if you want your organization to become a participant in CLARIN-NL well in advance of the submission deadline.

Applicants who are planning to submit a proposal are strongly advised to contact the CLARIN-NL Office for an eligibility check of their plans, or for assistance in finding suitable partners.

Eligible costs

- Personnel costs directly related to the project, in accordance with the *Akkoord NWO-VSNU 2008* (and any additions to it).[25]
- A fee of maximally 3.000 € per FTE per year (or a pro rata part for less than 1 FTE per year) for covering travel and subsistence costs
- The requested funding cannot exceed 120,000€

[25] http://www.nwo.nl/nwohome.nsf/pages/NWOP_67QK4E more specifically http://www.nwo.nl/files.nsf/pages/NWOA_7LYGWY/$file/Ondertekende_Overeenkomst_NWO_VSNU.pdf . See also http://www.nwo.nl/files.nsf/pages/SPES_5VEDDR/$file/NWO-Regeling%20Subsidies%20per%201%20mei%202011.pdf for general guidelines that CLARIN-NL will follow where applicable.

Evaluation procedure full proposals

All eligible full proposals submitted in this call will be presented to a panel of international experts in the humanities, language and speech technology and infrastructures (International Advisory Panel, IAP). The composition of the IAP can be found on the CLARIN-NL website. If the proposals require this, the CLARIN-NL Executive Board can decide to involve additional experts in the evaluation.

This international panel will assess and rank the eligible applications based on the assessment criteria relevant to this call and may formulate a set of recommendations for improving the individual proposals. The IAP's assessment and ranking will not in this stage be sent to the proposers, but the IAP will, if needed, formulate questions and remarks to the project proposers. The project proposers will get the opportunity to answer these questions and comment on the remarks. The response must have been received by the CLARIN-NL office before the deadline set for it (one will typically have about a week to make the response). The IAP's assessment and recommendations will be presented, together with the response of the project proposers to the National Advisory Panel (NAP). The members of the NAP who are not directly involved in the submitted proposals will also assess the applications, in particular with regard to national considerations and based on the project proposers' feedback, and it make its own recommendations. On the basis of the IAP's advice and the NAP's advice, the CLARIN-NL Board will finally determine which projects will be funded. The coordinator of each project will receive a message on the final decision together with the NAP's and IAP's assessment reports on the project.

Projects should start within three months after the applicant has received the formal notification of funding.

CLARIN-NL *consortium agreement*

More information as to which legal rules apply for this specific CLARIN-NL granting scheme are laid down in the *CLARIN-NL consortium agreement*, which can be found on the CLARIN-NL website (http://www.clarin.nl/node/72).

Timetable

Activity	**Date**
CLARIN-NL Fourth Call Open	Monday July 2, 2012
CLARIN-NL Fourth Call Information Session	Thursday August 30, 2012 (afternoon)
Deadline Proposal Submission	Wednesday September 26, 2012 13:00hrs CET

Feedback / Questions from IAP	Monday November 12, 2012
Response to the IAP feedback /questions	one week after the feedback has been sent by e-mail (normally: Monday November 19, 2012)
Decision by the Board	Mid December 2012

The exact deadlines of the assessments by the IAP and NAP, and the final decision by the board will be communicated later via the CLARIN-NL website. The target is to make the final decisions before the end of 2012.

CLARIN-NL Organization

The CLARIN-NL project is funded by NWO.

The CLARIN-NL project is coordinated by the Programme Director, prof.dr. J.E.J.M. Odijk, who is a member of the CLARIN-NL Executive Board. Drs. Jolien Scholten is the CLARIN-NL project secretary.

The International Advisory Panel (IAP) is a group of international experts in the areas of humanities, in particular linguistics, language and speech technology, and infrastructures for scientific research.

The National Advisory Panel is a group of national researchers representative for the fields of linguistics and humanities, language and speech technologies and infrastructures for scientific research.

The CLARIN-NL Board consists of national senior researchers with great expertise in governance and/or relevant technical expertise

The composition of these CLARIN-NL governance bodies can be found on the CLARIN-NL website.

CLARIN-NL Helpdesk

Contact the CLARIN-NL Helpdesk for any technical questions related to this call, e.g.

- Questions on metadata and CMDI
- Questions on data categories en ISOCAT
- Questions on standards supported in CLARIN
- Questions on web services and CLAM.
- Etc.

At the end of this document, one will find a list of acronyms and terms for technical notions with some explanation and references.

Answers to Frequently Asked Questions are provided on the Helpdesk FAQ section

Website: http://trac.clarin.nl/trac
e-mail: helpdesk@clarin.nl

CLARIN-NL Centres

A list of the current candidate CLARIN centres and their contact persons can be found on the CLARIN-NL website: http://www.clarin.nl/node/130

CLARIN-NL Office

Contact the CLARIN-NL office for any organizational or practical questions related to this call, e.g.

- Additional clarification
- Advice on eligibility of your plans
- Assistance with finding experts, data or technology required
- Assistance with finding project partners
- Assistance with selecting a CLARIN Centre.
- Etc.

The CLARIN-NL office also offers other forms of help with writing a project proposal. In particular, it is wise to have the CLARIN-NL office carry out a sanity check on a preliminary version of the proposal, so that evidently ineligible proposals and trivial mistakes in the proposals can be avoided.

Jan Odijk/Jolien Scholten
Tel: +31 30 253 6279
Fax: +31 30 253 6000
Address: Trans 10, 3512 JK Utrecht
E-mail: clarinnl@uu.nl
CLARIN-NL Website: http://www,clarin,nl .

Term	Expansion	Explanation	URL
API	Application Programmers Interface	see there	Wikipedia
Application Application software		a piece of software developed to for a specific task and a specific user group with a user interface specifically designed for the targeted user group	Wikipedia
Application Programmer		software engineer that develops an application and usually accomplishes this by making (partly) use of other software	
Application Programmers Interface		an interface that allows a piece of software to use data, procedures and methods from another piece of software	Wikipedia
Application Programmers Interface documentation		documentation of the Application Programmers Interface aimed at application programmers	
CLARIN centre		organisation that has the ambition to become a recognized CLARIN centre. See http://www.clarin.nl/node/130 for more details about CLARIN centres in the Netherlands	
CLARIN server		server hosted by or on behalf of a CLARIN centre that runs services accessible for the (CLARIN) research community via the CLARIN infrastructure	
CLARIN-compliant		For a detailed description see http://trac.clarin.nl/trac/wiki/WikiStart#CLARIN-compatible. By curating a resource as described in this call, you make it CLARIN-	

		compliant	
CMDI	Component-based MetaData Infrastructure	the approach to metadata advocated in CLARIN. For details, see http://www.clarin.eu/cmdi	
CMDI component		A metadata component defined in accordance with CMDI.	
CMDI profile		A metadata profile defined in accordance with CMDI	
Component-based MetaData Infrastructure		the approach to metadata advocated in CLARIN . See http://www.clarin.eu/cmdi	
core component		(of a software application) the whole application minus the user interface	
curation		explained in section 5.1	
data category		a name and associated information to express a concept	
data category registry		a data-base or directory where data categories can be stored and viewed	
demonstration scenario		a detailed description of (sequences of) actions a user can take to have the application applied to the research data and the corresponding system responses in order to get a representative picture of the functionality offered (see p. 145). The demonstration scenario should serve as an example of how to use the application and/or solve a particular type of problem.	
demonstrator		explained in section 5.2, p. 145, second paragraph	

Ethical Issues		Issues related to ethics that may arise by making data available to a wider audience, e.g. privacy violations etc.	
Formal description		a formal description is a description in a form defined by well-specified rules that make it possible to completely and unambiguously determine the meaning of the description from the form	
infrastructure		(in the CLARIN context) a combination of hardware and software that allows a researcher or research team to find and use language resources, apply tools to language resources, and store own data and metadata to make them available to the research community.	
Intellectual Property		a variety of intangible assets produced by intellectual labor, such as musical, literary, and artistic works; discoveries and inventions; and words, phrases, symbols, and designs	Wikipedia
Intellectual Property Rights		All the rights and restrictions that are associated with intellectual property: who is the owner, is the resource licensed, what is one allowed to do with a resource, etc etc. Common types of intellectual property rights include copyrights, trademarks, patents, etc.	Wikipedia
interoperability		resources (data and software tools) are interoperable when they can work together in a fully automated manner with minimal human intervention	
IPR	Intellectual Property Rights	see there	

language resource		data containing language or a software tool operating on or yielding language data	
metadata		Literally: data about data; in CLARIN usually restricted to descriptions of language resources (in CMDI format)	
metadata component		A metadata component is an XML structure containing metadata elements and (recursively) other metadata components; it is used to describe a related set of properties of a language resource.	
metadata harvesting		the process of obtaining metadata from a different server (metadata provider) over a network (e.g. to get these metadata available in central data catalog)	
metadata profile		A metadata profile is an XML document definition consisting of metadata components and is used to describe the properties of a language resource	
method		(in programming languages) a special type of procedure or function	
method invocation		(in programming languages) calling a procedure or function	
registry		(digital) data-base or directory to store and access information	
resource		used in this document mostly as equivalent to language resource	
resource curation		see: curation	

Service		(here used in a software context) a software program that provides certain functionality and communicates its input and output with other software programs	
Service Oriented Architecture		Software architecture that is designed around (distributed) services interacting with each other to provide certain functionality	
web application		an application with a web-based user interface	
Web service		service that communicates with other programs over a network (in particular, the internet)	
wrapper (function)		is a software component whose main purpose is to make the functions of a second software component available and does little or no additional computation	Wikipedia
XML	eXtensible Markup Language	is a markup language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable.	Wikipedia
XML element		a particular constituent in a document with XML markup	W3CSchools
XML Schema		A specific way of defining the structure, content and semantics of XML documents	W3C

CLARIN-NL Fourth Call: Closed Call

CLARIN-NL launches in its fourth call a Closed Call for project proposals. **This called is only open for researchers who have been explicitly invited to submit a project proposal for the closed call by the CLARIN-NL Board**. Information on the call will be available as of Monday July 2, 2012 and proposals can be submitted as of that data. The total budget for this closed call is limited to a maximum of € 400,000. The call is specifically open for proposals targeting resource curation projects or demonstrator projects. Such projects take existing data or tools as a basis and attempt to apply CLARIN-supported standards and best practices to make the data and tools CLARIN-compliant. Examples of such projects awarded in 2009 through 2011 can be found on the CLARIN-NL website.[26] Only those proposals that specifically target this priority will be eligible. The maximum budget per project is 80,000€.

Important Note It is an essential requirement that a CLARIN centre is involved in each resource curation or demonstrator project. It is therefore recommended to contact a CLARIN-centre as early as possible while preparing a project proposal. The list of CLARIN centres, the kinds of resources they focus on, and their contact persons can be found on the CLARIN web site, more specifically here.[27] Some aspects of this call text are technical in nature: the CLARIN centre will, inter alia, be able to explain and discuss these technical matters with you. At the end of this document, one will find a list of acronyms and terms for technical notions with some explanation and references.

Introduction

The first phase of the CLARIN-NL project focuses on the specification and design of the infrastructure. Obviously, the CLARIN infrastructure should be designed in such a way that it can incorporate the data and tools currently used by humanities researchers to address their research questions. On the other hand, the CLARIN infrastructure can only be successful if these data and tools meet certain requirements with regard to standards and interoperability, not only with regard to the data and tools themselves but also with regard to their visibility and accessibility. Many of the data and tools currently in use do not meet these requirements, or meet them only partially. One of the aims of the fourth call is to make existing data and tools compliant with the requirements on standards and interoperability currently advocated in CLARIN.

[26] http://www.clarin.nl/node/76

[27] http://www.clarin.nl/node/130

However, the list of standards, best practices and interoperability requirements currently advocated by CLARIN has not been completely fixed yet. There are good reasons for this: (1) there may be crucial data or tools for which none of the currently advocated standards or best practices is suited; (2) the currently advocated standards and best practices may be incomplete, insufficiently specific, inconvenient or even incompatible with crucial data and tools. Therefore, a second aim of the fourth call is to test whether and to which extent the currently existing data and tools can be made compatible with the standards and best practices currently advocated in CLARIN, and to get a good overview of any incompatibilities as well as suggestions for adaptations of these standards.

These matters can best be investigated by actually attempting the necessary conversions and adaptations. When applied to tools and other software this will result in CLARIN-compliant web applications that can also serve as showcases of the kind of functionality CLARIN aims to offer. A third aim of the fourth call is therefore to make sure that the applications developed become available in demonstrators which can help promote the particular applications but also the CLARIN infrastructure as a whole.

Finally, by investigating these matters using specific data and tools that are currently in use a lot can be learnt about requirements the CLARIN infrastructure should meet or desiderata that it should offer. The fourth aim of this call is therefore obtaining a detailed list of such requirements and desiderata.

Project Types

In the fourth call, CLARIN-NL therefore solicits (1) projects that carry out resource (data or tools) curation, and (2) demonstrator projects.

Curation projects

CLARIN-NL aims to support the curation of digital language resources (data and tools) so that these resources can participate in the CLARIN infrastructure, more easily be accessed by interested researchers via online methods and become part of appealing new applications. Many language resources are neither visible nor accessible. Visibility is mainly achieved by standardized metadata that are being harvested by service providers. Accessibility has many different aspects:

10. The resource needs to be stored at computers that are accessible via the internet.
11. The resource needs to be identified in a persistent manner.
12. The resource needs to be interpretable, which requires a format that adheres to best practices and it requires references to registries were the used concepts are defined.

Demonstrator projects

CLARIN-NL aims to support projects that create appealing showcases of functionality that the CLARIN infrastructure should support. Such projects should make available web applications that can be used as demonstrators of functionality that supports addressing research questions of the CLARIN-NL intended user group. The development of these demonstrators will also be used to inventory a list of requirements the CLARIN infrastructure should meet and desiderata it preferably should offer.

Goals

The goal of a **curation project** is:

- Adapting specific resources so that they are visible, uniquely referable and accessible via the web, and properly documented.

The goal of a **demonstrator project** is:

- Creating a documented web application starting from an existing tool or application that can be used as a demonstrator and function as a showcase of the type of functionality CLARIN will incorporate and support. Within the web application there must be a clear separation between the web-based user interface and the core component. For the latter an programming language API must be defined and documented.

Important goals **common** to both types of projects are:

- Applying standards and best practices and makes use of the suggested CLARIN architecture and agreements to understand their limitations and the requirements for extensions.
- Establishing requirements and desiderata for the CLARIN infrastructure

In both project types the use of CLARIN-supported standards and best practices is essential.

The selection of CLARIN-supported standards and best practices is currently ongoing and it is the intention that the projects supported in this call contribute to this. A preliminary set of candidates for CLARIN-supported standards and best practices is available,[28] and though it is by no means final we will refer to this list in this document by the term “CLARIN-standards”.

Roles

Four roles of persons involved in the projects can be distinguished: the *user*, the *data provider* (DP), the *technology provider (TP)*, and the *infrastructure specialist (IS)*. The *user* is a researcher from a linguistics or humanities institute who aims to investigate one or more specific research questions. The DP has a certain set of digital language-related data at his/her disposal (the ‘research data’) that can be used for addressing the research questions

[28] http://www.clarin.eu/recommendations

of the user. The TP has a certain technology (e.g. language or speech technology) at its disposal and a thorough understanding of this technology (e.g. because it was developed by the TP). This technology offers functionality that makes it possible to (better) address the user's research question by applying this functionality to the research data. The *infrastructure specialist* (IS) is a specialist who has a deep understanding of the CLARIN service-oriented architecture and its requirements, and/or a specialist in data and tool format standards supported in CLARIN. IS specialists usually get involved in the project via the *CLARIN-NL Helpdesk*.[29] In many cases the different roles of user, DP, TP and IS will be played by persons from different organizations, but they may originate from a single organization, and occasionally even be played by a single person. CLARIN-NL can offer assistance in bringing the right experts together, if desired.

User

The *user* is a researcher from a linguistics or humanities institute who aims to investigate one or more specific research questions. The project proposal should clearly describe the research question(s) of the user, and the research question(s) must be in the domain of research in the humanities in general and the study of language in particular. The project must be led by the user.

Data Provider

The DP has a certain set of digital language-related data at his/her disposal (the 'research data') that can be used for addressing the research questions of the user. The project proposal should clearly describe the research data the DP has at his/her disposal that can be used to address the research question(s), and how they can be used for this purpose. The research data must be existing digital language or language-related data. No new research data should be created in the project. The DP must have the right to make the research data available on a CLARIN server running at a dedicated centre. If the data are in a format that is not currently on the list of CLARIN standards, a resource curation project is in order. Otherwise, the data can be used in a demonstrator project. The project proposal should contain a detailed description of the research data, its current state and format, the plans to convert it if needed, justification for using different formats if applicable, and a detailed plan for dealing with the data and its metadata (see below). Any restrictions on the use of the data as well as any ethical issues that apply or may arise must be properly documented in the proposal.

The Technology Provider

The TP has a certain technology at his/her disposal that can be used as a basis for the development of a web-based application (possibly web-services based) and concomitant demonstrator, or that can be used for resource curation. Since the research data are language data, the technology will in most cases be language or speech technology.

[29] helpdesk@clarin.nl

The proposal should contain a detailed description of the available technology and its current status. It should make clear that the TP has a thorough understanding of this technology and describe how the TP obtained this understanding (e.g. because the TP developed the technology).

The intended use of the technology in the project should be described, as well as any extensions or modifications that have to be made to the technology in the project, and a plan to achieve this.

The TP must have the right to use this technology and indicate how it will be used in the project.

CLARIN-NL Helpdesk and Infrastructure Specialist

CLARIN-NL has set up a *Helpdesk* that project participants can turn to for all kinds of technical questions related to their project. If the relevant questions cannot be addressed by the *Helpdesk* functionality (FAQ section, etc.) or staff, the Helpdesk will involve infrastructure specialists. The infrastructure specialist (IS) is a specialist who has a deep understanding of the CLARIN service-oriented architecture and its requirements, and/or a specialist in data, metadata and tool format standards and best practices supported in CLARIN. The *Helpdesk*, and where needed IS specialists, will advise and assist the project partners.

Project Types

Resource Curation

Resource curation involves a number of different aspects:

16. The resource should be brought into a format that adheres to widely accepted standards and best practices currently considered as likely candidates by CLARIN.
17. Proper metadata descriptions need to be created and made available. They must be compliant with the CLARIN component metadata infrastructure (CMDI) and it should be possible to harvest and access them.
18. Metadata descriptions should include persistent identifiers that can be resolved and the CLARIN requirements should hold for the PID system.
19. All data categories used in the metadata and in the actual data (e.g. linguistic annotations) must be related to a CLARIN-recognized data category registry (currently only ISOCAT[30]), i.e. data categories used must be mapped to corresponding ISOCAT data categories where they exist in a formal way (e.g. via an XML Schema) and new data categories must be added to ISOCAT if they do not exist there yet.
20. Provide proper documentation of the resource, at least in English.

The plan for a curation project should describe in detail how these different aspects are going to be addressed in the project.

[30] http://www.isocat.org/

The results of these aspects should be tested by the project participants. Setting up tests for this should be included in the project plan and the results of these tests will be included in the project's success criteria. Example tests are e.g. a metadata harvesting test and formal procedures such as testing against an XML Schema.

The resulting resource and its metadata must be made available on a server of a recognized CLARIN centre. The project proposal must specify which (candidate) CLARIN centre this will be and concrete arrangements must have been made with this centre.

Demonstrator

In a demonstrator project a demonstrator is developed using a documented web-based application based on a technology that the TP currently has at his/her disposal. The development is carried out in close cooperation with the user

The project proposal must contain a detailed description of the targeted functionality, including input and output specifications, and how it can contribute to addressing the user's research questions. More generic functionality, i.e. functionality that can serve multiple different research questions from linguistics and humanities research will be preferred over less generic or completely idiosyncratic functionality. See below for additional criteria related to the functionality that will be used to rank proposals. Since a demonstrator project is short in duration and is relatively small, this functionality must already be available to the TP, though perhaps not in the form of a web-based application, and it perhaps only operates on data formats other than the ones listed in the CLARIN standards. The project proposal should contain a detailed description of the functionality in its current state, the targeted web-based application and its components, and a plan to achieve this. The application includes a web-based user interface that takes care of user interactions and method invocations to the core component. An Application Programming Interface (API) to the core component must be provided and documented. The TP must have the rights to make the targeted core component as well as the web-application available on a CLARIN server running at a dedicated centre.

The core component of the web-application must at least be able to operate on the research data and yield output in the formats agreed upon between user and TP. It is a pre if it can apply to other formats from the CLARIN standard list and yield additional output formats. The web-application and its core component will be used to obtain requirements and specifications of the architectural framework that is being worked out in CLARIN and may be used to test it.

A research data resource often consists of information of various kinds contained in multiple folders and multiple files of varying types. The information contained in such a resource can include documentation, source data, annotations of the source data, aggregate statistics tables on the source data and/or annotations, etc. Any tool (which eventually will very likely be integrated in the infrastructure as web service) should find out in a fully automated manner whether the research data selected by the user are appropriate input for it, and, if so, that it is applied to the right information (e.g. to the source data but not to the

documentation). To achieve this it will integrate wrappers that read and write metadata and provenance information provided by the CLARIN infrastructure. A popular wrapper, used a lot and supported in CLARIN-NL is CLAM, developed by Tilburg University and currently maintained by Radboud University Nijmegen.[31] Any requirements or desiderata that follow from this for metadata and data contents and formats should be properly documented in the documents with the requirements and desiderata for the CLARIN infrastructure.

The demonstrator consists, as a minimum, of a web application, the research data, and a demonstration scenario. A demonstration scenario is a detailed description of (sequences of) actions a user can take to have the application applied to the research data and the corresponding system responses in order to get a representative picture of the functionality offered. A movie or sequence of screen captures to illustrate the functionality is nice to have. The application will have to be installed on a CLARIN server, and the project proposal must contain a plan for doing this. It is the intention to have the demonstrator applications available for the lifetime of the CLARIN-NL project (2009-2014), so occasional support may be needed from the original developers even after the demonstrator project has finished.

The application must be tested with at least one of the common web browsers on the client side (MS IE, Firefox). Agreements about additional technical details (operating system, programming language, workspace requirements, etc) need to be made with the dedicated centre where the services should be executed.

Any vendor, platform or operating system dependent aspects of the application must be made explicit in the proposal and properly documented in the project.

The web-application and its core component should be properly documented, for users (user documentation), for application developers who want to use the core component (documentation of the API), and for technology developers who want to modify or extend the basic functionality of the application. The documentation, as well as the software user interface, must be provided at least in English.

Auxiliary Resources

The web-application may require data and other software (auxiliary resources) while running.

It must be documented which auxiliary data (e.g. a lexicon) and software (e.g. a library, converters) are needed during runtime for the application. The TP must have the right to make these auxiliary data and software available on a CLARIN server. Any restrictions on their usage (including costs) should be properly documented in the project proposal and in the documentation of the resulting application.

The application and the core component must be able to run on a dedicated CLARIN-server. An application or core component that can run only on a specific (non CLARIN) server (e.g.

[31] http://ilk.uvt.nl/clam

because it contains auxiliary resources that cannot be made available otherwise) is not acceptable.

Metadata

For the web application and its core component, the research data and all runtime auxiliary data used in the application, metadata descriptions must be made in accordance with the CLARIN metadata standard (CMDI).[32] CMDI provides a flexible component-based framework for dealing with metadata; the data and tools of the projects of this call may require the development of new CMDI components or the adaptation of existing components and thus can contribute to the further development of the CMDI framework.[33] Any required or desirable extension or modifications of the CMDI framework must be properly documented and be included in the CLARIN Requirements and Desiderata document.

Requirements and Desiderata for CLARIN infrastructure

One important result of both demonstrator and curation projects is a document or series of documents describing requirements and desiderata for the CLARIN infrastructure resulting from the experiences gained with the curation of the research data and/or tools, and with the development of the application, its core component and web-services derived from it. These requirements and desiderata can concern many aspects. The following is a non-exhaustive list of aspects that should be considered:

- Requirements for data formats and encoding standards
- Web-service wrappers
- Metadata elements, components and profiles
- Processing requirements
- Memory requirements
- Network Bandwidth requirements
- User workspace requirements
- API requirements (e.g. Calling conventions)
- IPR / restricted use / ethical issues requirements
- Documentation requirements
- Repository Requirements
- Requirements for registering and resolving PIDs
- Requirements related to semantic interoperability

Evaluation Criteria

A proposal submitted in the Closed Call will not compete with other proposals. Each proposal is evaluated in accordance with the criteria described here, and if this evaluation is positive, the project will be funded.

[32] http://www.clarin.eu/cmdi Login is required for editing.

[33] Existing components can be found in the Metadata Component Registry: http://catalog.clarin.eu/ds/ComponentRegistry/#

A proposal must describe a project that is compatible with the requirements mentioned in this call; in particular it must be a resource curation and/or a demonstrator project.

Proposals for projects will furthermore be evaluated according to the following more general criteria:

- **Quality**
  - Clarity and added value of the project proposal, in particular of the problem and the proposed approach
  - Suitability of the method and plan for the problem at hand
  - Feasibility of the project targets: can they be realized within the specified amount of time and with the instruments proposed?
  - Adequate balance between requested instruments and funds and proposed targets
  - Clearly specified and realistic work plan
  - Conformance to established standards and protocols as supported within CLARIN, or contribute to the development such standards and protocols.
- **Project Participants**
  - Competence of the participating partners (including their past performance);
  - Balanced cooperation and task assignments within the project. Justification of the composition of the team.
  - Availability of the infrastructure required for the project to be successful
  - Embedding of the work in other research programmes or projects, and/or additional funding from other funding sources is an advantage
- **User-orientation of the project**
  - Does the project address needs of the targeted infrastructure users (linguists and humanities researchers)?
  - Projects on tools or data that are widely in use in the targeted user community will score higher than projects focusing on lesser used tools and data.
  - More generic data or functionality, i.e. data or functionality that can serve multiple different research questions from linguistics and humanities research will be preferred over less generic or completely idiosyncratic functionality.
  - Is there cooperation with or support from the targeted (future) infrastructure users?
  - Is the resulting tool / service user-friendly, i.e. will non-technical linguistic and humanities researchers be able to use it?
  - Is dissemination of the results to the targeted users and (where appropriate) training of them planned?
- **Contribution to CLARIN-NL as a whole**
  - Conformance to the goals of CLARIN-NL in particular and CLARIN in general and the priorities set within them
  - Contribution to knowledge transfer and network creation. In particular, cooperation between the intended users (linguists and humanities researchers) and technology and service providers (researchers in language and speech technology, computer science, etc.) is an advantage.
- **Intellectual Property Rights and Synergy**

  - Each proposal must contain clear statements about the situation of the IPR of the data and tools/technologies used, and a detailed plan to resolve any open issues.
  - The project participants have the obligation and must therefore have the rights to incorporate the core data and tools used in a project into the CLARIN infrastructure (this is a sine qua non). There has to be a clear specification and justification of the use of any data or tools needed in the project that cannot be incorporated into the CLARIN infrastructure.
  - Each proposal must show that the submitters have adequate and up-to-date knowledge of data, tools and services that are already available, so that any duplication of effort can be avoided.
- **Formal compliance**
  - A proposal must meet the formal requirements imposed by the CLARIN-NL organization for proposals, such as
    - conformance to the prescribed format and proposal template
    - submission before the set deadline, using the means prescribed
    - conformance to the prescribed language of the proposal

In addition, projects that are part of or fit in with international cooperation with partners from CLARIN in other countries will be preferred (of course, foreign organizations will have to find their own funding). Finally, a project that does not meet the IPR-requirements stated or is insufficiently clear about it will be considered formally noncompliant.

Duration

The duration of the project must be justified. The default maximum duration is 12 months. Any duration longer than 12 months requires thorough justification.

Budget

The project budget must be in accordance with the tasks to be carried out, and this must be justified in the project proposal. The maximum budget is 80k€.

Intellectual Property Rights (IPR)

Ownership of all original data and software remains with the original owners.

An agreement must be in place between the owners of the original data and software and the project participants on the IPR of the adapted data and software before the submission date of a proposal if the owners of the original data and software are not identical to the project participants. If applicable, a copy of this agreement must be uploaded together with the project proposal. Otherwise ownership of the created adaptations and extensions will be with the creator(s).

The project participants have the obligation and therefore must have the rights to make the research data, the application, its core component, and any runtime auxiliary data or software available on a CLARIN server for use by researchers having access to the CLARIN infrastructure. This is a *sine qua non*. Any proposal not satisfying this requirement or being

insufficiently clear about this matter will be considered to be formally noncompliant and will be rejected on these grounds.

The project proposal should describe all issues related to IPR and present solutions for them. The relations between the partners in a project must be agreed upon in a consortium agreement before the start of the project.

Practical details
On the submission and evaluation procedure regarding the
CLARIN-NL Fourth Call: Closed Call

Call for Proposals

Information on the CLARIN-NL Fourth Call Closed Call is available as of Monday July 2, 2012, and proposals can be submitted as of that date. The total budget for this call is limited to a maximum of € 400,000. This call is specifically open for proposals targeting resource curation projects or demonstrator projects. Only those proposals that specifically target this priority will be eligible.

Full proposals must be submitted in English and in PDF format to the CLARIN-NL electronic proposal submission system using the prescribed template (which can be found on the CLARIN-NL website). The deadline for submitting full proposals in this call has been set for **Wednesday Sep 26, 2012 13:00 hours CET**.

The CLARIN-NL electronic proposal submission system can be accessed as of **Monday July 2, 2012** via the CLARIN-NL website. If the electronic proposal submission system would not work, contact the CLARIN-NL Office clarinnl@uu.nl.

Who can apply?

Applications can be submitted only by researchers who have been explicitly invited to submit a project proposal for the closed call by the CLARIN-NL Board.

Eligible costs

- Personnel costs directly related to the project, in accordance with the *Akkoord NWO-VSNU 2008* (and any additions to it).[34]
- A fee of maximally 3.000 € per FTE per year (or a pro rata part for less than 1 FTE per year) for covering travel and subsistence costs
- The requested funding cannot exceed 80,000€

Evaluation procedure full proposals

All eligible full proposals submitted in this call will be presented to a panel of international experts in the humanities, language and speech technology and infrastructures (International Advisory Panel, IAP). The composition of the IAP can be found on the CLARIN-NL website. If the proposals require this, the CLARIN-NL Executive Board can decide to involve additional experts in the evaluation.

[34] http://www.nwo.nl/nwohome.nsf/pages/NWOP_67QK4E more specifically http://www.nwo.nl/files.nsf/pages/NWOA_7LYGWY/$file/Ondertekende_Overeenkomst_NWO_VSNU.pdf . See also http://www.nwo.nl/files.nsf/pages/SPES_5VEDDR/$file/NWO-Regeling%20Subsidies%20per%201%20mei%202011.pdf for general guidelines that CLARIN-NL will follow where applicable.

This international panel will assess each eligible application based on the assessment criteria relevant to this call and may formulate a set of recommendations for improving the individual proposals. The IAP's assessment will not in this stage be sent to the proposers, but the IAP will, if needed, formulate questions and remarks for the project proposers. The project proposers will get the opportunity to answer these questions and to comment on the remarks. The response must have been received by the CLARIN-NL office before the deadline set for it (one will typically have about a week to make the response). The IAP's assessment and recommendations will be presented, together with the response of the project proposers to the National Advisory Panel (NAP). The members of the NAP who are not directly involved in the submitted proposals will also assess each application and the IAP's assessment and recommendations. On the basis of the IAP's advice and the NAP's advice, the CLARIN-NL Board will finally determine whether the project will be funded. The coordinator of each project will receive a message on the final decision together with the NAP's and IAP's assessment reports on the project.

Projects should start within three months after the applicant has received the formal notification of funding.

CLARIN-NL *consortium agreement*

More information as to which legal rules apply for this specific CLARIN-NL granting scheme are laid down in the *CLARIN-NL consortium agreement*, which can be found on the CLARIN-NL website (http://www.clarin.nl/node/72).

Timetable

Activity	**Date**
CLARIN-NL Fourth Call Open	Monday July 2, 2012
CLARIN-NL Fourth Call Information Session	Thursday August 30, 2012 (afternoon)
Deadline Proposal Submission	Wednesday September 26, 2012 13:00hrs CET
Feedback / Questions from IAP	Monday November 12, 2012
Response to the IAP feedback /questions	one week after the feedback has been sent by e-mail (normally: Monday November 19, 2012)
Decision by the Board	Mid December 2012

The exact date of the final decision by the board will be communicated later via the CLARIN-NL website.

CLARIN-NL Organization

The CLARIN-NL project is funded by NWO.

The CLARIN-NL project is coordinated by the Programme Director, prof.dr. J.E.J.M. Odijk, who is a member of the CLARIN-NL Executive Board. Drs. Jolien Scholten is the CLARIN-NL project secretary.

The International Advisory Panel (IAP) is a group of international experts in the areas of humanities, in particular linguistics, language and speech technology, and infrastructures for scientific research.

The National Advisory Panel is a group of national researchers representative for the fields of linguistics and humanities, language and speech technologies and infrastructures for scientific research.

The CLARIN-NL Board consists of national senior researchers with great expertise in governance and/or relevant technical expertise

The composition of these CLARIN-NL governance bodies can be found on the CLARIN-NL website.

CLARIN-NL Helpdesk

Contact the CLARIN-NL Helpdesk for any technical questions related to this call, e.g.

- Questions on metadata and CMDI
- Questions on data categories en ISOCAT
- Questions on standards supported in CLARIN
- Questions on web services and CLAM.
- Etc.

At the end of this document, one will find a list of acronyms and terms for technical notions with some explanation and references.
Answers to Frequently Asked Questions are provided on the Helpdesk FAQ section

Website: http://trac.clarin.nl/trac
e-mail: helpdesk@clarin.nl

CLARIN-NL Centres

A list of the current candidate CLARIN centres and their contact persons can be found on the CLARIN-NL website: http://www.clarin.nl/node/130

CLARIN-NL Office

Contact the CLARIN-NL office for any organizational or practical questions related to this call, e.g.

- Additional clarification

- Advice on eligibility of your plans
- Assistance with finding experts, data or technology required
- Assistance with finding project partners
- Assistance with selecting a CLARIN Centre.
- Etc.

The CLARIN-NL office also offers other forms of help with writing a project proposal. In particular, it is wise to have the CLARIN-NL office carry out a sanity check on a preliminary version of the proposal, so that evidently ineligible proposals and trivial mistakes in the proposals can be avoided.

Jan Odijk/Jolien Scholten
Tel: +31 30 253 6279
Fax: +31 30 253 6000
Address: Trans 10, 3512 JK Utrecht
E-mail: clarinnl@uu.nl
CLARIN-NL Website: http://www,clarin,nl .

Term	Expansion	Explanation	URL
API	Application Programmers Interface	see there	Wikipedia
Application Application software		a piece of software developed to for a specific task and a specific user group with a user interface specifically designed for the targeted user group	Wikipedia
Application Programmer		software engineer that develops an application and usually accomplishes this by making (partly) use of other software	
Application Programmers Interface		an interface that allows a piece of software to use data, procedures and methods from another piece of software	Wikipedia
Application Programmers Interface documentation		documentation of the Application Programmers Interface aimed at application programmers	
CLARIN centre		organisation that has the ambition to become a recognized CLARIN centre. See http://www.clarin.nl/node/130 for more details about CLARIN centres in the Netherlands	
CLARIN server		server hosted by or on behalf of a CLARIN centre that runs services accessible for the (CLARIN) research community via the CLARIN infrastructure	
CLARIN-compliant		For a detailed description see http://trac.clarin.nl/trac/wiki/WikiStart#CLARIN-compatible. By curating a resource as described in this call, you make it CLARIN-compliant	
CMDI	Component-based MetaData	the approach to metadata advocated in CLARIN. For details, see	

	Infrastructure	http://www.clarin.eu/cmdi	
CMDI component		A metadata component defined in accordance with CMDI.	
CMDI profile		A metadata profile defined in accordance with CMDI	
Component-based MetaData Infrastructure		the approach to metadata advocated in CLARIN . See http://www.clarin.eu/cmdi	
core component		(of a software application) the whole application minus the user interface	
curation		explained in section 5.1	
data category		a name and associated information to express a concept	
data category registry		a data-base or directory where data categories can be stored and viewed	
demonstration scenario		a detailed description of (sequences of) actions a user can take to have the application applied to the research data and the corresponding system responses in order to get a representative picture of the functionality offered (see p. 145). The demonstration scenario should serve as an example of how to use the application and/or solve a particular type of problem.	
demonstrator		explained in section 5.2, p. 145, second paragraph	
Ethical Issues		Issues related to ethics that may arise by making data available to a wider audience, e.g. privacy violations etc.	
Formal description		a formal description is a description in a form defined by well-specified rules that make it possible to completely and unambiguously determine the meaning of the	

		description from the form	
infrastructure		(in the CLARIN context) a combination of hardware and software that allows a researcher or research team to find and use language resources, apply tools to language resources, and store own data and metadata to make them available to the research community.	
Intellectual Property		a variety of intangible assets produced by intellectual labor, such as musical, literary, and artistic works; discoveries and inventions; and words, phrases, symbols, and designs	Wikipedia
Intellectual Property Rights		All the rights and restrictions that are associated with intellectual property: who is the owner, is the resource licensed, what is one allowed to do with a resource, etc etc. Common types of intellectual property rights include copyrights, trademarks, patents, etc.	Wikipedia
interoperability		resources (data and software tools) are interoperable when they can work together in a fully automated manner with minimal human intervention	
IPR	Intellectual Property Rights	see there	
language resource		data containing language or a software tool operating on or yielding language data	
metadata		Literally: data about data; in CLARIN usually restricted to descriptions of language resources (in CMDI format)	

metadata component		A metadata component is an XML structure containing metadata elements and (recursively) other metadata components; it is used to describe a related set of properties of a language resource.	
metadata harvesting		the process of obtaining metadata from a different server (metadata provider) over a network (e.g. to get these metadata available in central data catalog)	
metadata profile		A metadata profile is an XML document definition consisting of metadata components and is used to describe the properties of a language resource	
method		(in programming languages) a special type of procedure or function	
method invocation		(in programming languages) calling a procedure or function	
registry		(digital) data-base or directory to store and access information	
resource		used in this document mostly as equivalent to language resource	
resource curation		see: curation	
Service		(here used in a software context) a software program that provides certain functionality and communicates its input and output with other software programs	
Service Oriented Architecture		Software architecture that is designed around (distributed) services interacting with each other to provide certain functionality	
web application		an application with a web-based user interface	

Web service		service that communicates with other programs over a network (in particular, the internet)	
wrapper (function)		is a software component whose main purpose is to make the functions of a second software component available and does little or no additional computation	Wikipedia
XML	eXtensible Markup Language	is a markup language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable.	Wikipedia
XML element		a particular constituent in a document with XML markup	W3CSchools
XML Schema		A specific way of defining the structure, content and semantics of XML documents	W3C

Projects

Call 1

10-001 CKCC

Geleerdenbrievenproject - Circulation of Knowledge and Learned Practices in the 17th-century Dutch Republic

2010-02-01 – 2010-11-01

Project coordinator: Prof. dr. Wijnand Mijnhardt (Descartes Centre)

http://ckcc.huygens.knaw.nl/

Participant	Budget
Huygens ING	€59,250
Descartes Centre	€0
Total	**€59,250**

Abstract

A consortium of Dutch universities and cultural heritage institutions is building a web-based collaboratory (an online space for asynchronous collaboration) around a corpus of 20.000 letters of scholars who lived in the 17th-century Dutch Republic to answer the research question: how did knowledge circulate in the 17th century? Hereto, it will be necessary to analyze this large amount of correspondence systematically. Based on this (extendable) corpus, we will implement a content processing workflow that consists of iterative cycles of conceptual analysis, enrichment with several layers of annotation and visualization.

With advice from CLARIN-EU in the first stage of the project a demonstrator will be developed which implements techniques of keyword extraction (deadline: 1 October 2010) The second stage consists of evaluating existing more complex tools en techniques that can tackle one or more aspects of the targeted grammatical, content-related, and network complexity analysis, annotation, and visualization. The phase shall identify a set of tools that can be readily utilized in CKCC, as well as tools that need to be adapted or extended to the needs of CKCC; in short, by the end of this phase resources, requirements and risks shall become clear (deadline: December 2010).

In the third stage the collaboratory will be further developed according to the description in the CKCC project goals, centering around the technique of concept extraction (Deadline: 1 November 2012). These three stages constitute the Work Package Analysis Tools, the core of the CKCC project, for which the support of CLARIN-NL is requested. Other Work Packages provide data and software tools needed to create a complete system: the digital corpus of letters (WP6), the editing collaboratory that will contain the letters (WP1), and the archiving environment for data and software (WP2).

09-002 MIMORE

Microcomparative Morphosyntax Research Tool

2010-01-02 – 2010-06-30

Project coordinator: prof. dr. Sjef Barbiers (Meertens Institute)

http://www.meertens.knaw.nl/mimore/

Participant	Budget
Utrecht University (UiL-OTS)	€0

Meertens Institute	€54,952
Total	**€54,952**

Abstract

The demonstrator tool MIMORE will be based on three databases: (i) DynaSAND, a corpus of elicited speech and text collected between 2000-2005 to hart the syntactic variation at the clausal level in 267 dialects of Dutch spoken in the Netherlands, Belgium and North-West France; (ii) DiDDD, a corpus of elicited speech and text collected between 2005-2009 to chart the syntactic variation at the level of nominal groups in the same language area; (iii) MAND a corpus of elicited speech and text collected between 1980 and 1995 to chart morphological (word-level) variation. In the proposed tool the three databases will have a common search engine that makes it possible to investigate potential correlations between variables at the three different linguistic levels, cartographic functionality enabling the user to visualize these correlations and statistical functionality to analyse them.

09-003 SignLinC

Linking lexical databases and annotated corpora of signed Languages

2010-01-01 – 2011-01-01

Project coordinator: Dr. O.A. Crasborn (Radboud University Nijmegen)

http://www.ru.nl/sign-lang/projects/completed-projects/signlinc/

Participant	**Budget**
Nederlands Gebarencentrum	€10,318
Radboud University	€27,188
Max Planck Institute for Psycholinguistics	€18,825
Total	**€56,331**

Abstract

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. The first is a corpus of video texts that is already fully compliant with CLARIN standards, while the latter is an independently evolved Microsoft SQL database. Both are prototypical for the situation of signed language resources in the world: corpora of running signing have only recently been under development and typically use ELAN as the annotation tool, while lexical databases have typically evolved as standalone applications to produce dictionaries in books or on CD/DVD-ROMs. In order to establish the link in both directions a conversion of the lexical database to the CLARIN LMF standard for lexica is planned in the project. Further, enhancements to the existing CLARIN tools ELAN and LEXUS are created to start the exchange between the two domains.

09-007 TQE

Transcription Quality Evaluation

2010-01-01 – 2010-07-01

Project coordinator: Dr. H. Strik (Radboud University Nijmegen)

http://vps8639.xlshosting.net/TQE/

http://hdl.handle.net/1839/00-SERV-0000-0000-0005-6

Participant	Budget
Institute for Dutch Lexicology	€5,159
Radboud University	€41,342
Max Planck Institute for Psycholinguistics	€10,317
Total	**€56,818**

Abstract

The current proposal is about a completely automatic Transcription Quality Evaluation (TQE) tool. Input is a corpus with audio files and phone transcriptions (PTs). Audio and PTs are aligned, phone boundaries are derived, and for each segment-phone combination it is determined how well they fit together, i.e. for each phone a TQE measure (a confidence measure) is determined, e.g. ranging from 0-100%, indicating how well the fit is, what the quality of the phone transcription is. The output of the TQE tool will consist of a TQE measure and the segment boundaries for each phone in the corpus. 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 (L&ST) - in which audio and PTs are involved.

09-008 DUELME-LMF

Converting DUELME into LMF format

2010-01-01 – 2010-06-30

Project coordinator: prof.dr. J.E.J.M. Odijk (Utrecht University)

Participant	Budget
Utrecht University	€30,000
Institute for Dutch Lexicology	€30,000
Total	**€60,000**

Abstract

The goal of the project is develop a converter from DUELME into LMF format, and vice-versa (for a subset of lexicons in LMF-format), to apply the first converter to create a curated DUELME resource fully compliant with standards supported by CLARIN. A mapping will be defined between DUELME-specific data categories and (possibly newly created) ISOCAT data categories to ensure semantic interoperability of the curated resource with other resources and tools. A document will be produced in which limitations of and desiderata for the LMF standard, ISOCAT and other infrastructural aspects will be described.

09-011 TICCLops

Text-Induced Corpus Clean-up online processing system

2009-11-15 – 2010-05-15

Project coordinator: dr. Martin Reynaert (Tilburg University)

http://portal.clarin.inl.nl/ticclops

Participant	Budget
National Library of the Netherlands	€5,159
Institute for Dutch Lexicology	€5,159
Tilburg University	€43,877
Total	**€54,195**

Abstract

We propose a demonstration project which will allow CLARIN users to submit their corpora for fully automatic spelling correction and normalization by TICCLops, the online processing version of our core component TICCL. This system should be widely applicable in all manner of curation projects and lexicographical work.

09-012 AAM-LR

Automatic Annotation of Multi-modal Language Resources

2010-01-01 – 2010-06-30

Project coordinator: prof. dr. L. Boves (Radboud Universiteit)

http://hdl.handle.net/1839/00-SERV-0000-0000-004F-F

Participant	Budget
Radboud University Nijmegen	€26,839
Max Planck Institute for Psycholinguistics	€8,812
Leiden University	€20,635
Total	**€56,286**

Abstract

The AAM-LR project aims at building a demonstrator of a web service that will help filed researchers to annotate audio- and video-recordings. At the top level the service will mark the time intervals at which specific persons in the recording are speaking. In addition, the service will provide a global phonetic annotation, using language independent phone models and phonetic features. Speech will be separated from speaker noises such as laughing. The output of the web service will be fed into the ELAN/ANNEX editor, to facilitate further manual annotation. Integration in the ELAN/ANNEX framework will reduce the need for extensive user interface development. The annotations will conform to ISOCat and potential new categories will be added to ISOCat. The AAM-LR project is completely based on open domain software.

09-014 ADEPT

Assaying Differences via Edit-Distance of Pronunciation Transcriptions

2010-06-01 – 2010-11-30

Project coordinator: Dr. Charlotte Gooskens (University of Groningen)

http://www.gabmap.nl/

Participant	Budget
University of Groningen	€46,367
Meertens Institute	€10,500
Total	**€56,867**

Abstract
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. The C-based software on which the application is to be based has existed since about 2001, has been freely distributed since 2004, and is used modestly but profitably outside of Groningen (about 20 users). But it is too complex for many potential users, esp. dialectologists and second-language learning specialists. The goal of the proposed work is to develop and implement a graphical user interface and to make the string comparison facility available as a web application. This should enable wider experimentation with the techniques. We will consult with and collaborate with specialists for the expertise needed in establishing the application, to ensure that conforms to emerging standards, and also to attempt to use the needs of the application as a test case for formulating more general infrastructure requirements for CLARIN and CLARIN-NL.

09-015 INTERVIEWs

Curation of Interview Data
2010-01-01 – 2010-07-01
Project coordinator: Dr Henk van den Heuvel (Radboud University Nijmegen)
http://www.watveteranenvertellen.nl/annotationtool/
http://wwwlands2.let.kun.nl/spex/annotationtooldemo/#search-by-words

Participant	Budget
Data Archiving and Networking Services	€19,826
Veteran Institute	€8,261
Max Planck Institute for Psycholinguistics	€2,754
Radboud University Nijmegen	€27,188
Total	**€58,029**

Abstract
The INTER-VIEWS project will make a corpus of interview data available to the community of researchers in the humanities. For this type of research data a dedicated infrastructure does not exist, there is a lack of standardized metadata and lack of coordinated disclosure for research. The project will address these issues and provide solutions. The interviews that are made available in the INTER-VIEWS project originate from the Dutch Veterans institute and are considered as representative for other collections of interview data. The building blocks that are created, combined and optimized are durable storage, persistent identification, search and harvesting mechanisms, annotation tools, and privacy protection mechanisms. The aim of the project is to deliver generic, standardized solutions for the optimal scientific use of interview data. Thus, the project will enable researchers to review and share each other's views on the data; therefore, we call the project INTER-VIEWs. The building blocks of the infrastructure are based on results created in related previous projects. The corpus of interview data will be curated by DANS, acting as a Clarin A/B Centre.

09-016 WFT-GTB

Integrating the Woordenboek der Friese taal into the Geïntegreerde TaalBank

2009-09-01 – 2010-05-01
Project coordinator: drs. Hindrik Sijens (Fryske Akademy)

Participant	Budget
Fryske Akademy	€38,820
Institute for Dutch Lexicology	€17,178
Total	**€55,998**

Abstract
The data curation of the Wurdboek fan de Fryske Taal (WFT) database and demonstration of the data in the Geïntegreerde Taalbank (GTB) dictionary web application. The original dictionary data are stored in a BRS-database with almost no metadata added.
In this project the original data will be transferred into XML with TEIannotation. In order to achieve this result, preliminary work has to be done (additional work on the WFT-data, fixing mistakes and inconsistencies), a set of metadata is defined, a dictionary entry parser to convert the data to XML will be developed.
Before the TEI-annotated WFT can be demonstrated in the GTB, a list of word classes and an index of sources and references has to be compiled. In order to implement the WFT data into the GTB application, additional adaptations have to be made.
The TEI encoding scheme used will be fully documented. The back end (search layer) for retrieval in scholarly dictionaries will be fully documented as a web service. This will contribute to the establishment of a CLARIN standard for encoding of and retrieval in scholarly dictionaries as a linguistic resource.

09-017 Adelheid

A Distributed Lemmatizer for Historical Dutch
2010-01-01 – 2010-06-30
Project coordinator: Dr. Hans van Halteren (Radboud University Nijmegen)
http://adelheid.ruhosting.nl/

Participant	Budget
Radboud University Nijmegen	€52,652
Max Planck Institute for Psycholinguistics	€2,754
Total	**€55,405**

Abstract
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, using the most appropriate resources (such as lexica) for the text in question. The need to consistently use appropriate resources leads to the intuitively obvious strategy of placing this service in the Clarin infrastructure. For each specific text, the user can then select the best resources from those available in Clarin, wherever they might reside, and where necessary supplemented by own lexica. During the project a demonstrator for the distributed automatic lemmatization will be created, with some 14th century charters as test texts as well as corresponding resources.

09-018 TDS-Curator

A web-services architecture to curate the Typological Database System

2010-04-01 – 2011-05-31

Project coordinator: dr. Alexis Dimitriadis (Utrecht University)

http://tds2.dans.knaw.nl/TDS.html

Participant	Budget
Data Archiving and Networking Services	€27,329
Max Planck Institute for Psycholinguistics	€7,613
Meertens Institute	€17,400
Utrecht Univeristy	€5,159
Total	**€57,501**

Abstract

The Typological Database System (TDS) currently provides integrated access to multiple independently developed typological databases through a common web interface. While the system is currently operational, TDS Curator will make the TDS into a sustainable service that conforms to CLARIN infrastructural requirements. The system consists of an off-line back end that imports data and organizes it into an integrated system, and a real-time front end that manages user interaction and queries. At present, user access to the system is only possible through the interactive web interface; by implementing a web services architecture to properly separate the user interface from the data access layer, it will be possible for the latter to be queried by other components of the integrated environment envisaged by CLARIN. In this way, the data contained in the TDS can be transparently available to other tools and resources; the user interface itself would also be usable with other compatible data resources. TDS Curator will be transferred from the original developers to be hosted by DANS as an empirical test case for DANS serving as a Clarin A/B centre.

Call 2

10-008 COAVA

Cognition, Acquisition and Variation Tool

2011-01-01 – 2011-12-31

Project coordinator: Dr. Leonie Cornips (Meertens Institute)

http://www.meertens.knaw.nl/coavasite/?page_id=8

Participant	Budget
Tilburg University	€8,946
Meertens Institute	€57,449
Total	**€66,395**

Abstract

This proposal for a CLARIN project is targeted at interdisciplinary research into the relation between language acquisition and language variation by developing a tool for easily exploring the linguistic characteristics of objects. Tool development departs from tools developed in previous projects. The

project will demonstrate how the CLARIN infrastructure will enable eHumanities type research by making available datasets from traditionally distinct sub disciplines (first language acquisition and historical dialectology) in a standardized way and combine them with tools for data processing.

10-009 WIP

War in Parliament. The Second World War in Parliamentary Debates in the Netherlands.
2011-01-01 – 2011-12-31
Projectcoördinator: Dr. Hinke Piersma (NIOD)
http://www.clarin.nl/node/410
http://data.politicalmashup.nl/wip/

Participant	Budget
University of Amsterdam	€72,114
Data Archiving and Networking Services	€12,878
NIOD	€33,482
Total	**€118,473**

Abstract

References to the Second World War (WW II) shaped political debate in the Netherlands for many decades. However, we have no systematic knowledge of why, how often, when, by whom or from which political party, and in which context, these references were made. Nor do we know the meanings politicians ascribed to the war years, the lessons the war was supposed to teach, and how all of this influenced political decision-making. Answering these questions will help us better understand the complex legacies of WW II.[1]

The WIP project wants to bridge the gap between current historical and social science practices and the possibilities offered by using large corpora and language resources, in particular Clarin tools for Dutch. We will do this by making a much used dataset - de Handelingen der Staten-Generaal (Dutch Hansard) - compliant with Clarin, ISOCAT and ISO/TC 37/SC 4 standards. We will also create an advanced search engine for this dataset with an intuitive and powerful query language based on XPath, and output which can be fed directly into further analysis programs like SPSS. Integrating this technology with important historical research questions will directly contribute to new and innovative ways of writing about history.

The demonstration value of the project is enhanced by the production of an enriched publication in which we answer the research question based on the curated data. URLs in the paper link to research questions operationalized as XQueries which are executable on the curated data.

10-010 NEHOL

Negerhollands Database
2011-02-01 – 2012-02-01
Project coordinator: Prof. dr. P.C. Muysken (Radboud University Nijmegen)

Participant	Budget
Radboud University Nijmegen	€16,248
Max Planck Institute for Psycholinguistics	€26,644
Total	**€42,892**

Abstract

This Resource Curation Project aims to make available to the Clarin community the data from the Dutch-lexifier Creole language Negerhollands. Negerhollands, the now extinct Creole language of the Virgins Islands, is unique in that there is a rich digitized corpus of historical as well as almost contemporary texts available that hardly have been studied. In a previous NWO project the data were carefully edited and digitized, but so far they have remained unavailable online. The format to be followed will be that of the already existing SUCA database.

10-011 WAHSP

Towards a flexible and stable CLARIN-supported webapplication for historical sentiment mining in public media

2011-01-01 – 2011-12-31

Project coordinator: prof. dr. Toine Pieters (Utrecht University)

http://www.biland.nl/

http://dev.wahsp.nl/

Participant	Budget
Descartes Centre	€52,493
National Library of the Netherlands	€3,675
University of Amsterdam	€52,493
Huygens ING	€7,874
Total	**€116,535**

Abstract

This project aims at using and populating the basic CLARIN infrastructure to enable advanced forms of text mining in large historical datasets of newspapers and journals. The challenge is to convert a specific text mining technology, so-called 'sentiment mining', into an accessible CLARIN compliant web-application addressing research questions of the intended user group of historians and policy researchers. The demonstrator will build on the sentiment mining tools developed in the STEVIN DuOMAn project. The interdisciplinary project-team (historians, linguists, computer scientists) will tailor existing tools to the specific needs of digital humanities research, with a special focus on opinions/perceptions regarding the use and abuse of drugs between 1900 and 1945. The development of this demonstrator prototype will also be used to inventory a list of requirements the CLARIN infrastructure should meet and desiderata it preferably should offer. The demonstrator will be hosted by the Huygens Institute, acting as a CLARIN A/B centre.

10-012 INPOLDER

Integrated Parser and Lemmatizer Dutch in Retrospect

2011-02-01 – 2012-02-01

Project coordinator: Prof. Dr. Ans van Kemenade (Radboud University Nijmegen)

Participant	Budget
Meertens Institute	€46,244
Radboud University Nijmegen	€23,180
Total	**€69,424**

Abstract

In this project we aim to provide for a gap in the availability of syntactically analysed corpus material

for Dutch. While there is such material for Modern Dutch, as well as for historical versions of various neighbouring languages, it is sorely lacking for historical Dutch. We propose to repair this situation efficiently by bootstrapping off existing resources. The first necessary processing steps will be provided by the Adelheid tagger, which is currently being made available through the Clarin infrastructure. The syntactic parsing process will make use of a variant of the Penn-Helsinki parser for historical texts. This parser is trained on annotated corpus text by Dan Bikel's parser generator, and we estimate that with a limited amount of semi-automatic annotation we will be able to deliver a parser for historical Dutch with a quite respectable level of quality. This parser will be made available to the scholarly community through a web interface in such a way that a work flow can be set up which starts from raw text and allows tagging, lemmatizing and parsing, with optional manual correction at all interface points.

10-014 Arthurian Fiction

Arthurian Fiction in Medieval Europe: Narratives and Manuscripts
2011-03-01 – 2012-01-01
Project coordinator: dr Bart Besamusca (Utrecht University)
http://www.arthurianfiction.org/

Participant	Budget
Max Planck Institute for Psycholinguistics	€0
Huygens ING	€39,432
Utrecht University	€7,727
Total	**€47,158**

Abstract

The resources consist of two databases with data for literary research in the area of European Arthurian fiction. The databases are currently not available to researchers. Using the data and the technology of the project, scholars will be able to study their national literatures and book production from a European and interdisciplinary perspective. They will have access to data concerning a great many European languages and cultures and concerning manuscripts and their art. The users include literary historians in all European languages, cultural historians, medieval book specialists and art historians. Arthurian scholarship in particular will profit. The current datasets will be converted to an XML standard from which metadata will be extracted and made available to CLARIN harvesters. The demonstrator will allow users access to the data for searching as well as for adding and correcting records.

10-016 VU-DNC

VU Diachronic Newspaper Corpus
2011-03-01 – 2012-06-01
Project coordinator: Prof. dr. W. Spooren (VU University Amsterdam)
https://portal.clarin.inl.nl/vu-dnc/

Participant	Budget
VU University Amsterdam	€38,971
Institute for Dutch Lexicology	€2,266
Tilburg University	€17,646
Total	**€58,883**

Abstract
The VU-DNC project has four main aims: 1) to make a unique diachronic corpus of Dutch newspaper articles from five major Dutch newspapers from 1950/1951 and 2002 (2 MW) available to the community of researchers in the humanities, 2) to extend the linguistic annotation of discourse with encoding for lexico-grammatical features of subjectivity and quotations, 3) to create a gold standard benchmark that can be used for testing and training OCR-postcorrection tools, by aligning uncorrected and corrected versions of the digitized printed newspaper articles from 1950/51, and 4) to improve the development of metadata within CLARIN by mapping the data categories for the part of speech and lemma coding to the data category registry, and extending the ISOcat categories for the historical spelling variation, subjectivity and quotations.

10-022 C-DSD

Curating the Dutch Song Database
2011-03-01 – 2011-08-31
Project coordinator: prof. dr. E Stronks (Utrecht University)
http://www.clarin.nl/node/202

Participant	Budget
Utrecht University	€5,151
Meertens Institute	€24,747
Total	**€29,898**

Abstract
The Dutch Song Database (DSD) is a database in the field of Literary Studies. It contains (meta-)data on 140.000 songs and their 15.000 sources (songbooks, pamphlets, field recordings, etc.) from the Middle Ages to the present day. Built and rebuilt over a period of 25 years, consisting of four distinct datasets, this database was enlarged and enriched in many stages, with grants from (among others) NWO and OCW. The first online version of the DSD was published in 2007. As the DSD was built over a long period of time, parts of the metadata sets were adjusted to modern norms, but never the time was found to curate the database as a whole, that is to make it compliant to the current metadata standards and protocols. The ratio behind the analysis of data and the production of metadata in the DSD is internationally renowned and perceived as exemplary. Yet, because the DSD is not based upon now common, up-to-date standards for the encoding of metadata, the opportunities to export the DSD model are limited, as are the opportunities for (international) collaboration with other projects with similar source materials. The aim of the proposed project, 'Curating the Dutch Song Database' is to create new perspectives for international collaboration. The DSD will be employed in the CLARIN-EU Search & Develop's Multimedia/Multimodal Demonstrator, in which the DSD will participate.

10-023 IPROSLA

Integrating and publishing resources on sign language acquisition
2011-04-01 – 2011-12-31
Project coordinator: Prof. dr. P. (Paula) Fikkert (Radboud University Nijmegen)
http://www.ru.nl/sign-lang/projects/iprosla/

Participant	Budget

University of Amsterdam	€10,302
University of Applied Sciences Utrecht	€10,302
Radboud University Nijmegen	€40,487
Max Planck Institute for Psycholinguistics	€5,499
Total	**€66,590**

Abstract

This resource curation project aims to integrate two different data sets on sign language acquisition in one central archive. First, a diverse set of longitudinal data of deaf children from deaf and hearing parents that has been collected at the UvA in the last 20 years, and secondly, a new collection of longitudinal data collected at the RU from hearing children of deaf parents. Neither of the two has been properly documented with metadata descriptions, and neither of the two has been safely archived anywhere. Only of the latter, some CLARIN-compliant annotation is available. The goal of the project is to document the two data sets with CMDI, and archive them at the MPI language archive.

Call 3

11-006 FESLI

Functional Elements in Specific Language Impairment (Closed call)

2012-02-01 – 2013-01-31

Project coordinator: Prof.dr. Fred Weerman (University of Amsterdam)

Participant	**Budget**
University of Amsterdam	€45,715
Meertens Institute	€33,114
Total	**€78,829**

Abstract

Specific Language Impairment (SLI) is a developmental language disorder that is visible in particular in the acquisition of functional elements. In order to understand it better a comparison has to be made of the profile of functional elements in several groups of learners both qualitatively and quantitatively. To this end the FESLI project will make available Dutch corpora from monolingual and bilingual children with SLI, and corpora from typically developing bilingual children. Complemented with the existing CHILDES corpora from typically developing monolingual children evidence from all relevant learners groups will be available. Based on the technology from the COAVA project FESLI will develop additional tools that allow quantitative and qualitative comparisons of functional elements.

11-007 PILNAR

Pilgrimage Narratives: Creating a Corpus for Studying the Profile of the Modern Pilgrim (Closed call)

2012-04-01 – 2013-04-01

Project coordinator: Prof. dr. Paul Post (Tilburg University)

Participant	**Budget**
Tilburg University	€46,372
Meertens Institute	€33,114
Total	**€79,486**

Abstract
Pilgrimage narratives, especially travel accounts, have been used as a favorite source for research into ritual and religious dynamics for a long time. This project proposal is intended to establish a core corpus of modern pilgrimage narratives. It will consist of Dutch texts written after ca. 2000 that present the thoughts and impressions of pilgrims to Santiago de Compostela. This source has not or hardly been used for contemporary research in the cultural sciences. In previous exploratory research (Post 1992, 1994ab, 1998), it has become clear that the corpus of stories intended here is an excellent source for research into the profile (or, better, profiles) of the modern pilgrim. To open up this source and map these profiles, the heuristic tool, "Fields of the Sacred" will be used (Post 2010, 2011).

11-008 Gr-Ne

Grieks-Nederlands woordenboek (Closed call)
2012-04-01 – 2013-03-31
Project coordinator: Prof. dr. Ineke Sluiter (Leiden University)

Participant	Budget
Leiden University	€24,665
Amsterdam University Press	€6,068
University of Amsterdam	€42,055
Total	**€72,788**

Abstract
This project combines a resource curation and demonstrator project. It concerns the new dictionary (ancient)1 Greek-Dutch, currently under construction at the Leiden University Classics Department. [Resource curation:] For this resource (originally developed as book), a DTD will be developed and a substantial part of the data will be put into XML. The purpose is to make the data visible, uniquely referable and accessible via the web. Proper documentation will be provided. Since the dictionary is still under construction, it cannot be put online completely within the 12 months of the project. [Demonstrator:] However, as the second half of the project, we will develop an interface with the search functions ultimately envisaged for the whole online dictionary. Search functions will include searches for Greek lemmata; search of Greek declined or conjugated word-forms that will lead to the correct lemma ('parser'); searches for Dutch words leading to different Greek lemmata; etymological searches.

09-009 EMIT-X

Early-Modern Image and Text eXchange
2012-03-01 – 2013-03-01
Project coordinator: prof. dr. E. Stronks (Utrecht University)

Participant	Budget
Huygens ING	€17,486
Duke August Library Wolfenbüttel	€0
University of Illinois at Urbana-Champaign	€0
Utrecht University	€2,585
Total	**€20,071**

Abstract
In the Early-Modern period images and texts were thought of as closely related. In books, paintings and buildings they were often used jointly in an effort to talk both to the mind and to the senses. A prime example of this bimedial culture is the emblem book. Research of the intertwining of visual and textual traditions in the emblem has profited immensely from digitization efforts, for digital editions facilitate complex searches in large corpora of emblems. The Emblem Project Utrecht (EPU) digitized Dutch emblems books, and other corpora were published by groups elsewhere in Europe and the US. In a community effort coordinated in the Open Emblem Group, an exchange format has been designed to facilitate the aggregation of material from individual projects. EMIT-X will implement this exchange format in an OAI data provider, facilitating sharing and re-use of the EPU data and preparing the way for related projects elsewhere.

09-011 D-LUCEA

Database of the Longitudinal Utrecht Collection of English Accents
2012-04-01 – 2013-04-01
Project coordinator: Dr Hugo Quené (Utrecht University)

Participant	Budget
Utrecht University	€36,190
University College Utrecht	€10,340
Radboud University Nijmegen	€5,170
Max Planck Institute for Psycholinguistics	€16,557
Total	**€68,257**

Abstract
The proposed D-LUCEA project concerns the curation of a database of existing speech recordings of L1 and L2 speakers of English. The recorded speakers are students from an international student community where English is used as lingua franca. These students are being recorded longitudinally throughout their 3-year period on campus, using read and spontaneous speech in L1 and in L2 English (or in L1 English only). The proposed project aims to make the presently existing recordings compatible with CLARIN standards, and to make the resulting database accessible to researchers worldwide. This is achieved by creating and developing and verifying metadata, creating documentation, and by issueing persistent identifiers to primary data and metadata. The resulting database will be of interest for research and development in linguistics, language education (pronunciation training), speech technology (foreign accent detection, language recognition, speech recognition), and sociophonetics.

11-012 MIGMAP

Interactive migration maps for the 20th century
2012-03-01 – 2012-09-01
Project coordinator: dr. Gerrit Bloothooft (Utrecht University, Meertens Institute)
http://www.meertens.knaw.nl/migmap/

Participant	Budget
NIDI	€5,170
Meertens Institute	€33,114

Utrecht University	€15,510
Total	**€68,257**

Abstract

People migrate and take their social-cultural-linguistic identities with them. Since in their new environment this leads to interactions, knowledge of migration is of high interest to the understanding of, for instance, sociolinguistic and dialect diffusion processes. Based on the availability of places of birth and residence (in 2006) of the Dutch population (16 million alive, 6 million deceased but included) and their family relations from the Civil Registration, migrations patterns between municipalities (and immigration from abroad) can be presented over three generations in the 20th century. The project will develop a web application where the user first chooses generation (forward or backward in time) and gender, while the migration map of The Netherlands related to an interactively pointed municipality (or other aggregation unit) is shown. The existing map-making software module "Kaart" of the Meertens Institute will be transformed into a generic, standards-based tool for the creation and presentation of maps with complex spatio-temporal diffusion data in a user friendly and interactive way.

11-014 VK

Verrijkt Koninkrijk

2012-01-01 – 2012-12-20

Project coordinator: Dr Kees Ribbens (NIOD)

Participant	**Budget**
University of Amsterdam	€36,190
VU University Amsterdam	€28,435
Data Archiving and networking Services	€5,519
Meertens Institute	€5,519
NIOD	€35,508
Total	**€68,257**

Abstract

Dr Loe de Jong's *Het Koninkrijk der Nederlanden in de Tweede Wereldoorlog* remains the most appealing history of German occupied Dutch society (1940-1945). Published between 1969 and 1991, the 30 volumes still combine the qualities of an authoritative work for a general audience, and an inevitable point of reference for scholars. The aim of this project is twofold; in the demonstrator part of the project advanced tools and techniques are applied to gather data on De Jong's perception of the much debated issue of pillarization (Dutch: 'verzuiling') and group identity. In the resource curation part of the project the corpus will be enriched and made available to the CLARIN-community for further research. The overall budget for the project is € 119,993 and the partners are: NIOD Institute for War, Holocaust and Genocide Studies (NIOD), University of Amsterdam (UvA), Vrije

Universiteit Amsterdam (VUA), Meertens Institute and Data Archiving and Networked Services (DANS-KNAW).

11-017 BILAND

Towards a flexible and stable CLARIN-supported web-application for bilingual historical analysis of discourses in news media

2012-01-01 – 2013-12-31

Project coordinator: prof. dr. Toine Pieters (Utrecht University)

Participant	Budget
DIA	€1,104
NIOD	€1,104
Descartes Centre	€52,431
National Library of the Netherlands	€1,656
Staatsbibliothek zu Berlin	€0
Deutsche Nationalbibliothek	€0
University of Amsterdam	€57,950
Huygens ING	€5,519
Total	**€119,762**

Abstract

This project aims at using and populating the basic CLARIN infrastructure to enable bilingual and biscriptural advanced forms of discourse analysis in large historical datasets. The challenge is to convert a specific text mining technology, into an accessible CLARIN compliant web-application addressing research questions of the intended user group of historians. The demonstrator will further build on the text mining tools developed in previous CLARIN projects. The interdisciplinary project-team will tailor existing tools to the language specific needs of comparative historical research, with a special focus on the identity, intensity and location of discourses about heredity, genetics and eugenics in Dutch and German news media between 1863 and 1940. The challenge is to incorporate the semantics of two different languages and cultures. The development of this bilingual demonstrator will contribute to inventory a list of requirements the CLARIN infrastructure should meet and desiderata it preferably should offer within European contexts.

11-018 MultiCon

Multilayer Concordance Functions in ELAN and ANNEX

2012-01-01 – 2012-12-31

Project coordinator: Dr. O.A. Crasborn (Radboud University Nijmegen)

Participant	Budget
Max Planck Institute for Psycholinguistics	€49,671
Radboud University Nijmegen	€5,170

Total	**€54,841**

Abstract

Collocations generated by concordancers are a standard instrument in the exploitation of text corpora for the analysis of language use. Multimodal corpora show similar types of patterns, activities that frequently occur together, but there is no CLARIN-compliant tool that offers facilities for visualising such patterns. Examples include timing of eye contact with respect to speech, and the alignment of activities of the two hands in signed languages. This project proposes to enhance the standard CLARIN tools ELAN and ANNEX for multimodal annotation to address these needs, first of all by improving the query and concordancing functions, and secondly by generating visualisations of multilayer collocations that allow for intuitive explorations and analyses of such data. This will provide a boost to the linguistic fields of gesture and sign language studies, as it will improve the exploitation of multimodal corpora.

11-020 Cornetto-LMF-RDF

Curated Cornetto database in LMF and RDF

2012-04-01 – 2013-04-01

Project coordinator: Piek Vossen (VU University Amsterdam)

Participant	Budget
Institute for Dutch Lexicology	€55,190
VU University Amsterdam (LET)	€51,700
VU University Amsterdam (FEW)	€10,340
Total	**€117,230**

Abstract

This is a combined curation and demonstrator project in which the Dutch Cornetto database is converted to LMF and RDF and made available on a CLARIN Centre for efficient querying. As a semantic resource in which words and concepts are interlinked within the data and to other databases (e.g. wordnets in other languages and ontologies) this project will address many issues on the representation of meaning and user-queries to these data, such as the complex data structure (semantic and structural) and semantic linkage, such as hypernym chains of concepts or semantic typing of words. The project will combine a new release of Cornetto (version 2) with the data from DutchSemCor (a semantic annotation of text corpora) and a Dutch sentiment lexicon. The results are presented in LMF and the wordnet part also in RDF and SKOS. This bridges the standardization and metadata requirements of ISO and W3C.

11-022 PoliMedia

Interlinking multimedia for the analysis of media coverage of political debates

2012-03-01 – 2013-03-01

Project coordinator: Prof. dr. Henri Beunders (Erasmus University Rotterdam)

Participant	Budget
Delft University of Technology	€41,533
Erasmus University Rotterdam	€38,775
Netherlands Institute for Sound and Vision	€39,185
Total	**€119,493**

Abstract
Analysing media coverage across several types of media-outlets is a challenging task for (media) historians. Up until now, the focus has been on newspaper articles: being generally available in digital, computer-readable format, these can be studied relatively easily. Cross-media comparisons between different types of media-outlets have however rarely been undertaken, even though such comparisons have top priority on the wish-list of (media) historians as this could give better insight into the choices that different media-outlets make. A specific example of media coverage research investigates the coverage of political debates and how the representation of topics and people change over time. The PoliMedia project aims to showcase the potential of cross-media analysis for research in the humanities, by (i) curating automatically detected semantic links between four data sets of different media types, and (ii) developing a demonstrator application that allows researchers to deploy such an interlinked collection for quantitative and qualitative analysis of media coverage of debates in the Dutch parliament.

11-029 Namescape

Mapping the Landscape of Names in Modern Dutch Literature
2012-01-01 – 2013-01-01
Project coordinator: dr. Karina van Dalen-Oskam (Huygens ING)

Participant	Budget
University of Amsterdam	€44,402
Institute for Dutch Lexicology	€54,310
Huygens ING	€20,930
Total	**€119,642**

Abstract
Recent research has conclusively proven names in literary works can only put fully into perspective when studied in a wider context (landscape) of names either in the same text or in related material (the onymic landscape or "namescape"). Research on large corpora is needed to gain a better understanding of e.g. what is characteristic for a certain period, genre, author or cultural region. The data necessary for research on this scale simply does not exist yet. The proposed project aims to fill the need by annotating a substantial amount of literary works with a rich tag set, thereby enabling the participating parties to perform their research in more depth than previously possible. Several exploratory visualization tools will help the scholar to answer old questions and uncover many more new ones, which can be addressed using the demonstrator. The main tools will be made available as CLARIN compliant web services for use in other contexts.

11-031 DiscAn

Towards a Discourse Annotation system for Dutch language corpora (Closed call)
2012-04-01 – 2013-04-01
Project coordinator: Prof. Dr. T.J.M. Sanders (Utrecht University)

Participant	Budget
Utrecht University	€56,266
Max Planck Institute for Psycholinguistics	€16,527
Total	**€72,793**

Abstract
Although discourse is a crucial level in language and communication, existing corpora of Dutch language lack annotation at this level. The DiscAn project sets the first step to change this situation for Dutch, in line with international tendencies. It has four main goals: 1) to standardize and open up an existing set of Dutch corpus analyses of coherence relations and discourse connectives; 2) to develop the foundations for a discourse annotation system that can be used in Dutch natural language corpora; 3) to improve the metadata within CLARIN by investigating existing CMDI profiles or adding a new CMDI profiles especially suited for this type of analysis; 4) to inventorize the required discourse categories and investigate to what extent these could be included in ISOCAT categories for discourse that are currently being developed; 5) to further develop an interdisciplinary discourse community of linguists, corpus and computational linguists in The Netherlands and Belgium, in order to initiate further research on cross-linguistic comparison in a European context.

13-004 MinDisc

Minimal Set for Discourse Annotation
2013-03-01 – 2013-06-30
Project coordinator: prof. dr. T. Sanders (Utrecht University)

Participant	Budget
Utrecht University	€15,637
Total	**€15,637**

Abstract
This project has two goals: (i) The organisation of a workshop in June 2013 where international experts aim to develop an international ISO-standard for the annotation of discourse relations (as inventoried in the Call 3 project DiscAn). Furthermore (ii), a short empirical study will be carried out on the systematicity of discourse annotation. In this study a minimal set for discourse annotation (MinDisc) will be evaluated for inter-annotator agreement in lightly trained annotators (BA students).

12-002 LAISEANG

Language Archive of Insular South East Asia and West New Guinea
2012-11-01 – 2013-11-01
Project coordinator: Dr. Marian Klamer (Leiden University)

Participant	Budget
Leiden University	€67,760
Max Planck Institute for Psycholinguistics	€5,727
Total	**€73,487**

Abstract
The geographical region of insular South East Asia and New Guinea is well-known as an area of mega-biodiversity. Less well-known is the extreme linguistic diversity in this area: over a quarter of the world's 6000 languages are spoken here. As small minority languages, most of these will cease to be spoken in the coming few generations. This project will ensure the preservation of unique records of languages and the cultures encapsulated by them in the region. The language resources have been gathered by twenty linguists at, or in collaboration with Dutch universities over the last 40 years, and

will be compiled and archived in collaboration with The Language Archive (TLA) in Nijmegen. The resulting archive will constitute an unrivaled collection of multimedia materials and written documents from over 50 languages in Insular South East Asia and West New Guinea. At TLA, the data will be archived according to state-of-the-art standards (TLA holds the Data Seal of Approval): the component metadata infrastructure CMDI will be used; all metadata categories as well as relevant units of annotation will be linked to the ISO data category registry ISOcat. This guarantees the proper integration of the language resources into the CLARIN framework. Through the archive, future speaker communities and researchers will be able to plumb the materials for answers to their own questions, even if they do not themselves know the language, and even if the language dies.

Call 4

12-006 @PhilosTEI

TICCLing Philosophy: a TEI corpus-building workflow towards a new computational methodology for philosophy (Closed call)

2013-09-01 – 2014-08-30

Project coordinator: Dr. Arianna Betti (VU University Amsterdam)

Participant	Budget
VU University Amsterdam	€ t.b.a.
Huygens ING	€ t.b.a.
Tilburg University	€ t.b.a.
Total	**€ t.b.a.**

Abstract

The step to e-research in philosophy depends on the availability of high quality, easily accessible corpora in a sustainable format composed from multi-language, multi-script books from different historical periods. Corpora matching these needs are at the moment virtually non-existing. In this project we want to address this corpus building problem by developing and making available an open source, web-based, user-friendly workflow from textual digital images to TEI, based on an OCRopus/Tesseract webservice and a multilingual version of OCR-postcorrection webservice TICCLops. We shall demonstrate the tool on a multilingual, multi-script corpus of important 18th-20th-century European philosophical texts. These texts are of fundamental importance to understand the development of key scientific concepts such as explanation and truth in 18th-20th-century Europe. The tool will be of general interest and importance to solve problems of CLARIN-compliant corpora building.

12-005 QuaMeRDES

Quantitative Content Analysis of Media Researchers' Data (Closed call)

2013-05-01 – 2014-05-01

Project coordinator: Dr. Jasmijn van Gorp (Utrecht University)

Participant	Budget
Utrecht University	€31,218
Netherlands Institute for Sound and Vision	€11,713
University of Amsterdam	€75,977
Total	**€75,977**

Abstract
The QuaMeRDES demonstrator will enable quantitative content analysis of television and printed media. Key is providing meaningful links to various streams of information relevant to media studies scholars. This includes television broadcasts, newspaper archives, metadata records and also time-based information, notably subtitle files. QuaMeRDES will expand the existing tool MeRDES (Media Researchers' Data Exploration Suite) , developed in the NWO-CATCH project BRIDGE. MeRDES is an analysis tool that visualises trends in extensive catalogues maintained by archives. This tool will be expanded to support principles of quantitative content analysis and explore new data visualisation paradigms. The result will integrate wrappers that read and write metadata and provenance information provided by the CLARIN infrastructure. 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.

12-008 e-BNM+

Linked Data on Middle Dutch Sources Kept Worldwide
2013-05-01 – 2014-05-01
Project coordinator: Dr. André Bouwman (Leiden University)

Participant	Budget
Utrecht University	€5,203
Leiden University	€14,048
Huygens ING	€48,884
Total	**€68,135**

Abstract
The e-BNM is a database which collects and presents textual, codicological and historical information about thousands of Middle Dutch manuscripts kept worldwide. e-BNM is one of the most important sources of knowledge for all scholars working in medieval studies pertaining to The Netherlands and Flanders. The proposed project provides e-BNM with a much needed conversion into a flexible datastructure that will turn e-BNM into a key open access resource to which many other resources can be easily linked. Persistency and long term sustainability will be of main importance. The project will deliver a web application for consultation, using facetted search, and will enable collaborative editing.

12-010 VALID

Vulnerability in Acquisition: Language Impairments in Dutch. Curating five valuable data sets
2013-04-01 – 2014-04-01
Project coordinator: Dr. Jetske Klatter (Radboud University Nijmegen)

Participant	Budget
Radboud University Nijmegen	€41,519
Max Planck Institute for Psycholinguistics	€8,279
University of Amsterdam	€5,170
Utrecht University	€9,306
Total	**€64,274**

Abstract

Research groups from the universities of Nijmegen, Amsterdam (UvA), and Utrecht have decided to prepare a nationwide, open access multimedia archive of language pathology data collected in the Netherlands, primarily on Dutch. In this enterprise, as many other research groups as possible, both from universities and care and educational institutes, will be involved. Although data sharing is becoming increasingly important to widen the scope and depth of empirical research in all kinds of areas, no tradition has been established yet in language and speech pathology research. This means that in various places a wealth of relevant and precious language data exists that cannot be easily and optimally accessed and exploited. The aim of the present project is to curate five existing, digital data sets, in order to make them available for scientific research in CLARIN-compatible format. This is a first, major step in the development of a VALID data archive.

12-011 COBWWWEB

Connections Between Women and Writings Within European Borders

2013-04-01 – 2013-12-31

Project coordinator: Dr. Suzan van Dijk (Huygens ING)

Participants:

Participant	Budget
Huygens ING	€95,490
Total	**€95,490**

Abstract

Recent NWO and COST projects funded the collection of data in view of research on the reception and internationalization of women's literature in Europe up to the 20th century. The WomenWriters database currently contains references concerning 4.000 authors, their works and over 22.000 reception documents. To expand the research networks about female participation in the literary field we intend to: (1) connect this data 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. These standards will be implemented by international partners: the Selma Lager Archive and the Literature Bank in Sweden; the Google Grant project Collaborative Annotation of Digitalized Literary Texts in Madrid; the Serbian Književnstvo; the Norwegian Female Robinsonades and the Swiss Women in Arcadia (1690-1800) databases.

12-017 SHEBANQ

System for HEBrew Text: ANnotations for Queries and Markup

2013-05-01 – 2014-04-30

Project coordinator: Prof. dr. Wido van Peursen (VU University Amsterdam)

Participant	Budget
VU University Amsterdam	€60,910
Data Archiving and Networking Services	€40,274
Total	**€101,184**

Abstract

The WIVU Hebrew Text Database contains the Hebrew text of the Old Testament enriched with many linguistic features at the morpheme level up to the discourse level. This work of decades is currently represented in an object database that is optimized for linguistically relevant queries. However, this resource is not readily available to researchers, and moreover, work based on this resource cannot be linked to it on the web. The curation part of this project will create a durable representation of the contents of the WIVU database in LAF plus an annotatable Linked Data export in RDF. There will be persistent addresses for fine-grained fragments. The demonstrator part will be a web application that enables researchers to perform linguistic queries on the web resource and preserve significant results as annotations to this resource. The more than 100 different features will be defined in ISOcat and used by CMDI profiles.

12-019 DSS

Dutch Ships and Seamen

2013-04-01 – 2014-02-28

Project coordinator: Prof. dr. Lex Heerma van Voss (Huygens ING)

Participant	Budget
VU University Amsterdam	€42,925
Huygens ING	€55,512
Total	**€98,436**

Abstract

As a sea-faring nation, a large portion of Dutch history is found on the water. However, much of the digitized historical source material is still scattered across many databases and archives. This curation and demonstrator project aims to bring together the rich maritime historical data preserved in the many different databases. We propose a (semantic) web-based infrastructure that will house various maritime-historical datasets. We will provide a tool chain and methodology for converting legacy datasets. The infrastructure includes common vocabularies to normalize and enrich existing data. Links are established between the datasets and to other relevant datasets on the Web. Although the infrastructure will be set up to facilitate 25+ identified datasets, we initially populate the infrastructure with four selected datasets. These will allow us to investigate two case studies in order to answer the historical research question “To what extent did patterns of shipping and recruitment in the Dutch maritime sector change over the course of the 18th and 19th centuries?”

12-020 EXILSEA

Exploiting ISOcat's Language Sections in ELAN and ANNEX

2013-03-01 – 2014-05-01

Project coordinator: Dr. Onno Crasborn (Radboud University Nijmegen)

Participant	Budget
Radboud University Nijmegen	€53,794
Max Planck Institute for Psycholinguistics	€11,038
Total	**€64,832**

Abstract

With the growing amount of online language resources, the need to exploit these in innovative ways is also increasing. This project aims to strengthen the CLARIN infrastructures by extending ELAN and

ANNEX for the annotation and display of time---based resources such as audio and video with a referencing and note exchanging system. Precise and persistent references to time slices and annotations in the form of hyperlinks will be made possible. A system for exchanging notes containing these links and exploiting ISOcat will be created, allowing for efficient collaboration among researchers. ANNEX will be extended so that multiple references can be presented side---by---side, allowing for the display of multiple examples of the same phenomenon, for instance. While the system will be tested on the basis of linguistic data, the functionality can be exploited by any research that makes use of time---based resources, whether in the humanities or the social sciences. The demonstrator part will be a web application that enables researchers to perform linguistic queries on the web resource and preserve significant results as annotations to this resource. The more than 100 different features will be defined in ISOcat and used by CMDI profiles.

12-021 ColTime

Collaboration on Time--Based Resources

2013-03-01 – 2014-05-01

Project coordinator: Dr. Onno Crasborn (Radboud University Nijmegen)

Participants: Radboud University Nijmegen, Max Planck Institute for Psycholinguistics

Participant	Budget
Radboud University Nijmegen	€10,340
Max Planck Institute for Psycholinguistics	€52,431
University of Amsterdam	€5,170
Total	**€67,941**

Abstract

With the growing amount of online language resources, the need to exploit these in innovative ways is also increasing. This project aims to strengthen the CLARIN infrastructures by extending ELAN and ANNEX for the annotation and display of time---based resources such as audio and video with a referencing and note exchanging system. Precise and persistent references to time slices and annotations in the form of hyperlinks will be made possible. A system for exchanging notes containing these links and exploiting ISOcat will be created, allowing for efficient collaboration among researchers. ANNEX will be extended so that multiple references can be presented side---by---side, allowing for the display of multiple examples of the same phenomenon, for instance. While the system will be tested on the basis of linguistic data, the functionality can be exploited by any research that makes use of time---based resources, whether in the humanities or the social sciences. The demonstrator part will be a web application that enables researchers to perform linguistic queries on the web resource and preserve significant results as annotations to this resource. The more than 100 different features will be defined in ISOcat and used by CMDI profiles.

12-022 RemBench

A Digital Workbench for Rembrandt Research

2013-04-01 – 2014-03-31

Project coordinator: Prof. dr. Volker Manuth (Radboud University Nijmegen)

Participant	Budget
Radboud University Nijmegen	€63,717
Huygens ING	€50,500

Netherlands Institute for Art History	€6,968
Total	**€121,184**

Abstract
The goal of this project is to demonstrate how (art) historians can benefit from linking a pivotal CLARIN resource, namely eLaborate, which adheres to ISOCat, to resources created by museums, archives and libraries, which adhere to other standards for metadata. For that purpose, we intend to construct a demonstrator that connects a number of databases centred around the life and art of Rembrandt van Rijn. In this project we will demonstrate how an initial version of the Rembrandt Documents (RemDoc) database, built as an eLaborate application, can serve as an integrated tool for art history research by virtue of its coupling with related resources created by the Rijksbureau voor Kunsthistorische Documentatie (RKD), as well as with a standard university library catalogue.

12-013 OpenSoNaR

Online Personal Exploration and Navigation of SoNaR (Extended version)
2013-09-01 – 2013-08-31
Project coordinator: Prof. dr. Max Louwerse (Tilburg University)

Participant	**Budget**
Tilburg University TiCC	€98,582
Tilburg University TSH	€5,203
Institute for Dutch Lexicology	€41,107
Huygens ING	€2,601.50
Meertens Institute	€2,601.50
Total	**€150,095**

Abstract
The OpenSoNaR project will provide end users with the online means for extracting information from the SoNaR-500 reference corpus of contemporary written Dutch. This includes exploring the texts and navigating through the SoNaR-500 corpus by way of the metadata. The project makes the contents of the new SoNaR-500 reference corpus available to laymen and specialist researchers alike. Based on the desiderata of four distinct CLARIN-NL priority groups, access to the corpus for navigation, exploration and exploitation in an online environment will be through a front-end, to be called WhiteLab, providing a range of interfaces that provide user-driven functionality. The back-end is the new retrieval engine BlackLab developed by INL (Dutch Institute for Lexicology), designed to provide access to corpora for linguistic and lexicographical use in the CLARIN infrastructure.

2013 Projects

13-012 Lassy

Maintenance of the Lassy Word Relations Interface
Project Coordinator: prof.dr. Gertjan van Noord (RUG)
2013-11-01 – 2014-11-01

Organization	Total
RUG	€ 2,000
Grand Total	**€ 2,000**

Abstract
The project intends to revive the Lassy Word Relations Interface and make sure it keeps running during the project time.

2014 Projects

14-003 CMDI2RDF

Converting CMDI to RDF
Project Coordinator: dr. Menzo Windhouwer (DANS)
2014-05-01 – 2014-12-01

Organization	Total
DANS	€ 33,000
Grand Total	**€ 33,000**

Abstract
There is a continuous growing interconnected collection of linguistic data sources in the LOD cloud. This project will make the joint metadata domain of CLARIN take up its place in this cloud by making available an RDF representation and establishing links with existing RDF data sets. The obvious linking pin to many linguistic LOD sets will be based on language identifiers. But also the semantic facilities in the CMD Infrastructure, like ISOcat, RELcat (for predicates) and CLAVAS (for entities like organisations), can provide other, rich mappings which enlarge the connectivity of CMD records to the existing (linguistic) LOD cloud. Both common ways of data access in the LOD (downloading the resource, querying the resource via a SPARQL endpoint) will be provided by loading the RDF representations in a triple store. Next to these more machine-oriented interfaces it will be attempted to create an interactive user interface, a RDF browser. However, the CMD Infrastructure is very flexible and can lead to a wide variety of predicates, which might be hard to cope with by an off-the-shelf RDF browser with only a limited amount of tweaking. This triple store and, if successful, the new CMDI specific RDF browser can open up new ways of data discovery for CLARIN, both in the form of a web-interface and as an API

14-010 PaQu

Parse and Query
Project Coordinator: prof.dr. Gertjan van Noord (RU Groningen)
2014-07-01 – 2014-12-31

Organization	Total
RU Groningen	€ 6,667

Grand Total	**€ 6,667**

PaQu (Parse and Query) combines two existing tools: Alpino, the automatic parser for Dutch on the one hand, and the LASSY Word Relations search application (LWRS) on the other. For Alpino a web service has been made available in the TTNWW project. LWRS searches in LASSY-Small and the Spoken Dutch Corpus (CGN). In PaQu a user can provide a document (set), which is automatically parsed by the Alpino web service, after which the document can be queried with the LWRS search application. This will enable users to carry out queries into syntactic phenomena in the data that they need and use for their research, and are not bound the LASSY and CGN corpora only.

14-011 OpenConvert

Project Coordinator: Jan Theo Bakker (INL)
2014-09-01 – 2014-21-31

Organization	Total
INL	€ 11,246
Grand Total	**€ 11,246**

Abstract

OpenConvert aims to develop converters from formats often used in the real world into CLARIN standard formats. It will create an initial set of converters, more specifically to convert from plain text, HTML, ePub, Word 1997-2003, Word 2010, and ALTO into TEI or FoLIA.

14-012 AutoSearch

Search for token-based annotations in user-uploaded tagged corpora
Project Coordinator: Jan Theo Bakker (INL)
2014-09-01 - 2015-03-31

Organization	Total
INL	€ 10,039
Grand Total	**€ 10,039**

Abstract

Autosearch will enable users to upload a corpus, have it pos-tagged, after which the enriched corpus can be queries with a Modern Dutch Corpus-like search application. For PoS-tagging TTNWW will be used. The supported input formats are Folia, TEI, and plain text.

14-013 CORNETTO-LMF-II

Improved Interface for CORNETTO-LMF

Project Coordinator: prof.dr. Piek Vossen (Free University Amsterdam)

2014-09-01 – 2015-03-31

Organisation	Total
INL	€ 29,659.50
Free University	€ 5,094.00
Total	**€ 34,753.50**

Abstract

This project aims to improve the interface to the Cornetto application created in the Cornetto-LMF-RDF project and to improve the response times of the system.

2015 Projects

15-001 TPC

TaalPortaal with links to Corpora

Project Coordinator: Dr. Ton van der Wouden (Meertens Institute)

2015-05-01 – 2015-12-31

Organisation	Total
INL	€ 16,371
Meertens Institute	€ 2,547
Utrecht University	€ 16,709
RU Groningen	€ 16,709
Taalmonsters	€ 5,457
Total	**€ 57,792**

Abstract

The project creates queries in corpus search interfaces and links to them from the Taalportaal. Links will be made for example sentences, grammatical terms and names or descriptions of constructions. The queries will be offered to users of Taalportaal via standard internet hyperlinks. Clicking on such a link brings the user to a corpus query interface where the specified query will be executed. It will make a direct link between grammatical descriptions and actual data illustrating the description, thus enabling testing of the claims made. It will stimulate linguists to make use of corpus data in their research and will have the educational side effect that linguists will gradually become acquainted with such corpus query applications.

15-002 MDF

Meertens Deposition Facility

Project Coordinator: Dr. Marc Kemps-Snijders (Meertens)

2015-05-01 – 2015-12-31

Organisation	Total
Meertens Institute	€ 40,000
Total	**€ 40,000**

Abstract

The project is a pilot for upgrading the Meertens deposition possibilities both in the interest of its own researchers and of the Dutch research community. It brings Meertens in a good position to host data and make that data also available for use by the research community. This has become urgent because of the reduced activities of MPI/TLA in this respect. Access restrictions as specified by the owners will be honored. It will create additional options to ensure availability and accessibility of important data for the Dutch research community.

15-003 CorpusSearchWeb

Corpus Studio – web version

Project Coordinator: Dr. Erwin Komen (Meertens Institute)

2015-04-15 – 2015-12-31

Organisation	Total
Meertens Institute	€ 37,108
Total	**€ 37,108**

Abstract

The project aims to make a web version of the “Corpus Studio” programme. The web version will be a reduced version of the stand-alone version, but it will enable research on syntactically annotated resources in FoLiA or psdx xml format. The tool supports syntactic research, both synchronically and diachronically, by making it possible to query syntactically annotated corpora for syntactic constructions.

15-004 CARE

CuRation and integration of REgional dictionaries

Project Coordinator: Prof.dr. Nicoline van der Sijs (Nijmegen University)

2015-05-01 – 2015-12-31

Organisation	Total
Radboud University Nijmegen	€ 52,842
Total	**€ 52,842**

Abstract

Curation of various regional dictionaries of the Dutch language. Turning them into lexical databases. The project also lays a basis for connecting these databases to other regional databases such as eWND, DynaSand, and MIMORE. The curation of these regional dictionaries is crucial for the study of Dutch dialects and language variation in general.

15-005 OpenCGN

Incorporating CGN into OpenSONAR

Project Coordinator: Dr. Nelleke Oostdijk (Radboud University Nijmegen)

2015-05-01 – 2015-12-31

Organisation	Total
RUN	€ 39,865
Total	**€ 39,865**

Abstract

The OpenCGN project aims to provide access to the Spoken Dutch Corpus (CGN) through a web-based corpus exploitation environment, one which also gives access to the (written) SoNaR corpus. It will provide better empirical support for linguistic research by providing combined access to the SONAR written corpus and the Spoken Dutch Corpus.

Infrastructure Implementation Projects

10-002 TTNWW

TST tools voor het Nederlands als webservice in een workflow

2012-03-01 – 2012-09-30

Project coordinator: Ir. Marc Kemps-Snijders (Meertens Institute), drs. Ineke Schuurman (coordinator CLARIN Flanders)

Participant	Budget
Huygens ING	€10,000
Max Planck Institute for Psycholinguistics	€16,000
Meertens Institute	€74,000
University of Groningen	€70,000
Utrecht University	€30,000
Tilburg University	€75,000
Institute for Dutch Lexicology	€30,000
Radboud University Nijmegen	€25,000
University of Twente	€60,000
Aletta Institute for women's history	€5,000
Catholic Documentation Centre	€5,000
Total	**€400,000**

Abstract

The goal of the TTNWW project is to turn existing software components developed (inter alia) in the CGN project and the STEVIN programme into web services and to enable users to use these web services in a work flow system . It will run on servers of CLARIN centres. The goal is to offer this functionality to humanities researchers so that they can use it without requiring technical background or ad-hoc adaptations of data and/or software . The web services involve two modalities: text and speech.

09-001 MD

Metadataproject

2009-09-01 – 2010-03-31

Project coordinator: Ir. Daan Broeder (Max Planck Institute for Psycholinguistics)

Participant	Budget
Meertens Institute	€30,000
Max Planck Institute for Psycholinguistics	€15,000
Institute for Dutch Lexicology	€30,000
Total	**€75,000**

Abstract

In the CLARIN EU project a design was made for the so called CLARIN Metadata Infrastructure (CMDI) that should create a single interoperable domain of metadata descriptions for all resources housed at CLARIN centres. At the basis of this design is the concept of component metadata, these are bundles of metadata elements that describe related aspects of a resource. All metadata elements need to be

linked to a concept in the ISOCat Data Category Registry for semantic interoperability. Resource providers may reuse existing components or create their own and gather these into a metadata schema that they think best describes a specific resource type.
The CLARIN NL metadata project will, by using a preliminary rough implementation of a XML-schema and style-sheets toolkit (1) test the viability of this approach by describing the resources currently housed at two prospective CLARIN centres: INL and Meertens Institute. (2) It will provide a ready set of metadata components that can be used to seed the future CLARIN EU metadata component registry. (3) Provide best practice guidance for future users of the CMDI.

10-003 IIP

Infrastructuurimplentatieplan
2010-05-01 – 2013-05-01
Project coordinator: Ir. Daan Broeder (Max Planck Institute)

Participant	Budget
Meertens Institute	€192,000
Max Planck Institute for Psycholinguistics	€384,000
Institute for Dutch Lexicology	€192,000
Data Archiving and Networking Services	€30,000
Total	**€798,000**

Abstract
According to John Taylor an eScience scenario allowing researchers to make use of advanced features of the Internet can only be realized when they can make use of a new type of infrastructure. This new infrastructure should offer highly available and robust services to step away from the currently dominating download-first attitude. We can achieve this by offering a network of centers offering these services and taking care of data persistence and curation; developing a common and flexible framework of high quality metadata allowing easy discovery; setting up a federation based on single sign-on and single identity principles allowing researchers to easily navigate in such a domain and to create virtual collections; and by providing a joint framework for web services that allows researchers to easily include their smart algorithms and orchestrate them. All needs to be based on standards and best practices to achieve a high degree of integration of and interoperability between all types of language resources and tools. In this subproject we intend to implement (a first version of) such an infrastructure with MPI, INL, Meertens and DANS as initial centres.

12-003 CLARIN centre infrastructure Huygens ING

CLARIN centre infrastructure Huygens ING
2013-01-01 – 2013-10-01
Project coordinator: Dr. René Jongerius (Huygens ING)

Participant	Budget
Huygens ING	€52,974
Total	**€52,974**

Abstract
As a CLARIN-center, Huygens ING aims to enhance its eLaborate platform in 2012-13. This online work environment offers scholars tools for transcription, annotation and publication of texts.

Features like sharing metadata and archiving texts will be added to the current version. To make eLaborate fully CLARIN compliant a generic infrastructure is required for components like authentication, authorization, archiving and persistency. The implemented infrastructure will become the norm for new research projects regardless of the source of funding. As a CLARIN center Huygens ING thus strives for the structural growth of resources within the CLARIN domain.

10-004 S&D

Search and Develop

2010-05-01 – 2013-05-01

Project coordinator: Prof. dr. Hans Bennis (Meertens Institute)

Participant	Budget
Meertens Institute	€180,000
Max Planck Institute for Psycholinguistics	€216,000
Institute for Dutch Lexicology	€126,000
Data Archiving and Networking Services	€54,000
Total	**€576,000**

Abstract

This proposal describes a three-year program with the intention to develop a necessary functionality to allow a generic search on metadata and content. Such a project is ideal to achieve the required national cooperation between CLARIN centres as well, since it includes almost all the relevant infrastructural aspects. Moreover, it is in line with a proposal for a European project to develop a CLARIN European Demonstrator. In this project we intend to achieve three goals: (1) a generic search engine, (2) a national centre structure, and (3) a leading position in Europe.

MD4T

Metadata for Tools

2012-08-01 – 2013-02-01

Project coordinator: prof. dr. Jan Odijk (Utrecht University)

Participant	Budget
Utrecht University (UiL-OTS)	€18,000
Total	**€18,000**

Abstract

In the MD4T project CMDI metadata components and profiles will be created for describing software.

CLAM-S (11-032, 13-001, 13-013)

CLAM Support

2011-10-01 – 2014-12-31

Project Coordinator: prof. dr. Antal van den Bosch (Radboud University Nijmegen)

Participant	Budget
Radboud University Nijmegen	€18,400
Total	**€18,400**

Abstract
In CLAM-S support is given for the CLAM web service wrapper, and since 2013 also for the FoLia text corpus format.

ISOCAT Coordinator (11-033)

ISOCAT Coordinator
2011-03-01 – 2015-04-01
Project coordinator: prof. dr. Jan Odijk (Utrecht University)

Participant	Budget
Utrecht University (UiL-OTS)	€48,000
Total	**€48,000**

Abstract
The ISOCAT coordinator coordinates activities related to semantic interoperability. This includes the organisation of semantic interoperability documents. papers, tutorials, workshops as well as make recommendations to CLARIN-NL participants.

11-002 CLAVAS

CLARIN Vocabulary Service
2011-06-01 – 2012-05-31
Project coordinator: drs. Hennie Brugman (Meertens Institute)

Participant	Budget
Meertens Institute	€30,000
Total	**€30,000**

Abstract
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, CLAVAS built 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. The CLARIN Vocabulary Service is a running instance of the OpenSKOS exchange and publication platform for SKOS vocabularies. This OpenSKOS instance currently publishes SKOS versions of three vocabularies:

1. ISO-639-3 language codes, as published by SIL.
2. Closed and simple Data Categories from the ISOcat metadata profile.
3. A manually constructed and curated list of Organizations, based on the CLARIN *VLO*.

13-011 MCC

Metadata Content Coordinator
2014-01-01 – 2014-06-14
Project coordinator: dr. Marc Kemps-Snijders (Meertens Institute)

Participant	Budget
Meertens Institute	€17,000
Total	**€17,000**

Abstract
The quality of the harvested metadata is preventing users to use resource discovery tools efficiently. The lack of quality has many causes as (1) insufficient filled in information: lacking owner name, lacking content language etc. (2) spelling varieties and errors (3) lacking collection records i.e. no metadata record for the whole collection exists (4) the CMDI metadata schema used to describe the data are not optimal. A partial solution is expected from having a metadata content coordinator: a person that inspects the harvested CLARIN NL metadata, analyses this and gives feedback to the CLARIN centers providing the metadata about specific problems and makes general recommendations for improvement. Also the metadata content coordinator will liaise with the colleagues from the other national CLARIN projects to come to EU wide recommendations.

14-014 OpenSKOS

2014-07-14 – 2015-01-31
Project coordinator: drs. Hennie Brugman (Meertens Institute)

Participant	Budget
Meertens Institute	€35,635
Total	**€35,635**

Abstract
For CLARIN it has become imperative to provide an alternative for the current ISOcat software stack. The OpenSKOS software, previously developed in CATCH+, is considered a viable successor for the current ISOcat implementation. This project will deliver an implementation of the OpenSKOS software as a service taking into account the CLARIN requirements.

Centres of Expertise

CLARIN-NL has set up two (virtual) centres of expertise. Centres of expertise foster and further elaborate expertise that is considered important for the CLARIN infrastructure.

Data Curation Centre (10-025, 11-005)

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, ran, after a preliminary study in 2010, from November 2011 until the end of 2013. A guideline for the curation of data has been created, so that other projects can benefit from this. Here's a list of the data curated by DCS:

- LESLLA
- IPNV Interviews with veterans
- DBD/TCULT
- Woordenboek Gelderse Dialecten, Rivierengebied
- Woordenboek Gelderse Dialecten, Veluwe part)
- Curation organisation names for CLAVAS / OpenSkos
- Six dialect dictionaries from Brabant
- WLD and WBD part III
- Roots of Etnolect (collection Linda van Meel)
- Traces of Contact (collection Kofi Yakbo)

Participant	Budget
Radboud University Nijmegen	€174,807
Total	**€174,807**

Centre for the study of the Dutch language and culture across time (13-003)

Early in 2012 it was decided to consider Nederlab as the second (virtual) CLARIN Centre of Expertise. The *Nederlab* project is funded independently and aims to provide data and tools for the longitudinal study of the Dutch language and culture. It has also been supported financially by CLARIN-NL (from 2013-01-01 through 2014-12-31) and has been set up 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. Within the Nederlab project large amounts of historical data are curated, and their metadata created or curated. A dedicated search application has been constructed to search in the data, their metadata and in annotations on these data on multiple tiers. A first version of the Nederlab application was demonstrated at the CLARIN-NL final event, and extended

versions that incorporate parts of the multitier annotation search facilities at various workshops in 2015.

Participant	Budget
Meertens Institute	€150,000
Total	**€150,000**

User Survey

In order to get an idea about the current state of digitization in the humanities, (the use of digital tools and/or resources), the wishes and fears of the humanity scholars, and eventually the reasons why it sometimes took so long before digital initiatives took place, a User Survey was carried out in the first year of CLARIN.

This User Survey took the form of interviews with selected researchers from a wide variety of disciplines. The reason to do the interviews was because it was expected that most humanities researchers would have insufficient knowledge to assess the potential of the CLARIN infrastructure i.e. tools, resources, and services without any explanation of the ideas behind CLARIN. Besides this "evangelisation" of the CLARIN-idea, the interviews were an outstanding opportunity to inventory the (digital) needs and wishes of the humanity scholars, to get insight into their actual research questions. Moreover, we tried to get insight in their data needs: what data did they have, was it digital and if so in which format? Which data did they want to have? What did they do with "used data" (data used for an already finished PhD or published experiment)? Would they consider to deposit their data to a digital repository, and if so, under what conditions?

Therefore, this User Survey resulted in an overview of the available and desired data and tools for the humanity scholars and in a raised awareness of the CLARIN-NL project and the CLARIN infrastructure. Nine of the (app. 36) interviewed researchers actually got involved in one or more CLARIN-NL subprojects.

13-005 EduClarin (and 13-014, 14-005 – 14-009)

CLARIN and Education
2013-05-01 – 2014-12-31
Project coordinator: Dr. ir. G. Bloothooft (Utrecht University)

Participant	**Budget**
Utrecht University	€30,000
Total	**€30,000**

Abstract
When exploring the CLARIN-NL site(s), it becomes clear that access to project results is not easy. Although many projects may demonstrate the usefulness and possibilities of the CLARIN-NL programme, and stimulate students to further exploration, there is need for guidance and the development of educational modules that can be implemented in programmes of study. Such an action could imply:

(1) Identification of projects in CLARIN-NL that are well suited to be studied in the frame work of a Digital Humanities curriculum. Since students in these curricula may have various levels of back ground knowledge, special attention should be given to the requirements needed for an effective study and exploration of resources and tools.

(2) For each selected project, an introductory web chapter should be written, including suggestions for exploration. These could be, for instance, (1) step-by-step procedures that learn the contents and possibilities of the resource, (2) a series of questions to be studied with the help of the resource(s) and tools, with provided answers, and (3) more advanced assignments for small investigations.
The educational component of CLARIN-NL should be accessible to everyone as part of the CLARIN-NL website.

Six topics were selected for the development of an education package. The resulting packages are:

- The GrETEL Module Search in Treebanks: PDF; URL (link is external)
- The MIMORE Educational Module Linguistic Microvariation: PDF; URL (link is external)
- ePistolarium: URL (link is external) both in English (link is external) and in Dutch (link is external)
- Arthurian fiction (in Dutch): PDF; URL (link is external)
- Dutch Song Database (Liederenbank) (in Dutch); PDF
- Emblem project educational package

Events

Organized

2009

27 May	CLARIN-NL Kick-off Meeting	Utrecht, Netherlands
1 Jul	Infosessie CLARIN-NL Eerste Oproep	Nijmegen, Netherlands

2010

08 Jan	Relation Registry Workshop	Nijmegen, Netherlands
19 Feb	1e CLARIN-NL Meeting	Utrecht, Netherlands
25 March	ISOCAT tutorial	Utrecht, Netherlands
27 May	Metadata Tutorial	Nijmegen, Netherlands
26 Aug	CLARIN-NL Info Session Second Call	Amsterdam, Netherlands
21 Sept	ISOCAT workshop	Nijmegen, Netherlands
28 Oct	2e CLARIN-NL Meeting	Utrecht, Netherlands

2011

9 Feb	CLARIN-NL Call 2 Kick-off	Utrecht, Netherlands
10 May	ISOCAT Workshop	Utrecht, Netherlands
25 Aug	CLARIN-NL Call 3 Info Session	Utrecht, Netherlands
24 Oct	ISOCAT Workshop, part 2	Utrecht, Netherlands

2012

6 March	CLARIN Kick-off	Hilversum, Netherlands
20 March	ISOCAT Tutorial 2012 (I)	Utrecht, Netherlands
19 Jun	ISOCAT Tutorial 2012 (II)	Utrecht, Netherlands
13 Sept	CMDI Tutorial	Nijmegen, Netherlands
10 Oct	ISOCAT Tutorial 2012 (III)	Utrecht, Netherlands

2013

10 Feb	ISOCAT Workshop	Utrecht
21 Feb	Call 4 Kick-off Meeting	Utrecht
4 Jun	CMDI Interoperability Workshop	Utrecht
20 Jun	ISOCAT Tutorial	Utrecht
4 Sep	CLARIN-NL IAP Demonstration Event.	Utrecht
12 Sep	Semantic interoperability Workshop	Utrecht
14 Oct	CMDI Future Workshop	Utrecht

2014

13 Feb	'The CLARIN Infrastructure / GrETEL', MA Rotation Lecture	Utrecht
4 Mar	CMDI Tutorial	Utrecht
31 Mar	Searching for Constructions with GrETEL, Syntax Interface Meeting	Utrecht
23 Apr	Your Examples in GrETEL I	Utrecht
20 May	Your Examples in GrETEL II	Utrecht
23-27 Jun	LOT Summerschool CLARIN Course	Nijmegen
12 Aug	Tutorial Gabmap	Groningen
14 Oct	CLARIN and GrETEL: Lecture for the UU Digital Humanities Batchelor Course	Utrecht
17 Dec 14 - 20 Jan 15	Corpus Analysis Course	Utrecht

2015

12- 16 Jan	LOT Winterschool Course: linguistic research Using CLARIN	Amsterdam
13 Mar	CLARIN-NL Final Event and CLARIAH Kick-off	Hilversum
19 Mar	***Leren werken met een nieuwe zoektool voor de Short-Title Catalogue Netherlands***	Den Haag
13 Oct	Gastcollege DH Minor: CLARIN	Utrecht
10 Nov	MA-Rotation Lecture: CLARIN	Utrecht
Nov 2015 – Jan 2016	Corpus Analysis Course	Utrecht

Supported

2009

Date	Description	Location	Form of Support
18 Dec	Dag van de Fonetiek	Utrecht, Netherlands	presentation
18 Dec	Anéla Taalgala	Utrecht, Netherlands	presentation

2010

Date	Description	Location	Form of Support
11-22 Jan	LOT Winterschool	Leiden, Netherlands	Course on Linguistic

			Infrastructure and Microvariation
25-26 Jan	CLARIN Demonstrator Workshop	Amsterdam, Netherlands	Financial support for invited speakers
5 Feb	CLIN20	Utrecht, Netherlands	10k euro + plenary presentation on CLARIN-NL
4-5 Aug	Relish Workshop	Nijmegen, Netherlands	Funding travel and subsistence of some participants

2011

Date	Description	Location	Form of Support
17 Jan	CMDI developers meeting	Nijmegen, Netherlands	Subsidy of €228.50
9 May	Federated Search & EDC Workshop	Nijmegen, Netherlands	Subsidy of €3,000
19 May	Kick-off meeting VU-DNC	Amsterdam, Netherlands	Subsidy of €116.15
23 Jun	Workshop E-History	Amsterdam, Netherlands	Subsidy of €423.85

2012

Date	Description	Location	Form of Support
29 oct	CLARIN's Turn to the literary text	Den Haag, Netherlands	Subsidy of €2,761
31 Jan	Lustrumcongres Alfainformatica en oratie Gertjan van Noord	Groningen, Netherlands	Subsidy of €1,500
20 Jan	CLIN	Tilburg, Netherlands	Subsidy of €2500
30 Nov	A Shared LMF Core for Sign Language Lexicons	Ravenstein, Netherlands	Subsidy of €6840

2013

Date	Description	Location	Form of Support
12 Apr	CLARIN Literary Workshop	Amsterdam	financial
17 Jan	CLIN	Hengelo	financial

2014

Date	Description	Location	Form of Support
14 Jan	THATCamp	Den Haag	financial
17 Jan	CLIN	Leiden	financial

2015

Date	Description	Location	Form of Support
5 Feb	CLIN25	Antwerpen	financial
13 Apr	Workshop STCN	Den Haag	financial
25 Sep	DRONGO Festival	Utrecht	financial (via UU)
16 Nov	Workshop Morphosyntactische verrijking van 17e eeuwse teksten	Utrecht	Website, announcements, active participation
18 Dec	CLIN26	Amsterdam	financial

Attended

2009

30 Jan	DANS Digital Preservation Workshop	The Hague, Netherlands
4-5 Apr	CLARIN/FLaReNet Usage Scenario Meeting	Athens, Greece
11-12 May	CLARIN Consortium Meeting	Barcelona, Spain
4 Jun	KNAW Symposium Computational Humanities: Bridging the gap between the Humanities and the Computational Sciences	Amsterdam, Netherlands
15-16 Jun	Workshop Small Tools for Cross-Linguistic Research	Utrecht, Netherlands
30 Sep	CLARIN - FLaReNet workshop Standards in the LRT field	Helsinki, Finland
1-2 Oct	NEERI09 – interaction between research infrastructures on common topics	Helsinki, Finland
21-23 Oct	Working together to strengthen Research in Europe	Brussels, Belgium
2 Nov	SILT-FLaReNet Workshop Toward an Operational Definition of Interoperability for Language Technology	Waltham, US
5 Nov	CLARIN centres workshop	Prague, Czech Republic
2 Dec	Symposium - Door Data Gedreven, nieuwe inzichten in de wetenschap door innovatief gegevensgebruik	The Hague, Netherlands

2010

25-26 Jan	Workshop on CLARIN EDC (European Demonstrator Case)	Amsterdam, Netherlands
5 Feb	CLIN 20	Utrecht, Netherlands
16-23 May	LREC 2010	Valletta, Malta
11 Jun	Symposium "Toekomst voor Digitale Geesteswetenschappen"	Utrecht, Netherlands
17-18 Jun	NO-CLARIN	Oslo, Norway
28 Jun	Ed Hovy Symposium	Utrecht, Netherlands
1 Jul	DigiHist: Studiedag Digitalisering Historische Handschriften	Leuven, Belgium
7-8 Jul	Digital Humanities 2010	London, UK
13 Sep	BIG Grid User event	Amsterdam, Netherlands
18-22 Oct	SDH2010 + NEERI	Vienna, Austria
2 Dec	Conferentie Grootschalige onderzoeksfaciliteiten	Leiden, Netherlands

2011

17-18 Jan	CMDI workshop	Nijmegen, Netherlands
21 Jan	Oratie Paul Wouters	Leiden, Netherlands
11 Feb	CLIN	Gent, Belgium
20 Feb	DARIAH workshop	London, UK
08 March	Virtual Organization Workshop	Mannheim, Germany
28 March	EHumanities Symposium	Wassenaar Netherlands
30 March	Diachroon Corpus	The Hague, Netherlands

01 April	TTNWW workshop	Tilburg, Netherlands
07 April	SURFnet event about data management	Utrecht, Netherlands
12-13 April	EPIC User Forum	Amsterdam, Netherlands
09 May	EU workshop Federated Search	Nijmegen, Netherlands
18 May	SURFshare Open OnderzoeksData Dag	Amsterdam, Netherlands
26-27 May	FlareNet workshop	Venice, Italy
08 Jun	Dealing with data	The Hague, Netherlands
9-10 Jun	Federated Identity Management Workshop	Genève, Swiss
2-3 Sept	Peter Wittenburg International workshop	Kleve, Germany
7-9 Sept	Panel Digging into Data	London, UK
12-13 Sept	IRISC workshop	Helsinki, Finland
29 Sept	Alfalab Symposium	Amsterdam, Netherlands
10 Oct	Creatieve industrie en wetenschappelijk onderzoek	Amsterdam, Netherlands
1 Nov	CATCH meeting at HITIME	Amsterdam, Netherlands
5-19 Nov	IJCNLP Chiang Mai	Chiang Mai, Thailand
9 Dec	SARA 40	Amsterdam, Netherlands

2012

14 Dec	Oratie Rens Bod	Amsterdam, Netherlands
11 Dec	Humanities Lecture Frans Wieringa	Utrecht, Netherlands
3-4 Dec	CATCH+ / DEN Digital Heritage Conference	Rotterdam, Netherlands
29 Oct	CLARIN's Turn to the literary text	The Hague, Netherlands
24-28 Oct	CLARIN Integration Conference	Sofia, Bulgary
23 Oct	Kick-off CLICK-NL	Eindhoven, Netherlands
23 Oct	Digital Humanities Lecture Linguistic Research and The CLARIN Infrastructure	Utrecht, Netherlands
22-23 Oct	International Workshop on Open Access	Rotterdam, Netherlands
19 Oct	Kick-off Digital Humanities Amsterdam, KNAW	Amsterdam, Netherlands
18 Oct	PhD Defense Folkert de Vriend	Nijmegen, Netherlands
8 Oct	Presentation Project 'Gekaapte Brieven'	Rotterdam, Netherlands
3-4 Oct	SURF Relatiedagen	Noordwijkerhout, Netherlands
7 Sept	Afscheidscollege Geert Booij	Leiden, Netherlands
7 Sept	Symposium Geert Booij	Leiden, Netherlands
3 Sept	Opening Academisch jaar	Utrecht, Netherlands
3 Sept	Opening letterenbibliotheek	Utrecht, Netherlands
31 August	Presentation Network Cultural Heritage / CLICK-NL	Hilversum, Netherlands
30 August	CLARIN informatiesessie Call 4	Amsterdam, Netherlands
20-21 Jun	META-FORUM 2012	Brussels, Belgium
4 Jun	Workshop Open Access, Museon	The Hague, Netherlands
20-27 May	LREC	Istanbul, Turkey
23 March	Meeting with CLARIN-CZ/LINDAT (Jan Hajič, Eva	Prague, Czech Republic

	Hajičová)	
22 March	Lecture in Prague	Prague, Czech Republic
6 March	CLARIN Kick-off	Hilversum, Netherlands
31 Jan	Lustrumcongres Alfa-Informatica en oratie Gertjan van Noord en Johan Bos	Groningen, Netherlands
20 Jan	CLIN	Tilburg, Netherlands

2013

13 Jan	8th International Digital Curation Conference 2013	Amsterdam, Netherlands
17-18 Jan	CLIN	Enschede, Netherlands
23 Jan	The PoliMedia symposium: linking political debates and media	Amsterdam,Netherlands
21 Feb	CLARIN Kick-off	Utrecht, Netherlands
28 Feb	SURF Research and Innovation Event	The Hague, Netherlands
6 March	Route 2020: the Future for Large-Scale Research Infrastructures	Amsterdam, Netherlands
18 March	Research Data Alliance Launch and First Plenary	Götenborg, Sweden
21 mar	NIAS DH Workshop	Wassenaar, NL
10 Apr	eHumanities Address Software and Tools	Amsterdam, NL
12 Apr	CLARIN Literary Workshop	Amsterdam, NL
24 Apr	CLARIN Federated Content Search workshop	Copenhagen, DK
25 Apr	CLARIN Centre Assessment Workshop	Copenhagen, DK
17 May	OpenSKOS and controlled vocabularies meeting	Utrecht, NL
18 Jul	Language Diversity Congress	Groningen, NL
10 Sep	Open Research Data on the Horizon … and beyond - RDA Europe Workshop prior to the RDA Plenary in Washington	Garching, US
20 Sep	Digital Humanities Summerschool	Leuven, BE
30 Sep	Workshop: CLARIN, Standards and the Text Encoding Initiative	Rome, IT
17-18 Oct	DARIAH/DASISH Access and Authentication Infrastructures Workshop	Köln, DE
20-22 Oct	CLARIN ANNUAL Meeting	Prague, CZ
28-30 Oct	2nd EUDAT Conference (October 28-30, Rome)	Rome, IT
26-27 Nov	Annual Meeting General Assembly	Wrocław, PO

2014

14-15 Jan	THATCamp	Den Haag, NL
17 Jan	CLIN	Leiden, NL
10-11 Apr	'Mining Digital Repositories'	Den Haag, NL
25-31 May	LREC 2014	Reykjavik, IS
6 Jun	Bessensap	Utrecht, NL
23 Jun	CLARIN std. cie / ISO-TC37/SC4	Berlin, NL
23-Oct	CLARIN Conference	Soesterberg, NL
26 Nov	CLARIN ERIC General Assembly	Tübingen, DE
1-2 Dec	Clarino Meeting[35]	Tromsø, NO
8-9 Dec	Exploring Historical Sources with Language Technology: Results and Perspectives	Den Haag, NL
15-16 Dec	NeDiMAH	Den Haag, NL

[35] Jan Odijk is member of the CLARINO Advisory Board.

2015

5 Feb	CLIN25	Antwerpen, NL
7 Feb	TIN-dag	Utrecht, NL
24 Mar	Digitale Historische Kranten als Big Data	Den Haag, NL
26 Mar	CLARIN-TCC Creative Camp	Amsterdam, NL
1 Apr	Symposium *Luid zingend op een ijsschots de zomer tegemoet: de toekomst van de geesteswetenschappen*	Amsterdam, NL
9 Apr	Invited Lecture at the Austrian Centre for Digital Humanities	Vienna, AT
28 Apr	Digital Humanities Spring School (Keynote)	Leuven, BE
20-24 Sep	PARSEME Workshop	Iaşi, RO
25-26 Sep	DRONGO-festival	Utrecht, NL
15-17 Oct	CLARIN Conference	Wrocław, PO
30 Oct	CLARIN-TCC Creative Camp	Den Haag, NL
16 Nov	Workshop Morphosyntactische verrijking van 17e eeuwse teksten	Utrecht, NL
19-20 Nov	CLARIN General Assembly	Copenhagen, DK
18 Dec	CLIN26	Amsterdam, NL

Meetings

2009

19 Jan	CLARIN-NL preparatory committee	Utrecht, Netherlands
21 Jan	NGT	Utrecht, Netherlands
28 Jan	NWO	The Hague, Netherlands
Mar	Institut für Deutsche Sprache	Mannheim, Germany
11 Mar	MPI	Nijmegen, Netherlands
15 Apr	DANS	Utrecht, Netherlands
19 May	Alfalab	Utrecht, Netherlands
17 Sep	NWO	The Hague, Netherlands
6 Oct	Geleerdenbrievenproject (CKCC)	Utrecht, Netherlands
9 Oct	EWI	Brussels, Belgium
9 Dec	Ministerie van OCW on ERIC	The Hague, Netherlands

10 Dec	ISOCAT Meeting	Nijmegen, Netherlands
10 Dec	INL	Leiden, Netherlands

2010

05 Jan	KNAW / CLARIN-NL	Amsterdam, Netherlands
28 Jan	LOT-Bestuursvergadering	Utrecht, Netherlands
12 Feb	CLARIN – T4ME (META-NET) Meeting (at FLaReNet Forum}	Barcelona, Spain
17 March	Meeting with Bart Besamusca	Utrecht, Netherlands
26 April	Huygens	The Hague, Netherlands
28 May	Preparation Symposium “Toekomst Digitale GeestesWetenschappen”	Utrecht, Netherlands
02 Jun	IIP Meeting	Utrecht, Netherlands
23 Jun	eData & research Meeting (Martijn de Groot)	Utrecht, Netherlands
24 Jun	MIMORE Meeting	Utrecht, Netherlands
30 Jun	SURFNet Meeting	Utrecht, Netherlands
24 Aug	DEN Meeting	The Hague, Netherlands
01 Sept	ADEPT Meeting	Groningen, Netherlands
08 Sept	CLARIN NTU ERIC	Roosendaal, Netherlands
16 Sept	Meeting with Wijnand Mijnhardt about DGO	Utrecht, Netherlands
22 Sept	Meeting with José de Kruif	Utrecht, Netherlands
23 Sept	IIP + S&D Meeting	Utrecht, Netherlands
04 Oct	TDS Project	The Hague, Netherlands
06 Oct	Libratory	Leiden, Netherlands
13 Oct	Metadata ISO	Berlin, Germany
26 Oct	DGO	Utrecht, Netherlands
09 Nov	CMDI-DEN meeting	Nijmegen, Netherlands
10 Nov	Beeld & Geluid	Hilversum, Netherlands
22 Nov	DGO	Utrecht, Netherlands
25 Nov	IIP + S&D Meeting	Utrecht, Netherlands
8 Dec	Meeting with Steven Krauwer	Utrecht, Netherlands

2011

19 Jan	Tweede Kamer	The Hague, Netherlands
19 Jan	Meeting CLARIN European Demonstrator Case	Nijmegen, Netherlands
26 Jan	Prep ERIC SC	The Hague, Netherlands
15 Mar	ED&R Bestuur	The Hague, Netherlands
17-18 Mar	CLARIN ERIC SC	The Hague, Netherlands
24 Mar	CLARIN-NL interview	Utrecht, Netherlands
29 Mar	eHumanities meeting	Wassenaar, Netherlands
6 April	TAB Taalportaal	Leiden, Netherlands
15 April	ERIC SC 2	The Hague, Netherlands

27 April	ED&R Bestuur	The Hague, Netherlands
28 April	CLARIN ERIC meeting with NTU (a.o.)	Roosendaal, Netherlands
4 May	EC meeting	Brussels, Belgium
13-15 Jun	TC37/SC4 meeting	Seoul, Korea
16 Jun	KNAW Meeting	Amsterdam, Netherlands
21 Jun	CLARIN-NL / OCW-NWO	The Hague, Netherlands
22 Jun	CLARIAH Meeting	Amsterdam, Netherlands
28-30 Jun	Meeting of CLARIN National coordinators	Budapest, Hungary
11 Aug	CLARIAH meeting	Amsterdam, Netherlands
23-24 Aug	CMDI dev meeting	Nijmegen, Netherlands
26 Aug	CLARIAH meeting	Amsterdam, Netherlands
12 Sept	IPROSLA meeting	Nijmegen, Netherlands
15 Sept	Meeting	Amsterdam, Netherlands
21 Sept	Prep meeting Creatieve Industrie	Nijmegen, Netherlands
28 Sept	Meeting Nationaal Archief	The Hague, Netherlands
29 Sept	Final presentation Alfalab	Amsterdam, Netherlands
1-31 Oct	Many meetings with several applicants about Nederlab	Netherlands
10 Oct	Meeting about Creatieve Industrie	Amsterdam, Netherlands
14 Oct	Meeting UBU	Utrecht, Netherlands
14 Oct	Valerie Frissen Creatieve Industrie	The Hague, Netherlands
17 Oct	Meeting about digitization and CLARIN	The Hague, Netherlands
31 Oct	Meeting Raad van Toezicht Catch+	Amsterdam, Netherlands
31 Oct	CLARIAH meeting	Utrecht, Netherlands
10 Nov	INPOLDER meeting	Amsterdam, Netherlands
11 Nov	Meeting about Creatieve Industrie	Amsterdam, Netherlands
17 Nov	Meeting about The Language Archive	Amsterdam, Netherlands
20 Nov	CLARIN- NL/ DE Meeting	Utrecht, Netherlands
23 Nov	ArthurianFiction meeting	The Hague, Netherlands
28 Nov	Meeting about Centre for Digital Humanities	Amsterdam, Netherlands
30 Nov	CLARIAH Meeting	Utrecht, Netherlands
2 Dec	Talk with Rijksdienst voor Cultureel Erfgoed	Amsterdam, Netherlands
9 Dec	CLARIAH Meeting	The Hague, Netherlands
12 Dec	Reviews for Digital Humanities 2012	Amsterdam, Netherlands
13 Dec	Meeting Stuurgroep Catch+	The Hague, Netherlands
15 Dec	Meeting about The Language Archive	Amsterdam, Netherlands
19 Dec	Meeting about Early Dutch Books on Line	Amsterdam, Netherlands
21 Dec	Meeting about Dienstenmodel Catch+	The Hague, Netherlands

2012

18 Dec	Meeting CLARIN educatie	Utrecht, Netherlands

11 Dec	Bijeenkomst 'Technologie & digitalisering in de wetenschap' op Ministerie van OCW	The Hague, Netherlands
30 Oct	Digital Humanities curriculum	Utrecht, Netherlands
23 Oct	FESLI meeting	Amsterdam, Netherlands
19 Oct	Brainstorming PMMD conferentie	Amsterdam, Netherlands
16 Oct	Meeting Leo Lentz	Utrecht, Netherlands
16 Oct	Meeting software metadata	Utrecht, Netherlands
26 Sept	Ambassadeursbijeenkomst	Utrecht, Netherlands
14 Sept	FESLI meeting	Amsterdam, Netherlands
12 Sept	Meeting CLARIAH evaluation	Groningen, Netherlands
3 Sept	CLARIN ERIC celebration	Utrecht, Netherlands
20 Sept	Presentation SURF report	Utrecht, Netherlands
27 August	Meeting Computational Humanities	Amsterdam, Netherlands
22 August	ISOCAT meeting	Utrecht, Netherlands
21 August	Meeting Bountouridis	Utrecht, Netherlands
14 August	Meeting Metadata4Tools	Utrecht, Netherlands
14 August	Meeting UU financial dept.	Utrecht, Netherlands
9 August	Meeting UBU	Utrecht, Netherlands
26 Jun	FESLI meeting	Amsterdam, Netherlands
12 Jun	Taalportaal Meeting	Leiden, Netherlands
7 Jun	NWO bijeenkomst over call CI Topsector	Amsterdam, Netherlands
5 Jun	Cornetto-LMF-RDF Kickoff meeting	Amsterdam, Netherlands
30 May	Clariah Meeting	The Hague, Netherlands
16 May	Meeting UBU	Utrecht, Netherlands
8 May	Clariah Meeting	The Hague, Netherlands
18-19 April	CLARIN ERIC GA en Coördinators meeting	The Hague, Netherlands
12 April	Meeting Els Kloek	The Hague, Netherlands
5 April	GrNe Kick-off	Leiden, Netherlands
4 April	Meeting Huygens ING IVM Centre plan	Utrecht, Netherlands
23 March	Meeting with CLARIN-CZ / LINDAT	Prague, Czech Republic
9 March	Nederlab Site Visit	Amsterdam, Netherlands
2 March	Infrastructuur toekenningsbijeenkomst	Utrecht, Netherlands
18 Jan	GrNe bespreking	Utrecht, Netherlands
18 Jan	PILNAR bespreking	Utrecht, Netherlands
9 Jan	Clariah Interview	Utrecht, Netherlands
4 Jan	Clariah Meeting	The Hague, Netherlands

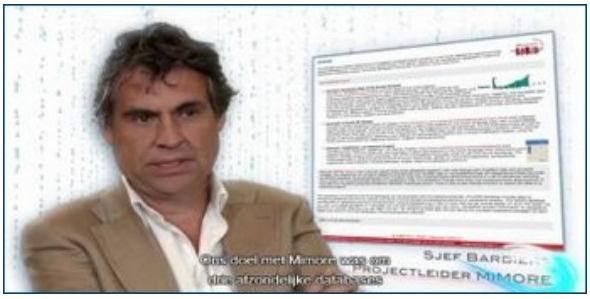
2013-2015

No records systematically kept

Media

Movies

Under the authority of CLARIN-NL 12 movies have been developed in which projects are explained.



MIMORE

WFT-GTB

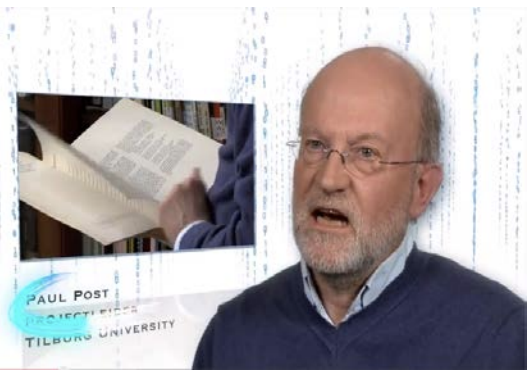
S&D

SignLinC

Arthurian Fiction

WIP

Geleerdenbrieven



PilNar

-

[PoliMedia](#) [D-Lucea](#)

The movies on DSS and OpenSONAR are planned to be available as of Mid 2016.

See http://www.clarin.nl/node/403

Several subprojects have made movies of their own:

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

Newsflash

Every year several NewsFlashes have been issued:

- NewsFlash 1 – 2009 June 23th
- NewsFlash 2 – 2009 July 2nd
- NewsFlash 3 – 2009 August 13th
- NewsFlash 4 – 2009 December 15th
- NewsFlash 5 – 2010 January 6th
- NewsFlash 6 – 2010 March 5th
- NewsFlash 7 – 2010 April 27th
- NewsFlash 8 – 2010 July 6th
- NewsFlash 9 – 2010 August 30th

- NewsFlash 10 – 2010 September 23th
- NewsFlash 11 – 2010 October 13th
- NewsFlash 12 – 2010 November 25th
- NewsFlash 13 – 2010 December 21th
- NewsFlash 14 – 2011 January 26th
- NewsFlash 15 – 2011 April 21th
- NewsFlash 16 – 2011 July 1st
- NewsFlash 17 – 2011 September 5th
- NewsFlash 18 – 2011 December 23th
- NewsFlash 19 – 2012 February 24th
- NewsFlash 20 – 2012 April 26th
- NewsFlash 21 – 2012 June 18th
- NewsFlash 22 – 2012 July 2nd
- NewsFlash 23 – 2012 August 16th
- NewsFlash 24 – 2012 September 28th
- NewsFlash 25 – 2012 October 24th
- NewsFlash 26 – 2012 December 21th
- NewsFlash 27 – 2013 February 11th
- NewsFlash 28 (16-07-2013)
- NewsFlash 29 (28-01-2014)
- NewsFlash 30 (23-05-2014)
- Newsflash 31 (??)
- NewsFlash 32 (12 jan 2015)
- NewsFlash 33 (13 feb 2015)
- NewsFlash 34 (6 mrt 2015)
- NewsFlash 35 (9 sep 2015)
- NewsFlash 36 (21 okt 2015)

Press

2010

[WFT-GTB] Wurdboek van de Fryske Taal online", article at website 'De Nederlandse Vereniging', September 6, 2010.

[WFT-GTB]"Woordenboek der Friese taal online", article at Kennislink, July 16, 2010.

[WFT-GTB] "Wetenschappelijk Woordenboek van de Friese taal", article at 'Blog van de Historische Vereniging Noordoost-Friesland' by Hans Zijlstra, July 7, 2010.

[WFT-GTB]"40.000 kaartjes voor het woordje 'de' staan online", article about WFT-GTB in Friesch Dagblad, July 6, 2010.

[WFT-GTB] "25-dielich Wurdboek fan de Fryske Taal online", article at 'Hallo Fryslân', July 6, 2010.

[WFT-GTB] “Het laatste grote woordenboek op papier?”, article at NRC website by Ewoud Sanders, July 5, 2010.

'Naar een betere toegang tot tekst- en spraaktools'. Article by Marc Kemps-Snijders en Ineke Schuurman in E-Data & Research 5.1, 2010. [URL]

2011

[GABMAP] “Het beeld en de streektaal”, article about the subproject GABMAP by Erica Renckens in E-Data & Research 6(2), p.3.

[CLARIN-NL] "CLARIN-NL - Common Language Resources and Technology Infrastructure in Nederland", focus article by Inge Angevaare in E-Data & Research, 6(1), June 2011, p.6.

[TICCLops] “Oude teksten bruikbaar voor nieuw onderzoek”, article about the subprojects TICCLops and Adelheid by Erica Renckens in E-Data & Research 6(1), June 2011, p.3.

[WIP] "Dat doet denken aan de oorlog", article about the subproject WIP by Erica Renckens in E-Data & Research 6(2), October 2011, p. 1 [PDF].

[CLARIN-NL] "Het digitale drama", article about digitization in the Netherlands by Karel Berkhout in NRC, September 10, 2011 [URL] and various responses, in particular the response by Jan Odijk.

[CLARIN-NL] "CLARIN-NL - Common Language Resources and Technology Infrastructure in Nederland", focus article by Inge Angevaare in E-Data & Research, 6(1), June 2011, p.6. [URL] [PDF].

[WIP] "Oorlog in de Kamer", article about the subproject WIP (War in Parliament) by Maarten Marx in NRC, March 3, 2011 [URL].

[CLARIN-NL] "Negen nieuwe CLARIN projecten van start", article by Erica Renckens in E-Data & Research 5(4), March 2011, p. 7 [URL] [PDF].

[CLARIN-NL] "CLARIN-NL deelnemer in eData & Research", article by Erica Renckens in E-Data & Research 5(4), March 2011, p. 1 [URL] [PDF].

[WIP] "Waarom politici graag over de oorlog praten", article about the project WIP by Bart Funnekotter in NRC, February 25, 2011 [URL] [pdf] [WIP website].

2012

[CLARIN-NL] “De vraag van morgen moet je vandaag kunnen beantwoorden”, interview with Peter Wittenburg by Heidi Berkhout and Erica Renckens in E-Data & Research 7(1), October 2012, p.5.

[CLARIN-NL] “Alle teksten onder één virtuleel lab”, article by Erica Renckens in E-Data & Research 7(1), October 2012, p.6.

[CLARIN-NL] “Nog meer slimme tools voor geesteswetenschap”, article by Peter Boot in E-Data & Research 6 (3), February 2012, p.2.

[PILNAR] "Wat bezielt de pelgrim", article by Erica Renckens in E-Data & Reserach 7(1), October 2012, p.3 [URL] [PDF].

[Arthurian Fiction] "Manuscripten koning Arthur op tafel", article by Peter Boot in E-Data & Research 7(1), October 2012, p.1 [URL] [PDF].

[DCS] "Een mooi tweede leven voor oude data", article by Erica Renckens in E-Data & Research 6(4), June 2012, p.1 [URL] [PDF].

[CLARIN-NL] "Variatie in derde ronde CLARIN-NL", article by Erica Renckens in E-Data & Research 6(4), June 2012, p.2 [URL] [PDF].

[CLARIN-NL] "DASISH brengt ESFRI initiatieven bij elkaar", article by Eric Balster in E-Data & Research 6(4), June 2012, p.6 [URL] [PDF]

[CLARIN-NL] "CLARIN krijgt ERIC -status", article by Erica Renckens in E-Data & Research 6(3), February 2012, p.6 [URL] [PDF].

[WFT-GTB] "Wurdboek in Taalbank INL", article by Erica Renckens in E-Data & Research 6(3), February 2012, p.6 [URL] [PDF].

[VK] "Zoeken door de hele Loe de Jong", article by Peter Boot in E-Data & Research 6(3), February 2012, p.3 [URL] [PDF].

2013

[CLARIN-NL] “Workshop CLARIN over relatie met literatuurstudie”, article by Peter Boot in E-Data & Research 7(2), February 2013, p.2.

[VU-DNC] “Online krantencorpus maakt vergelijkend onderzoek mogelijk”, article about subproject VU-DNC by Erica Renckens in E-Data & research, February 2013, p.3.

[CLARIN-NL] “Archieven Beeld en Geluid opengesteld”, article by Erica Renckens in E-Data & Research, February 2013, p.6.

[CLARIN-NL] "Bouwen aan een grootse toekomst", article by Erica Renckens in E-Data & Research 8(1), October 2013, p.2 [URL] [PDF].

[CLARIN-NL] "Taalportaal bundel grammaticale kennis", article by Erica Renckens in E-Data & Research 8(1), October 2013, p.7 [URL] [PDF].

[GrNe]"Een modern woordenboek voor het Oudgrieks", article by Erica Renckens in E-Data & Research 8(1), October 2013, p.8 [URL] [PDF]

[CKCC] "Online grasduinen in brieven van geleerden", article by Erica Renckens in E-Data & Research 7(3), June 2013, p.1 [URL] [PDF].

[CLARIN-NL] "Online krantencorpus maakt vergelijkend onderzoek mogelijk", article by Erica Renckens in E-Data & Reserach 7(2), February 2013, p.3 [URL] [PDF].

2014

[CLARIAH] Searching beyond the borders of topic and time. An integrated approach for data and tools in the humanities'. 'Article by Patricia Alkhoven in *eData & Research* Special issue 2014 p 3-3.

[CLARIN-NL] 'We moeten nog leren werken met Big Data'. Article by Inge Angevaere in *eData & Research*, 8.3, p. 5-5 Jun 2014

[Nederlab] 'Onderzoekers geven verlanglijst aan Nederlab', Article by Peter Boot in *eData & Research*, 8.2, p. 2-2 Feb 2014.

[CLARIN-NL] 'Eerste Nederlandse THATCamp blijkt leerzame unconference. Article by Peter Boot in *eData & Research*, 8.2, p. 6-6 Feb 2014.

[COBWWWEB] 'Over vrouwen en het literaire veld'. Article by Suzan van Dijk, & W. Ravenek in *eData & Research*, 8.3, p. 6-6 Jun 2014

[CLARIN-NL] 'Why the Social Sciences and Humanities need research infrastructures'. Article by Peter Doorn in *eData & Research* Special issue 2014 p. 1-1

[CLARIN-NL] 'e-Humanities ook voor historische taalkunde'. Article by Erica Renckens in *eData & Research*, 8.3, p. 2-2 Jun 2014

[CLAPOP] 'CLAPOP toont resultaten CLARIN-NL'. Article by Erica Renckens in *eData & Research*, 8.3, p. 7-7 Jun 2014 [URL]

[CLARIN-NL] 'Digital Humanities cursussen op een rij'. Article by Erica Renckens in *eData & Research*, 8.3, p. 7-7 Jun 2014

[Nederlab] 'Nederlab', Article by Nicoline van der Sijs in *eData & Research* Special issue 2014, p. 3-3

[CLARIN-NL] "Hoe Honkvast zijn Groningers", article by Erica Renckens in E-Data & Research 8(3), June 2014, p. 7. [URL] [PDF]

[CLARIN-NL] "Voor al haar mantelzorgen", article by Ewoud Sanders in NRC, 14 April 2014 [PDF].

[CLARIN-NL] "Online gateway to language resources", article by Jan Odijk in E-Data & Research Special Issue, 2014, p.2 [URL] [PDF].

[CLARIAH] "Searching beyond the borders of topic and time", article by Patricia Alkhoven in E-Data & Research Special Issue, 2014, p.3 [URL] [PDF].

[WAHSP] "WASHP-tool peilt sentiment in twee talen", article by Erica Renckens in E-Data & Research 8(2), February 2014, p.3 [URL] [PDF].

[Namescape] "Letterkundig namenlandschap", article by Peter Boot in E-Data & Research 8(2), February 2014, p.7 [URL] [PDF].

[Polimedai] "Eerste prijs voor online tool met linked data", article by Erica Renckens in E-Data & Research 8(2), February 2014, p.8 [URL] [PDF].

2015

[SHEBANQ] 'Online SHEBANQ wint DH-award'. Article by Erica Renckens in eData & Research, 9.3 June 2015. [URL]

[CLARIN-NL] 'Educatieve modules CLARIN-NL'. Article by Erica Renckens in eData & Research, 9.3 June 2015 [URL]

[DSS] 'Duik in de maritieme data: CLARIN-NL project Dutch Ships and Sailors koppelt en verrijkt'. Article by V. de Boer, M. van Rossum, J. Leinenga en R. Hoekstra in eData & Research, 9.3, June 2015.[URL]

[CLARIN-NL] 'In het woud van tools en data'. Article by Jan Odijk and Daan Broeder in eData & Research, 9.2, Feb 2015.[URL]

[Rembench] 'Alles over Rembrandt digitaal doorzoekbaar: RemBench laat online mogelijkheden kunsthistorische data zien' Article by Ilja Nieuwland in eData & Research 9.2, Feb 2015. [URL]

[CLARIN-TCC] 'Nieuwe Tools voor Europese Data: Aan de slag met Linked Data tijdens Creative Camps'. Article by Erica Renckens in eData & Research 9.2, Feb 2015 [URL]

Publications

2010

Broeder, D. (2010), "Use of SLICS in the CLARIN use case", abstract and poster at EGEE Uppsala, April 12, 2010.

Broeder, D., Kemps-Snijders, M., Van Uytvanck, D., Windhouwer, M., Withers, P., Wittenburg, P., Zinn, C. (2010). "A data category registry- and component-based metadata framework." In Calzolari, N., Maegaard, B., Mariani, J., Odijk, J., Choukri, K., Piperidis, S., Rosner, M. and Tapias, D. (eds), Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC 2010), Valetta, Malta: European Language Resources Association (ELRA), pp. 43-27.

Heeringa, W. (2010), "Gabmap", a practical at the Vrije Universiteit Amsterdam, as part of the course 'Inleiding taalvariatie - Haagse Harrie, Mehmet Pamuk en hun kornuiten, September 29, 2010.

Heuvel, H. van den, Horik, R. van, Scagliola, S., Sanders and E. Witkamp, P. (2010), "The Veteran Tapes: Research Corpus, Fragment Processing Tool, and Enhanced Publications for the e-Humanities". In Calzolari, N., Maegaard, B., Mariani, J., Odijk, J., Choukri, K., Piperidis, S., Rosner, M. and Tapias, D. (eds), Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC 2010), Valetta, Malta: European Language Resources Association (ELRA).

Heuvel, H. van den (2010), "Interviews, geknipt voor onderzoek". In: DIXIT 2010, jg 7, nr 1. pp. 20-21.

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

Odijk, J. (2010), "The CLARIN-NL Project", CLARIN Newletter 9-10, March-JUne 2010, pp 3-4. http://www.clarin.eu/newsletter/3280

Odijk, J. (2010), 'Het CLARIN-NL project', DIXIT 7(1) dec 2010, pp. 32-33.

Roorda, D., Bos, E-J. and Van den Heuvel, C.M.J.M. (2010), “Letters, Ideas and Information Technology: Using digital corpora of letters to disclose the circulation of knowledge in the 17th century”, paper/presentation held at Digital Humanities 2010, London, July 9, 2010.

Sijens, H. and Depuydt, K. (2010), "Wurdboek fan de Fryske taal/Dictionary of the Frisian Language online: new possibilities, new opportunities", in: Dykstra, A. and Schoonheim, T. (eds.), Proceedings of the XIV Euralex International Congress, Leeuwarden, 2010, pp. 726-732.

Van Uytvanck, D., Zinn, C., Broeder, D., Wittenburg, P., Gardelleni, M. (2010). "Virtual language observatory: The portal to the language resources and technology universe. In Calzolari, N., Maegaard, B., Mariani, J., Odijk, J., Choukri, K., Piperidis, S., Rosner, M., and Tapias, D. (eds), Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC 2010), European Language Resources Association (ELRA), pp. 900–903.

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.

2011

Cornips, L., Kemps-Snijders, M. Snijders, M., Swanenberg, J. and Vriend, F. de (2011), "Bridging the Gap between First Language Acquisition and Historical Dialectology with the Help of Digital Humanities". In: Proceedings of the 2nd Conference on Supporting Digital Humanities (SDH 2011).

Nerbonne, J., Colen, R., Gooskens, C., Kleiweg, P and Leinonen, T. (2011), "Gabmap — A Web Application for Dialectology". In: Nerbonne, J., Grondelaers, S., Speelman, D. and Perea, M. (eds), Dialectologia, Special Issue II (Spec. Iss. Production, Perception and Attitude), pp.65-89.

Odijk, J. (2011), "CLARIN Developments in the Countries: the Netherlands", CLARIN NewsFlash March 2011, p. 2.

Odijk, J. (2011), "User Scenario Search", internal CLARIN-NL document, April 13, 2011.

Odijk, J. (2011). "CLARIN-NL Story: How to successfully convince your government", CLARIN Newsletter 13 (June 2011), p. 5.

Odijk, J. and A. van Hessen (2011), "Sharing Resources in CLARIN-NL", Proceedings of the Language Resources, Technology and Services in the Sharing Paradigm workshop at IJCNLP 2011, pp. 98-106 (November 2011), Chiang Mai Thailand.

Snelders, S., Pieters, T. and Laan, F. (2011), "Lekensamenvatting WAHSP". Internal document.

Wittek, P. and Ravenek, W. (2011), “Supporting the Exploration of a Corpus of 17th-Century Scholarly Correspondences by Topic Modeling”. In: B. Maegaard (Ed.), Supporting Digital Humanities 2011: Answering the unaskable. Copenhagen, Denmark.

2012

Broeder, D., Uytvanck, D. van, Gavrilidou, 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.

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.

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.

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.

Juric, D., Hollink, L., & Houben, G. (2012). Bringing parliamentary debates to the Semantic Web. DeRiVE workshop on Detection, Representation, and Exploitation of Events in the Semantic Web.

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.

Odijk, J.E.J.M. (2012, October 19), Proposal for a Usage Scenario and Theme for presentation at the PMMD conference, Amsterdam.

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.

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.

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.

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.

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.

2013

Boer, Victor de, Johan van Doornik, Lars Buitinck, Maarten Marx, Tim Veken, Kees Ribbens. Linking the Kingdom: Enriched Access To A Historiographical Text. Proceedings of KCAP 2013, Banff, Canada, 23-26 June 2013. [PDF]

Eijnatten, J. van, Pieters, T., & Verheul, J. (2013). Big data for global history: The transformative promise of digital humanities. *BMGN-Low Countries Historical Review*, *128*(4).

Halteren, H. van & Rem, M. (2013). Dealing with orthographic variation in a tagger-lemmatizer for fourteenth century Dutch charters. *Language resources and evaluation*, *47*(4), 1233-1259.

Huijnen, P., Laan, F., De Rijke, M., & Pieters, T. (2013). A digital humanities approach to the history of science. In *Social Informatics* (pp. 71-85). Springer Berlin Heidelberg.

Juric, D., Hollink, L., & Houben, G. (2013). Discovering links between political debates and media. The 13th International Conference on Web Engineering (ICWE'13). Aalborg, Denmark.

Kemman, M. J., & Kleppe, M. (2013). PoliMedia - Improving Analyses of Radio, TV & Newspaper Coverage of Political Debates. In T. Aalberg & E. Al. (Eds.), TPDL2013, LCNS 8092 (pp. 409-412). Springer-Verlag Berlin Heidelberg.

Kemman, M. J., Kleppe, M., & Maarseveen, J. (2013). Eye Tracking the Use of a Collapsible Facets Panel in a Search Interface. In T. Aalberg & E. Al. (Eds.), TPDL2013, LCNS 8092 (pp. 405-408). Springer-Verlag Berlin Heidelberg.

Odijk, J. (2013) 'Common Lab Research Infrastructure for the Arts and Humanities (CLARIAH), in *Unchartered Frontiers: the Netherlands' Roadmap for Large-Scale Research Facilities*, publication of the Dutch Ministry of Education, Culture and Science, p. 17.

Odijk, J. (2013). DUELME: Dutch Electronic Lexicon of multiword Expressions. In G. Francopoulou (Eds.), *LMF - Lexical Markup Framework* London, UK / Hoboken, US: ISTE / Wiley.

Oostdijk, N., van den Heuvel, H., & Treurniet, M. (2013). The CLARIN-NL Data Curation Service: Bringing Data to the Foreground. *International Journal of Digital Curation*, *8*(2), 134-145.

Rooij, O. de, Kenter, T., & de Rijke, M. xTAS and ThemeStreams. *proceedings dir 2013*, 2.

Schlieper, H. (2013). International Workshop Report: Collaborating in Networks: Preparing Future Projects, Westfalische Wilhelms-Universitat Munster, Germany, 3-5 April 2013. *J. Res. Gender Stud.*, *3*, 157.

Sijs, Nicoline van der (2013) 'Open Dialectendata', NRC 2013-05-11. [pdf]

Sijs, N. van der (2013), 'Sluitdagen en afpakschuur: de Indisch-Nederlandse woordenschat', in: Over tienduizend dingen. Bijdragen voor Reinier Salverda, 273-286.

Sijs, N. van der (2013), 'De digitalisering van het Nederlandstalige erfgoed', in: Nederland in ideeën. 101 denkers over inzichten en innovaties die ons land verander(d)en, red. Mark Geels en Tim van Opijnen, Amsterdam: Maven Publishing, 130-133.

Schuth, A., Hofmann, K., Whiteson, S., & de Rijke, M. (2013, November). Lerot: An online learning to rank framework. In *Proceedings of the 2013 workshop on Living labs for information retrieval evaluation* (pp. 23-26). ACM.

Stipriaan, R. van (2013), '5949 dagen Groot-Nederland: sleutelen aan een collectieve identiteit tussen 1815-1830, en daarna', in: Ons Erfdeel 3, 30-37.

Van Uytvanck, D. (2013) 'ICT Tools enable eHumanities top-research', in Unchartered Frontiers: the Netherlands' Roadmap for Large-Scale Research Facilities, publication of the Dutch Ministry of Education, Culture and Science, p. 18-19.

Wright, S. E., Windhouwer, M., Schuurman, I., & Kemps-Snijders, M. (2013). Community efforts around the ISOcat Data Category Registry. In *The People's Web Meets NLP* (pp. 349-373). Springer Berlin Heidelberg.

Wynne, M. (2013). The role of CLARIN in digital transformations in the humanities. *International Journal of Humanities and Arts Computing*, *7*(1-2), 89-104.

Zaagsma, G. (2013). On digital history. *BMGN-Low Countries Historical Review*, *128*(4).

2014

Amigó, E., Carrillo-de-Albornoz, J., Chugur, I., Corujo, A., Gonzalo, J., Meij, E., ... & Spina, D. (2014). Overview of RepLab 2014: author profiling and reputation dimensions for Online Reputation Management. In *Information Access Evaluation. Multilinguality, Multimodality, and Interaction* (pp. 307-322). Springer International Publishing.
http://nlp.uned.es/~damiano/pdf/replab2014overview.pdf

Baltussen, L. B., Blom, J., Medjkoune, L., Pop, R., Van Gorp, J., Huurdeman, H., & Haaijer, L. (2014). Hard Content, Fab Front-End: Archiving Websites of Dutch Public Broadcasters. *Alexandria*, *25*(1), 69-91.

Boer, Victor de, Jur Leinenga, Matthias van Rossum and Rik Hoekstra. 'Dutch Ships and Sailors Linked Data Cloud', Accepted for publication in *Proceedings of the International Semantic Web Conference (ISWC 2014)*, 19-23 October, Riva del Garda, Italy, 2014. [Draft]

Bravo Balado, A. *Information extraction on newspaper archives for historical research. A Dutch maritime history case study*. M.Sc. thesis VU University Amsterdam (forthcoming), 2014.

Broeder, D. *et al.* Experiences with the ISOcat Data Category Registry. *Proceedings of LREC 2014*, Reykjavik, Iceland. [URL]

Broeder, D., & Lannom, L. W. (2014). Data Type Registries: A Research Data Alliance Working Group. D-Lib Magazine, 20(1), 2. http://dialnet.unirioja.es/servlet/articulo?codigo=4572541&orden=1&info=link

Buccio, E. D., Nunzio, G. M. D., & Silvello, G. (2014). A linked open data approach for geolinguistics applications. International Journal of Metadata, Semantics and Ontologies, 9

Cai, F., Liang, S., & de Rijke, M. (2014, July). Personalized document re-ranking based on bayesian probabilistic matrix factorization. In Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval (pp. 835-838). ACM. http://ilps.science.uva.nl/sites/ilps.science.uva.nl/files/sigir2014-personalisationFei.pdf

Chuklin, A., & de Rijke, M. (2015). The Anatomy of Relevance: Topical, Snippet and Perceived Relevance in Search Result Evaluation. arXiv preprint arXiv:1501.06412. http://arxiv.org/pdf/1501.06412

Chuklin, A., Zhou, K., Schuth, A., Sietsma, F., & de Rijke, M. (2014, July). Evaluating intuitiveness of vertical-aware click models. In Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval (pp. 1075-1078). ACM. http://staff.science.uva.nl/~mdr/Publications/sigir2014-sp-clickmodels.pdf

Chuklin, A., Schuth, A., Zhou, K., & Rijke, M. D. (2015). A Comparative Analysis of Interleaving Methods for Aggregated Search. ACM Transactions on Information Systems (TOIS), 33(2), 5. http://ilps.science.uva.nl/sites/ilps.science.uva.nl/files/tois2015-interleaving-for-aggregated-search.pdf

Crasborn, O. & H. Sloetjes, Improving the Exploitation of Linguistic Annotations in ELAN. Proceedings of LREC 2014, Reykjavik, Iceland. [URL]

Crasborn, O., Hulsbosch, M., Lampen, L., & Sloetjes, H. (2014). New multilayer concordance functions in ELAN and TROVA. In Tilburg Gesture Research Meeting [TiGeR 2013]. http://pubman.mpdl.mpg.de/pubman/item/escidoc:1760119/component/escidoc:1760120/crasborn.pdf

van Dalen-Oskam, K., de Does, J., Marx, M., Sijaranamual, I., Depuydt, K., Verheij, B., & Geirnaert, V. (2014). Named entity recognition and resolution for literary studies. *Computational Linguistics in the Netherlands Journal*, *4*, 121-136.

Dijk, E., & Doorn, P. (2014). Providing access to research data, publications and current research information at Data Archiving and Networked Services-DANS. In *Conference on Grey Literature and Repositories* (p. 39).

Dorofeeva, A. (2014). Towards Digital Humanities Tool Criticism. https://openaccess.leidenuniv.nl/bitstream/handle/1887/29597/MA_thesis_Dorofeeva_TowardsDigitalHumanitiesToolCriticism.pdf?sequence=1

Durco and Windhouwer, The CMD Cloud. Proceedings of LREC 2014, Reykjavik, Iceland. [URL]

Gompel, M. van (2014). CLAM: Computational Linguistics Application Mediator – Documentation – LST Technical Report Series LST-14-02 Nijmegen, 2014

Gompel, M. van (2014). FoLiA: Format for Linguistic Annotation. http://antalvandenbosch.ruhosting.nl/folia.pdf

Gompel, M. van and M. Reynaert - CLAM: Quickly deploy NLP command-line tools on the web. Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics, 2014.

Gompel, M. van & M. Reynaert (2014), 'FoLiA: A practical XML Format for Linguistic Annotation - a descriptive and comparative study', in: Computational Linguistics in the Netherlands Journal 3.

Van Gorp, J. (2014). modes of self-representation: Visualized identities of former yugoslav migrant women in the netherlands. Crossings: Journal of Migration & Culture, 5(1), 153-171.

Graus, D., van Dijk, D., Tsagkias, M., Weerkamp, W., & de Rijke, M. (2014, July). Recipient recommendation in enterprises using communication graphs and email content. In Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval (pp. 1079-1082). ACM. http://graus.nu/paper/sigir2014-recipientrecommendation.pdf

Graus, D., Odijk, D., Tsagkias, M., Weerkamp, W., & de Rijke, M. (2014, July). Semanticizing search engine queries: the University of Amsterdam at the ERD 2014 challenge. In Proceedings of the first international workshop on Entity recognition & disambiguation (pp. 69-74). ACM. http://web-ngram.research.microsoft.com/ERD2014/Docs/submissions/erd14_submission_8.pdf

Graus, D., Tsagkias, M., Buitinck, L., & de Rijke, M. (2014). Generating pseudo-ground truth for predicting new concepts in social streams. In Advances in Information Retrieval (pp. 286-298). Springer International Publishing. https://staff.science.uva.nl/m.derijke/Publications/ecir2014-fp-ookb.pdf

Heuvel, H. van den, Sanders, E., Klatter, J., Hout, R. van, Fikkert, P., Baker, A., Jong, J. de, Wijnen, F. & Trilsbeek, P. (2014). Data curation for a VALID Archive of Dutch Language Impairment Data. In: *Dutch Journal of Applied Linguistics*, 3(2), pp. 127-135. DOI:10.1075/dujal.3.2.02heu

Hofmann, K., Whiteson, S., Schuth, A., & de Rijke, M. (2014). Learning to rank for information retrieval from user interactions by K. Hofmann, S. Whiteson, A. Schuth, and M. de Rijke with Martin Vesely as coordinator. ACM SIGWEB Newsletter, (Spring), 5. https://staff.science.uva.nl/s.a.whiteson/pubs/hofmannsigweb14.pdf

Hofmann, K., Schuth, A., Bellogin, A., & de Rijke, M. (2014). Effects of Position Bias on Click-Based Recommender Evaluation. In Advances in Information Retrieval (pp. 624-630). Springer International Publishing. http://www.anneschuth.nl/wp-content/uploads/2014/01/ecir2014-sp-interleaving.pdf

Huijnen, P., Laan, F., de Rijke, M., & Pieters, T. (2014). A Digital Humanities Approach to the History of Science. In Social Informatics (pp. 71-85). Springer Berlin Heidelberg. http://ilps.science.uva.nl/sites/ilps.science.uva.nl/files/histoinformatics2013-wahsp.pdf

Huijnen, P., Laan, F., de Rijke, M., & Pieters, T. (2014). A Digital Humanities Approach to the History of Science Eugenics Revisited in Hidden Debates by Means of Semantic Text Mining. SOCIAL INFORMATICS, SOCINFO 2013, 8359, 70-84. http://dspace.library.uu.nl/bitstream/handle/1874/303426/histoinformatics2013_wahsp.pdf?sequence=1

Klatter, J., Hout, R. van, Heuvel, H. van den, Fikkert, P., Baker, A., Jong, J. de, Wijnen, F., Sanders, E. & Trilsbeek, P. (2014). Vulnerability in Acquisition, Language Impairments in Dutch: Creating a VALID Data Archive. *Language Resources and Evaluation Conference Proceedings* (LREC) May 26-31, 2014, Reykjavik. pp. 357-364.

Kleppe, M., Hollink, L., Kemman, M., Juric, D., Beunders, H., Blom, J., Oomen, J., Houben, G.-J. (2014). PoliMedia - Analysing Media Coverage of Political Debates By Automatically Generated Links to Radio & Newspaper Items. In M. D'Aquin, S. Dietze, H. Drachsler, M. Guy, & E. Herder (Eds.), Proceedings of the LinkedUp Veni Competition on Linked and Open Data for Education. Geneva, Switzerland: CEUR-WS.

Lefortier, D., Serdyukov, P., Romanenko, F., & de Rijke, M. (2014). Blending Vertical and Web Results. In Advances in Information Retrieval (pp. 184-196). Springer International Publishing. https://staff.science.uva.nl/m.derijke/Publications/ecir2014-fp-verticals.pdf

Liang, S., Ren, Z., & de Rijke, M. (2014). The impact of semantic document expansion on cluster-based fusion for microblog search. In Advances in Information Retrieval (pp. 493-499). Springer International Publishing. https://staff.fnwi.uva.nl/s.liang/uploads/Main/HomePage/ecir2014.pdf

Liang, S., Ren, Z., & de Rijke, M. (2014, July). Fusion helps diversification. In Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval (pp. 303-312). ACM. https://staff.science.uva.nl/m.derijke/Publications/sigir2014-fp-fusion.pdf

Liang, S., Ren, Z., & de Rijke, M. (2014, August). Personalized search result diversification via structured learning. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 751-760). ACM.

Odijk, J. 'CLARIN-NL: Major Results', in Proceedings LREC 2014, Reykjavik, May 2014, pp 2187-2193 [pdf] [URL]

Odijk, J. 'The CLARIN infrastructure in the Netherlands: What is it and how can you use it?', unpublished paper, Utrecht University [pdf] [URL]

Odijk, J. 'The CLARIN infrastructure in the Netherlands: Design and Construction', unpublished paper, Utrecht University [pdf] [URL]

Odijk, J. 'Discovering Resources in CLARIN: Problems and Suggestions for Solutions', internal CLARIN Report, Utrecht University [pdf] [URL]

Oostdijk, N. & H. van den Heuvel, The Evolving Infrastructure for Language Resources and the Role for Data Scientists. Proceedings of LREC 2014, Reykjavik, Iceland. [URL]

Pander Maat, H., Kraf, R., Van den Bosch, A., Dekker, N., Van Gompel, M., Kleijn, S., Sanders, T., and Van der Sloot, K. (2014). T-Scan: A new tool for analysing Dutch Text. Computational Linguistics in the Netherland Journal, 4, pp. 53-74.]

Peetz, M. H., Meij, E., & de Rijke, M. (2014). Using temporal bursts for query modeling. Information retrieval, 17(1), 74-108. http://edgar.meij.pro/wp-content/papercite-data/pdf/irj-2013-peetz.pdf

Piersma, H., Tames, I., Buitinck, L., Doornik, J., & Marx, M. (2014). War in Parliament: What a Digital Approach Can Add to the Study of Parliamentary History. *Digital Humanities Quarterly*, *8*(1).

Ponstein, R. *Reconciling Dutch ships and sailors*. M.Sc. thesis VU University Amsterdam (forthcoming), 2014

Postma, M., & Vossen, P. (2014). Open Source Dutch WordNet. http://kyoto.let.vu.nl/~postma/odwn/Documentation/odwn_documentation.pdf

Rehm, G., Uszkoreit, H., Odijk, J. E. J. M., & Zabarskaite, J. (2014). The Strategic Impact of META-NET on the Regional, National and International Level. Proceedings of LREC 2014 http://dspace.library.uu.nl/bitstream/handle/1874/295272/405_Paper.pdf?sequence=1

Ren, Z., Peetz, M. H., Liang, S., van Dolen, W., & de Rijke, M. (2014, July). Hierarchical multi-label classification of social text streams. In Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval (pp. 213-222). ACM. http://staff.science.uva.nl/~sliang/uploads/Main/Publications/sigir2014-classification.pdf

Reynaert, M. Synergy of Nederlab and @PhilosTEI: Diachronic and Multilingual Text-induced Corpus Clean-up. Proceedings of LREC 2014, Reykjavik, Iceland. [URL]

Reynaert, M. (2014). TICCLops: Text-Induced Corpus Clean-up as online processing system. COLING 2014, 52. http://www.aclweb.org/anthology/C/C14/C14-2.pdf#page=64

Reynaert, M. (2014, May). On OCR ground truths and OCR post-correction gold standards, tools and formats. In Proceedings of the First International Conference on Digital Access to Textual Cultural Heritage (pp. 159-166). ACM http://www.nederlab.nl/wp/wp-content/uploads/2014/03/Datech2014-poster.pdf

Reynaert, M., van de Camp, M., Taalmonsters, D., & van Zaanen, M. (2014). OpenSoNaR: user-driven development of the SoNaR corpus interfaces. COLING 2014, 124. http://www.aclweb.org/anthology/C/C14/C14-2.pdf#page=136

Roorda, D. (2015). The Hebrew Bible as Data: Laboratory-Sharing-Experiences. arXiv preprint arXiv:1501.01866. http://arxiv.org/pdf/1501.01866

Roorda, D., Kalkman, G., Naaijer, M., & van Cranenburgh, A. (2014). LAF-Fabric: a data analysis tool for Linguistic Annotation Framework with an application to the Hebrew Bible. arXiv preprint arXiv:1410.0286. http://arxiv.org/pdf/1410.0286

Sanders, E. et al. The Dutch LESLLA Corpus, Proceedings of LREC 2014, Reykjavik, Iceland. [URL]

Schuth, A., Sietsma, F., Whiteson, S., & de Rijke, M. (2014). Optimizing Base Rankers Using Clicks. In Advances in Information Retrieval (pp. 75-87). Springer International Publishing. http://staff.science.uva.nl/~mdr/Publications/ecir2014-fp-bm25.pdf

Severyn, A., Moschitti, A., Tsagkias, M., Berendsen, R., & de Rijke, M. (2014, July). A syntax-aware re-ranker for microblog retrieval. In Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval (pp. 1067-1070). ACM. http://ilps.science.uva.nl/sites/ilps.science.uva.nl/files/sigir2014-sp-microblog.pdf

Sijs, V. D. N. (2014). De voortzetting van de historische taalkunde met andere middelen. http://depot.knaw.nl/14952/1/2014_Sijs_Voortzetting_historische_taalkunde_oratie.pdf

Sijs, Nicoline van der (2014), 'Less is more in digitaliseringsland', in: Nieuw Letterkundig Magazijn 32, 1, 26.

Sijs, Nicoline van der (2014), 'Systematisch onderzoek naar Nederlandse contactvariëteiten', in: Taal & Tongval 66.

Sijs, Nicoline van der (2014), 'Besteed de geesteswetenschappen niet uit!', in: Nieuw Letterkundig Magazijn 32, 2.

van Sluijs, R. (2014). What's Past Is Past: Variation in the Expression of Past Time Reference in Negerhollands Narratives. Journal of Germanic Linguistics, 26(03), 272-321.

Stronks, E. (2015). 'Dees kennisse zuldy te kope vinnen'. Liedcultuur en de waarde van 'know how'in de vroegmoderne Republiek. De Zeventiende Eeuw. Cultuur in de Nederlanden in interdisciplinair perspectief, 30(2), 147-167. http://www.de-zeventiende-eeuw.nl/index.php/dze/article/view/9932/10409

Voskarides, N., Odijk, D., Tsagkias, M., Weerkamp, W., & de Rijke, M. (2014). Query-dependent contextualization of streaming data. In Advances in Information Retrieval (pp. 706-712). Springer International Publishing. http://wouter.weerkamp.com/downloads/ecir2014-query-dependent-contextualization.pdf

De Vriend, F., Broeder, D., Depoorter, G., van Eerten, L., & Van Uytvanck, D. (2013). Creating & testing CLARIN metadata components. *Language resources and evaluation*, *47*(4), 1315-1326.

Windhouwer, Menzo, and Matej Durco. 2014. "Lightweight Semantics in the CLARIN Infrastructure." (2014). EUON Workshop, Amsterdam, the Netherlands.

Windhouwer, M., & Schuurman, I. (2014, May). Linguistic resources and cats: how to use ISOcat, RELcat and SCHEMAcat. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)* (pp. 3806-3810).

Wright, S. E., Windhouwer, M., Schuurman, I., & Broeder, D. (2014). Segueing from a Data Category Registry to a Data Concept Registry. In Terminology and Knowledge Engineering 2014 (pp. 11-p). https://hal.archives-ouvertes.fr/hal-01005840/document

Zeldenrust, D. A. (2014). Access to data, The Soundbites collection of the Meertens Institute and a flexible approach to the curation and dissemination of humanities digital resources. In Digital Humanities Benelux Conference.

Zeldenrust, D. A. (2014). The Exploration and Visualisation of Big Data in the Humanities, Comparing Data Representation of a Large-scale E-infrastructure (CLARIN) and a Dedicated Virtual Research Environment (Soundbites). In European Social Science History Conference 2014 (ESSHC).

Zoghi, M., Whiteson, S. A., de Rijke, M., & Munos, R. (2014, February). Relative confidence sampling for efficient on-line ranker evaluation. In Proceedings of the 7th ACM international conference on Web search and data mining (pp. 73-82). ACM. https://staff.fnwi.uva.nl/s.a.whiteson/pubs/zoghiwsdm14.pdf

2015

Bergmann, L. (2015). Taalvaardigheden in Narratieven bij Kinderen met SLI. MA Thesis, RU Nijmegen.

Bouma, G., van Koppen, J. M., Landsbergen, F., Odijk, J. E. J. M., van der Wouden, T. & van de Camp, M. (2015) 'Enriching a Descriptive Grammar with Treebank Queries', In : *Proceedings of the Fourteenth International Workshop on Treebanks and Linguistic Theories (TLT14)*, 14, p. 13-25. [url] [pdf]

Bron, M., Van Gorp, J., & Rijke, M. (2015). Media studies research in the data-driven age: How research questions evolve. *Journal of the Association for Information Science and Technology*.

Chuklin, Aleksandr, et al. "A comparative analysis of interleaving methods for aggregated search." *ACM Transactions on Information Systems (TOIS)* 33.2 (2015): 5.

Van Gorp, Jasmijn, et al. "Digital Media Archaeology: Uncovering the digital tool AVResearcherXL." *VIEW Journal of European Television History and Culture* 4.7 (2015): 38-53.

Heuvel, H. van de , N. Oostdijk, E. Sanders & V. De Lint. 2015. Data curations by the Dutch Data Curation Service: Overview and future perspective. *Selected Papers from the CLARIN 2014 Conference*,54-62.

Kemman, M. 2015. User Required? On the Value of User Research in the Digital Humanities. *Selected Papers from the CLARIN 2014 Conference*, 63-74.

Klamer, M. (2015). Language as Time Machine. Oratie University of Leiden.

Odijk, J. (2015) 'Linguistic Research with PaQU' *Computational Linguistics in The Netherlands journal* 5, p. 3-14 [pdf] [url]

Odijk, J. (2015), 'Metadata Curation Strategy', manuscript Utrecht University. [pdf] with 2 appendixes: resource type [xlsx] and modality [xlsx]

Odijk, J. (2015) 'Linguistic Research with PaQU' *Computational Linguistics in The Netherlands journal* 5, p. 3-14 [pdf] [url]

Odijk, J. (ed.) 2015. Selected Papers from the CLARIN 2014 Conference, October 24-25, 2014, Soesterberg, The Netherlands. Linköping" Linköping Electronic Conference Proceedings 116. http://www.ep.liu.se/ecp/article.asp?issue=116

Post, P. (2015). Pelgrimsverhalen–pelgrims verhalen: Mapping the field. *Jaarboek voor liturgie-onderzoek*, 31.

Van der Sijs, Nicoline. "TS Tools: Nederlab voor tijdschriftonderzoek." *Tijdschrift voor Tijdschriftstudies* 37 (2015): 53-62.

Stronks, Els. "'Dees kennisse zuldy te kope vinnen'. Liedcultuur en de waarde van 'know how'in de vroegmoderne Republiek." *De Zeventiende Eeuw. Cultuur in de Nederlanden in interdisciplinair perspectief* 30.2 (2015).

2016

Barbiers, Sjef, Marjo van Koppen, Hans Bennis, Norbert Corver (2016). MIcrocomparative MOrphosyntactic REsearch (MIMORE): Mapping partial grammars of Flemish, Brabantish and Dutch. doi:10.1016/j.lingua.2015.10.018. to appear in *Lingua*.

Brugman, Hennie, Martin Reynaert, Nicoline van der Sijs, René van Stipriaan, Erik Tjong Kim Sang and Antal van den Bosch (2016), 'Nederlab: Towards a Single Portal and Research Environment for Diachronic Dutch Text Corpora', accepted for publication in the LREC 2016 proceedings.

Cornips, Leonie, Jos Swanenberg, Wilbert Heeringa, Folkert de Vriend (2016). The relationship between first language acquisition and dialect variation: Linking resources from distinct disciplines in a CLARIN-NL project. doi:10.1016/j.lingua.2015.11.007. to appear in *Lingua*.

Heuvel, Henk van den and Nelleke Oostdijk (2016), 'Falling silent, lost for words ... Tracing personal involvement in interviews with Dutch war veterans', accepted for publication in the LREC 2016 proceedings.

Leinonen, Therese, Çağrı Çöltekin, John Nerbonne (2016). Using Gabmap. doi:10.1016/j.lingua.2015.02.004. to appear in *Lingua*.

Odijk, J. *(2016)* 'A Use Case for Linguistic Research on Dutch with CLARIN', in K. De Smedt (ed.), *Selected Papers from the CLARIN Annual Conference 2015,* 45-61. [Abstract and Fulltext]

Odijk, J. (2016), 'Linguistic Research using CLARIN – Introduction'. doi:10.1016/j.lingua.2016.04.003. To appear in *Lingua.*

Odijk, J. (2016), 'CLARIAH in the Netherlands', accepted for publication in the LREC 2016 proceedings.

Reynaert, Martin (2016), 'OCR Post-Correction Evaluation of Early Dutch Books Online – Revisited', accepted for publication in the LREC 2016 proceedings.

Scholman, Merel CJ, Jacqueline Evers-Vermeul, and Ted JM Sanders. "Categories of coherence relations in discourse annotation: Towards a reliable categorization of coherence relations." *Dialogue & Discourse* 7.2 (2016): 1-28.

Schuurman, Ineke, Menzo Windhouwer, Oddrun Ohren, Daniel Zeman (2016), 'CLARIN Concept Registry: The New Semantic Registry', in K. De Smedt (ed.), *Selected Papers from the CLARIN Annual Conference 2015,* 62-70. [Abstract and Fulltext]

Sijs, Nicoline van der, Eric Sanders, Henk van den Heuvel & Aukje Borkent (2016), 'Curation of Dutch Regional Dictionaries', accepted for publication in the LREC 2016 proceedings.

Sluijs, Robbert van, Margot van den Berg, Pieter Muysken (2016). Exploring genealogical blends: The Surinamese Creole cluster and the Virgin Island Dutch Creole cluster. doi:10.1016/j.lingua.2015.12.004 . to appear in *Lingua*.

Windhouwer, Menzo , Marc Kemps-Snijders, Paul Trilsbeek, André Moreira, Bas Van der Veen and Guilherme Silva (2016). 'FLAT: constructing a CLARIN compatible home for language resources', accepted for publication in the LREC 2016 proceedings.

Wieling, M. , E. Sassolini, N. Cucurullo, S. Montemagni . 2016. ALT Explored: Integrating an online dialectometric tool and an online dialect atlas. Accepted for publication in the LREC 2016 proceedings.

Wouden, Ton van der, Gosse Bouma, Matje van de Camp, Marjo van Koppen, Frank Landsbergen, Jan Odijk (2016). 'Enriching a Grammatical Database with Intelligent Links to Linguistic Resources', in K. Desmedt (ed.), *Selected Papers from the CLARIN Annual Conference 2015,* 108-117. [Abstract and Fulltext]

Wouden, Ton van der et al. Het Taalportaal: Een nieuwe wetenschappelijke grammatica voor het Nederlands en het Fries. To appear in *Nederlandse Taalkunde*, 2016

Presentations

2009

Barbiers, S., Kunst, J.P. (2009), "DynaSAND and DynaMAND", presentation held at the CLARIN-NL Kick-off Meeting, Utrecht, May 27, 2009

Bennis, H. (2009), "Geography, databases, theoretical linguistics and CLARIN-NL", presentation held at ICLAVE 5, Copenhagen, June, 2009

Bennis, H. (2009), "Digitale geesteswetenschappen: hoe en waarom?", presentation held at KNAW Symposium Computational Humanities: Bridging the gap between the Humanities and the Computational Sciences, Amsterdam, June 4, 2009

Bennis, H. (2009), "CLARIN-NL and dialect data", presentation held at the Grand Meeting of ScanDiaSyn, Alvdalen, Sweden, August, 2009

Broeder, D. (2009), "Component Metadata, DatCats + VLO", presentation held at the CLARIN-NL Info Session, MPI, Nijmegen, July 1, 2009

Hessen, A. van (2009), "Verteld Verleden", presentation held at the Nationale Seminar AV-Archieven, Beeld & Geluid, Hilversum, November 26, 2009

Hessen, A. van (2009), "Educational and Training Activities in CLARIN-NL", presentation held at the CLARIN Consortium Meeting, Barcelona, Spain, May 11,2009.

Hessen, A. van (2009) "CLARIN: wat is dat?", presentation held at the Dag van de Fonetiek, Utrecht, December 18, 2009

Jung, D., Wittenburg, P. (2009), "Taaldocumentatie en Infrastructuur", presentation held at the CLARIN-NL Kick-off Meeting, Utrecht, May 27, 2009

Kemps-Snijders, M. (2009), "Web-services and WorkFlow", presentation held at the CLARIN-NL Info Session, MPI, Nijmegen, July 1, 2009

Kemps-Snijders, M. (2009), "e-Infrastructures, the CLARIN perspective", presentation held at the SURF Foundation Autumn School, Leiden, November 3,2009

Kemps-Snijders, M. (2009), "Format Interoperability, Standardization", presentation held at the CLARIN-NL Info Session, MPI, Nijmegen, July 1, 2009

Krauwer, S. (2009), "CLARIN-EU: Waar staan we?", presentation held at the CLARIN-NL Kick-off Meeting, Utrecht, May 27, 2009

Odijk, J. (2009), "CLARIN-NL", presentation held at the CLARIN-NL Kick-off Meeting, Utrecht, May 27, 2009

Odijk, J. (2009), "CLARIN-NL First Call: Oproep voor Demonstrator & Resource Curation projecten", presentation held at the CLARIN-NL Kick-off Meeting,Utrecht, May 27, 2009

Odijk, J. (2009), "Data Categories and ISOCAT: some remarks from a simple linguist", presentation held at FLaReNet/CLARIN Standards Workshop, Helsinki, September 27, 2009

Odijk, J. (2009), "Data Categories and their Semantics", presentation held at the SILT/FLaReNet Workshop, Brandeis University, Waltham, USA, November1, 2009

Odijk, J. et al. (2009), "Data Categories and their Semantics: Summary", presentation held at the SILT/FLaReNet Workshop, Brandeis University, Waltham, USA, November 2, 2009

Uytvanck, D. van (2009), "AAI Federation Planning", presentation held at the CLARINNL Kick-off Meeting, Utrecht, May 27, 2009

Veenendaal, R. van (2009), "Klaar? In!", presentation held at the CLARIN-NL Kick-off Meeting, Utrecht, May 27, 2009

Windhouwer, M. (2009), "Semantic Interoperability, ISOcat", presentation held at the CLARIN-NL Info Session, MPI, Nijmegen, July 1, 2009

Wittenburg, P. (2009), "Centres in CLARIN", presentation held at the CLARIN-NL Info Session, MPI, Nijmegen, July 1, 2009

Wittenburg, P. (2009), "Persistent Identifier", presentation held at the CLARIN-NL Info Session, MPI, Nijmegen, July 1, 2009

2010

Bennis, H.J. (2010), "Linguistics, databases and Open Access", presentation held at the Grand Meeting ScanDiaSyn, Tromsø, Norway, June 8, 2010.

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

Bennis, H. (2010), "Syntactic Variation", Presentation held at SDH 2010, Vienna, Austria. 19 October. http://ztwweb.trans.univie.ac.at/sdh2010/

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.

Broeder, D. (2010), "CLARIN Infrastructure Vision (and some real needs)", presentation held at SARA Amsterdam, March 8, 2010.

Broeder, D. (2010), "CMDI Software Components", presentations/demo CMDI for CLARIN A librarians, Vienna, March 29-30, 2010.

Broeder, D. (2010), "CLARIN AAI, Web Services Security Requirements", presentation held at the Web-services security workshop, co-organized with BiG-Grid at NIKHEF Amsterdam, April 27, 2010.

Broeder, D. (2010), "A Data Category Registry- and Component-based Metadata Framework", presentation held at LREC 2010, May 19, 2010, Valetta, Malta.

Broeder, D. (2010), "CMDI - CLARIN Component Metadata Infrastructure", presentation held at the CLARIN CMDI workshop, Max Planck Instituut, Nijmegen, May 27, 2010.

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

Broeder, D. (2010), "Advanced Metadata Usage", presentation held at the APA workshop "Research metadata in context", Max Planck Instituut, Nijmegen, September 7-9, 2010.

Broeder, D. (2010), "Flexible Syntax and Concept Registries as a basis for Metadata", presentation held at the APA workshop "Research metadata in context", Max Planck Instituut, Nijmegen, September 7-9, 2010.

Broeder, D. (2010), "Metadata Component Framework - Possible Standardization Work within TC37/SC4", presentation held at the ISO TC37/SC4 meeting, Berlin, October 12-13, 2010.

Crasborn, O., Hulsbosch, M., Sloetjes, H., Schermer, T. and Harmsen, H. (2010), “Linking ELAN and LEXUS”, poster presentation at the 4th workshop of the Sign Linguistics Corpora Network, "Exploitation of Sign Language Corpora", Berlin, Dec. 3-4, 2010.

Grégoire, N. (2010), "DuELME LMF project", presentation held at the Workshop "D-Spin meets RELISH: standardizing lexicon structures", University of Frankfurt, April 26, 2010.

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", Max Planck Instituut, Nijmegen, August 4, 2010.

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

Hessen, A. van & J. Odijk (2010). CLARIN-NL. Poster at Conferentie Grootschalige Infrastructuren 'Closing the Deal', LUMC, Leiden, December 2, 2010.

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

Heuvel, H. van den (2010), “'INTER-VIEWs'. Curation of Interview Data". Presentation held at the first CLARIN-NL meeting, Feb 19, 2010.

Heuvel, H. van den (2010), "The VeteranTapes: Research Corpus, Fragment Processing Tool, and Enhanced Publications for the e-Humanities", poster presentation at LREC2010, Malta, May 20, 2010.

Heuvel, H. van den (2010), "INTER-VIEWs, Curation of interview data", presentation held at the ISOcat Workshop, Nijmegen, September 21, 2010.

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.

Kemps-Snijders, M., Wright, S.E., Windhouwer, M. (2010), "ISOcat Introduction", presentation held at the ISOcat Tutorial, Utrecht, March 25, 2010.

Kemps-Snijders, M., Windhouwer, M. (2010), "ISOcat Annotation Tutorial", presentation held at the ISOcat Tutorial, Utrecht, March 25, 2010.

Kemps-Snijders, M., Windhouwer, M. (2010), "ISOcat Datamodel Tutorial", presentation held at the ISOcat Tutorial, Utrecht, March 25, 2010.

Kemps-Snijders, M., Windhouwer, M. (2010), "ISOcat Hands-on Tutorial", presentation held at the ISOcat Tutorial, Utrecht, March 25, 2010.

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

Leinonen, T. (2010), "RuG/L04 - en webbapplikation", presentation held at the Nordic Congress of Dialectologists, Uppsala, August 20, 2010.

Leinonen, T. (2010), "Gabmap − doing dialect analysis on the web", presentation held at the Tagung des Forums Sprachvariation, Erlangen, October 15, 2010.

Leinonen, T. (2010), "Dialektometrins bidrag till dialektologin", presentation held at the Föreningen för nordisk filologis höstsymposium: "Nya perspektiv på dialekter", Tvärminne (Finland), November 26-27, 2010.

Leinonen, T. (2010), "Gabmap − doing dialect analysis on the web", presentation for researchers at the Dialect dictionary of Finnish dialects and Dictionary of Swedish dialects in Finland, The Research Institute for the Languages of Finland, Helsinki, November 30, 2010. [no slides; but see http://www.gabmap.nl/~app/doc/tutorial/]

Leinonen, T. (2010), "Gabmap − doing dialect analysis on the web", presentation at the computer department at The Research Institute for the Languages of Finland, Helsinki, December 1, 2010. [no slides; but see http://www.gabmap.nl/~app/doc/tutorial/]

Leinonen, T. (2010), "Gabmap − doing dialect analysis on the web", presentation held at the Gabmap tutorail at the University of Potsdam, Potsdam, December 7, 2010.

Odijk, J. (2010), ""Relations between Data Categories, presentation held at the CLARIN Relation Registry Workshop, MPI, Nijmegen, January 8, 2010.

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

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

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.

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.

Odijk, J. (2010), "CLARIN-NL: What is going on?", presentation held at the CLARIN-NL Second Open Call Info Session, Amsterdam, August 26, 2010.

Odijk, J. (2010), "CLARIN-NL Second Open Call", presentation held at the CLARIN-NL Second Open Call Info Session, Amsterdam, August 26, 2010.

Odijk, J. (2010), "ISOCAT problems encountered at in DUELME-LMF", presentation held at the ISOcat Workshop, Nijmegen, September 21, 2010

Odijk, J. (2010), "Proposed solutions for problems encountered in DUELME-LMF", presentation held at the ISOcat Workshop, Nijmegen, September 21, 2010

Odijk, J. (2010). Contribution to the Panel Session on Open and Persistent Access, SDH 2010, Vienna, October 20, 2010.

Poorter, G. de (2010), “Metadataproject”, presentation held at the 1st CLARIN-NL Event, Utrecht, February 19th, 2010.

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.

Ravenek, W. (2010), "Language technology in the CKCC project", presentation held at Supporting Digital Humanities 2010, Wien, November 19, 2010.

Roorda, D., Bos, E-J. and Van den Heuvel, C.M.J.M. (2010), “Letters, Ideas and Information Technology: Using digital corpora of letters to disclose the circulation of knowledge in the 17th century”, paper/presentation held at Digital Humanities 2010, London, July 9, 2010.

Scagliola, S. (2010), "Interview Project Dutch Veterans", presentation held at the Second Military Oral History Conference "Between Memory and History", Victoria BC, Canada, May 5-7, 2010.

Scagliola, S. (2010), "Interview Project Dutch Veterans", presentation held at XVI International Oral History Conference, Prague, Czech Republic, July 7, 2010.

Scagliola, S. (2010), "The Veteran Tapes : An Inquiry into the benefits, disadvantages and preconditions for the reuse of interviews as sources of scientific knowledge", presentation held at CLARIN/DARIAH Supporting the Digital Humanities, Vienna, October 19-20, 2010.

Scagliola, S. (2010), "Interviewproject Nederlandse Veteranen", presentation held at Symposium DANS/Kwalon/VI, "Wat Veteranen Vertellen", Amsterdam, October 27, 2010.

Scagliola, S. (2010), "Speech retrieval for interview data; Technology push versus academic demand", presentation held at "ACM Multimedia 2010 - Workshop Searching Spontaneous Conversational Speech", Florence, October 29, 2010.

Scagliola, S. (2010), "Hergebruik van interviews: hoe anticipeert de verzamelaar op de behoeften van verschillende soorten onderzoekers?", presentation held for the working group Oral History Huizinga Instituut, Amsterdam, November 26, 2010.

Scagliola, S. (2010), "De biografie van het Interviewproject Nederlandse Veteranen (IPNV); Empowerment, community building of data-verzameling voor onderzoekers?", presentation held at Jubileum Aletta, centrum voor vrouwengeschiedenis, Amsterdam, December 3, 2010.

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.

Sijens, H. and Depuydt, K. (2010), "Wurdboek fan de Fryske taal/Dictionary of the Frisian Language online: new possibilities, new opportunities", presentation held at the XIV Euralex International Congress, Leeuwarden, July 6, 2010.

Sijens, H. and Depuydt, K. (2010), "CLARIN-NL Project WFT-GTB", presentation held at the CLARIN meeting, Utrecht, October 28, 2010.

Van den Heuvel, C.M.J.M. (2010), “Visualizing scholarly debates in 17th Century Correspondences”, invited lecture held at the Stanford Humanities Center, Stanford – Palo Alto, April 18, 2010.

Van den Heuvel, C.M.J.M. (2010), “Notes and Nodes; Annotations, Visualizations and Interfaces to e-research from a historical perspective”, invited lecture held at the Graduate School for Education and Information, Information School Colloquium, UCLA, Los Angeles, October 7, 2010.

Van den Heuvel, C.M.J.M. (2010), “Creating Histories using Information and Experiences; Annotations, models and visualizations in historical research and cultural heritage”, invited lecture held at the School of Social Sciences, Humanities and Arts. Center for Research in the Humanities and Arts, University of California, Merced, October 13, 2010.

Van Uytvanck, D. (2010), "Metadata and DCR", presentation held at the ISOcat Tutorial, Utrecht, March 25, 2010.

Vriend, F. de (2010), "Creating & Testing CLARIN Metadata Components", presentation held at LREC 2010, Valetta, Malta, May 18, 2010.

Vriend, F. de (2010), "MIMORE", presentation held at the CLARIN-NL Second Open Call Info Session, Amsterdam, August 26, 2010.

Windhouwer, M. (2010), "ISOcat status", presentation held at the ISOcat Workshop, Nijmegen, September 21, 2010.

Windhouwer, M. (2010), "TDS Curator", presentation held at the ISOcat Workshop, Nijmegen, September 21, 2010.

Windhouwer, M. (2010), "Possible solutions ISOcat", presentation held at the ISOcat Workshop, Nijmegen, September 21, 2010.

2011

Barbiers, S. (2011), "MIMORE", presentation held at the Call 3 information session, Utrecht, August 25, 2011.

Bennis, H. (2011), "Populating the infrastructure: the case of the Netherlands", presentation held at the CLARIN Coordinators Meeting, Budapest, June 29 and 30, 2011.

Berg, M. van den and Sluijs, R. van (2011), "NEHOL - NEgerHOLlands Database", presentation held at the CLARIN-NL Kick-off meeting Call 2, Utrecht, February 9, 2011.

Broeder, D (2011), "CMDI - CLARIN Component Metadata Infrastructure", presentation held at the CMDI Tutorial, Nijmegen, January 17, 2011.

Broeder, D. (2011), "Een kijkje in de CLARIN keuken", presentation held at SURFnet event about data management, Utrecht, April 7, 2011.

Broeder, D. (2011), "Introduction, Federated Search as a EU CLARIN demonstrator", presentation held at the EU workshop Federated Search, Nijmegen, May 9, 2011.

Broeder, D. (2011), "NP 24622 CMDI-1Metadata Component Framework - New Standardization Work within TC37/SC4", presentation held at the ISO TC37/SC4 meeting, Seoul, June 13, 2011.

Broeder, D. (2011), "Populating the infrsatructure using standards", presentation held at the CLARIN Coordinators Meeting, Budapest, June 29 and 30, 2011.

Bruin, M. de (2011), "C-DSD - Curating the Dutch Song Database", presentation held at the CLARIN-NL Kick-off meeting Call 2, Utrecht, February 9, 2011.

Caspel, P. van (2011), "CLARIN-NL HELPDESK", presentation held at the CLARIN-NL Kick-off meeting Call 2, Utrecht, February 9, 2011.

Cornips, L., Kemps-Snijders, M., Snijders, M., Swanenberg, J. & Vriend, F. de (2011), "Bridging the gap between first language acquisition and historical linguistics with the help of digital humanities", presentation held at Supporting Digital Humanities, University of Copenhagen, November 17, 2011.

Crasborn, O. (2011), "Support for sign language resources in the Netherlands and beyond in CLARINNL", presentation of SignLinC and IPROSLA (Call 2) held at the CLARIN-D kick-off meeting, Leipzig, July 22, 2011.

Dalen-Oskam, K. van and Besamusca, B. (2011), "Arthurian Fiction in Medieval Europe: Narratives and Manuscripts", presentation held at the CLARIN-NL Kick-off meeting Call 2, Utrecht, February 9, 2011.

Dalen-Oskam, K. van (2011), "ArthurianFiction", presentation held at the Call 3 information session, Utrecht, August 25, 2011.

Duin, P. (2011), CMDI Component Registry", presentation held at the CMDI Tutorial, Nijmegen, January 17, 2011.

Duin, P. (2011), CMDI Virtual Language Observatory Faceted Browsing", presentation held at the CMDI Tutorial, Nijmegen, January 17, 2011.

Ďurčo, M. (2011), CMDRSB", presentation held at the CMDI Tutorial, Nijmegen, January 17, 2011.

Fikkert, P. (2011), "Integrating and publishing resources on sign language aquisition (IPROSLA), presentation held at the CLARIN-NL Kick-off meeting Call 2, Utrecht, February 9, 2011.

Gerritsen, G.H. (2011), "Geleerdenbrievencorpus en analysetools", presentation held at the workshop 'Diachroon Corpus', Den Haag, March 30, 2011.

Halteren, H. van (2011), "Processing 14th century Dutch text through Clarin, step 1: Adelheid", presentation held at the CLARIN-NL Kick-off meeting Call 2, Utrecht, February 9, 2011.

Halteren, H. van (2011), "Adelheid: Tagging and Lemmatizing Historical Dutch Texts through the Clarin Infrastructure", presentation held at Galatea II, Antwerpen, February 18, 2011.

Halteren, H. van (2011), "Dealing with Medieval Text", presentation held for the course "Language Engineering Applications" at KU Leuven, Leuven, February 21, 2011.

Kemenade, A. van Postma, G. (2011), "International Perser and Lemmatizer of Dutch in Retrospect (INPOLDER)", presentation held at the CLARIN-NL Kick-off meeting Call 2, Utrecht, February 9, 2011.

Kemps--Snijders, M. (2011), "The Dutch Search and Development project and the Meertens search engine", presentation held at the EU workshop Federated Search, Nijmegen, May 9, 2011.

Kemps-Snijders, M. (2011), "Search and Develop", presentation held at the Call 3 information session, Utrecht, August 25, 2011.

Kleiweg, P. and Leinonen, T. (2011), "Gabmap - Doing dialect analysis on the web", poster at ICLAVE 6, Freiburg, June 30, 2011.

Leinonen, T. (2011), "Dialektgeografiska kartor i Gabmap", tutorial at the workshop vid Svenska litteratursällskapet i Finland r.f., Helsinki, November 23 and 25, 2011.

Nerbonne, J. (2011), 'Gabmap: A Web Application for Analyzing Linguistic Variation', Tutorial given at Digital Humanities 2011, Stanford, California, USA, May 19, 2011. https://dh2011.stanford.edu/?page_id=521

Nerbonne, J., Wieling, M. and Heeringa, W. (2011), "Gabmap - A web application for measuring and visualizing distances between language varieties", workshop at the international conference Methods in Dialectology XIV, London, Ontario, August 2-6, 2011.

Nerbonne, J. (2011), "Gabmap, een webapplicatie voor dialectologie", presentation held at the Call 3 information session, Utrecht, August 25, 2011.

Nerbonne, J. (2011), "Gabmap", tutorial in connection with the "Workshop on comparing approaches to measuring linguistic differences", Gothenburg, Sweden, October 24-26, 2011.

Odijk, J. (2011), "Introduction to Language and Speech Technology", presentations held at the LOT Winterschool, Amsterdam, January, 2011.

Odijk, J. (2011), "Call 2 Kick-off - Introduction", presentation held at the CLARIN-NL Kick-off meeting Call 2, Utrecht, February 9, 2011.

Odijk, J. (2011), "Linguistic Research in the CLARIN Infrastructure", presentation for the KNAW eHumanities Workshop, NIAS, Wassenaar, Mar 29, 2011. Abstract contained in eHumanities Brainstorm Booklet.

Odijk, J. (2011), "CLARIN-NL Call 3", presentation held at the Call 3 information session, Utrecht, August 25, 2011.

Odijk, J. (2011), "Requirements and desiderata", presentation held at the Call 3 information session, Utrecht, August 25, 2011.

Odijk, J. and A. van Hessen (2011), "Sharing Resources in CLARIN-NL", Presentation held at the Language Resources, Technology and Services in the Sharing Paradigm workshop at IJCNLP 2011 (12 November 2011), Chiang Mai Thailand.

Piersma, H. (2011), "War in parliament", presentation held at the CLARIN-NL Kick-off meeting Call 2, Utrecht, February 9, 2011.

Pieters, P. (2011), "WAHSP", presentation held at the Call 3 information session, Utrecht, August 25, 2011.

Ravenek, W. and Van den Heuvel, C.M.J.M. (2011), “Circulation of knowledge: Integrated access to texts and metadata”, presentation held at the International Workshop Representing the Republic of Letters, Huygens Institute for the History of the Netherlands (in cooperation with Descartes Centre UU), Den Haag, June 30 – July 1, 2011.

Ravenek, W. (2011), “Supporting the Exploration of a Corpus of 17th-Century Scholarly Correspondences by Topic Modeling”, presentation held at Supporting Digital Humanities 2011, Copenhagen, November 18, 2011.

Schuurman, I. and Windhouwer, M. (2011), "ISOcat introduction", presentation held at the ISOcat Workshop, Utrecht, May 10, 2011.

Schuurman, I. and Windhouwer, M. (2011), "Content of the Data Category Registry", presentation held at the ISOcat Workshop, Utrecht, May 10, 2011.

Schuurman, I. and Windhouwer, M. (2011), "Data Category specifications", presentation held at the ISOcat Workshop, Utrecht, May 10, 2011.

Schuurman, I. and Windhouwer, M. (2011), "RELcat and friends", presentation held at the ISOcat Workshop, Utrecht, May 10, 2011.

Schuurman, I. and Windhouwer, M. (2011), "ISOcat: known issues", presentation held at the ISOcat Workshop, Utrecht, May 10, 2011

Schuurman, I. (2011), "CLARIN-NL: Dealing with ISOcat", presentation held at the CLARIN Coordinators Meeting, Budapest, June 29 and 30, 2011.

Schuurman, I. and Windhouwer, M. (2011), "CLARIN-NL ISOcat workshop 2011, part 2", presentation held at the ISOcat workshop 2011, part 2, Utrecht, October 24, 2011.

Schuurman, I. and Windhouwer, M. (2011), "Bulk loading ISOcat data categories with the Data Category Interchange Format", presentation held at the ISOcat workshop 2011, part 2, Utrecht, October 24, 2011.

Snelders, S. and T. Pieters (2011), "WAHSP: Towards a flexible and stable CLARIN-supported web-application for historical sentiment mining in public media", presentation held at the CLARIN-NL Kick-off meeting Call 2, Utrecht, February 9, 2011.

Snelders, S. and T. Pieters (2011), "WAHSP: Towards a flexible and stable CLARIN-supported open-source web-application for historical sentiment mining in public media", presentation held at the Call 3 information session, Utrecht, August 25, 2011.

Snelders, S., Odijk, D. and Pieters, T. (2011), "Historical Sentiment Mining in Public Media: Problems and Challenges of E-Historical Methodology", presentation held at the Workshop E-History NIOD, Amsterdam, 23 juni, 2011.

Stehouwer, H. (2011), "Implementing SRU/CQL on Trova", presentation held at the EU workshop Federated Search, Nijmegen, May 9, 2011.

Strik, H. (2011), "TQE: Transcription Quality Evaluation", presentation held at the CLARIN-NL Kick-off meeting Call 2, Utrecht, February 9, 2011.

Trippel, T. (2011), "CMDI use in the NaLiDa project", presentation held at the CMDI Tutorial, Nijmegen, January 17, 2011.

Van den Heuvel, C.M.J.M. (2011), “Themes and Debates Emerging from 17th Century Correspondences of Dutch Scholars" Circulation of Knowledge: Visualizations and Interfaces for Research,” presentation held at the Mapping the Republic of Letters Conference, Fondazione Cini, San Giorgio Maggiore, Venice, March 16-18, 2011.

Van den Heuvel, C.M.J.M. and Coleman, C.N. (2011), “Visualizing Uncertainty and Complexity: Humanistic Methods for Mapping the Intellectual Geography of the Early Modern World”, presentation held at the conference ‘Intellectual Geography: Comparative Studies, 1550-1700’, St Anne’s College, University of Oxford, September 5-7, 2011.

Van den Heuvel, C.M.J.M. (2011), “Visualizing Universes of Knowledge” Winsor Lecture- Graduate School for Library and Information Sciences, invited lecture held at the University of Illinois, Urbana-Champaign, October 5, 2011.

Van den Heuvel, C.M.J.M. (2011), “Circulation of Knowledge and Learned Practices in the 17th century Dutch Republic. A Web-based Collaboratory on Correspondences. Expectations-Experiences-Recommendations. Invited lecture held at the CATCH-Meeting HiTIME: e-History colloquium, International Institute of Social History, November 1, 2011.

Vriend, F. de (2011), "COAVA", presentation held at the CLARIN-NL Kick-off meeting Call 2, Utrecht, February 9, 2011.

Windhouwer, M. (2011), "The ISO-DCR", presentation held at the CMDI Tutorial, Nijmegen, January 17, 2011.

Withers, P. (2011), "ARBIL, a tool for organising corpus/language data", presentation held at the CMDI Tutorial, Nijmegen, January 17, 2011.

2012

Bennis, H. (2012), “Introduction”, presentation held at the Call 4 Information Session, Utrecht, August 30, 2012.

Beunders, H. & Kemman, M. (2012), “PoliMedia: Interlinking multimedia for the analysis of media coverage of political debates”, presentation held at the Call 3 kick off meeting, Utrecht, March 6, 2012.

Bloothooft, G. (2012), “MIGMAP”, presentation held at the Call 3 kick off meeting, Utrecht, March 6, 2012.

Boot, P. (2012), “EMIT-X”, presentation held at the Call 3 kick off meeting, Utrecht, March 6, 2012.

Bruin, M. (2012), “C-DSD”, presentation held at the Call 4 Information Session, Utrecht, August 30, 2012.

Crasborn, O. (2012), “MultiCon”, presentation held at the Call 3 kick off meeting, Utrecht, March 6, 2012.

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

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

Dalen-Oskam, K. van (2012), “Namescape”, presentation held at the Call 3 kick off meeting, Utrecht, March 6, 2012.

Doornik, J. van & Boer, V. de (2012), “VK”, presentation held at the Call 4 Information Session, Utrecht, August 30, 2012.

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.

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

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.

Jong, J. de (2012), “FESLI”, presentation held at the Call 3 kick off meeting, Utrecht, March 6, 2012.

Kemps-Snijders, M. (2012), “TTNWW”, presentation held at the Call 4 Information Session, Utrecht, August 30, 2012.

Klijn, E. (2012), “Verrijkt Koninkrijk. Het digitale voortbestaan van Loe de Jongs Tweede Wereldoorlog”, presentation held at the Call 3 kick off meeting, Utrecht, March 6, 2012.

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.

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

Odijk, 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.

Odijk, J.E.J.M. (2012), “Linguistic Research and the CLARIN Infrastructure”, Utrecht, Digital Humanities Lecture, Utrecht, October 23, 2012.

Odijk, J.E.J.M. (2012), “Introduction”, presentation held at the Call 3 kick off meeting, Utrecht, March 6, 2012.

Odijk, J.E.J.M. (2012), “CLARIN-NL en Call 4”, presentation held at the Call 4 Information Session, Utrecht, August 30, 2012.

Oostdijk, N. (2012), ‘Data Curation Service’, presentation held at the meeting of the Dutch Society of Phonetic Sciences, 13 December 2012 in Utrecht.

Oostdijk, N. & Heuvel, H. van den (2012), ‘Introducing the CLARIN-NL Data Curation Service’, presentation held at the Eight International Conference on Language Resources and Evaluation (LREC'12) May 23rd, 2012. Istanbul, Turkey.

Piersma, H. (2012), “WiP”, presentation held at the Call 3 kick off meeting, Utrecht, March 6, 2012.

Pieters, T. (2012), “BILAND”, presentation held at the Call 3 kick off meeting, Utrecht, March 6, 2012.

Pieters, T. and Snelder, S. (2012), "Towards a flexible and stable CALRIN-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.

Post, P. (2012), “PILNAR: The profile of the modern pelgrim”, presentation held at the Call 3 kick off meeting, Utrecht, March 6, 2012.

Quené,H. & Orr, R. (2012), “Database of the longitudinal Utrecht Collection of English Accents (D-LUCEA)”, presentation held at the Call 3 kick off meeting, Utrecht, March 6, 2012.

Sanders, T. (2012), “DiscAn: why and how we would want to annotate corpora on the discourse level”, presentation held at the Call 3 kick off meeting, Utrecht, March 6, 2012.

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.

Schuurman, I. (2012), “CLARIN? ISOcat!”, presentation held at the Call 3 kick off meeting, Utrecht, March 6, 2012.

Schuurman, I. (2012), “ISOcat”, presentation held at the Call 4 Information Session, Utrecht, August 30, 2012.

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

Schuurman, I. & M. Windhouwer (2012), “Call 3 ISOcat follow-up”, presentation held at the ISOCAT Call 3 Follow-up Workshop, Utrecht October 10, 2012.

Sluis, R. van (2012), “NEHOL”, presentation held at the Call 4 Information Session, Utrecht, August 30, 2012.

Sluiter, I. (2012), “Een Grieks-Nederlands woordenboek voor de 21e eeuw”, presentation held at the Call 3 kick off meeting, Utrecht, March 6, 2012.

Swanenberg, J., Cornips, L.M.E.A., Vriend, F. de & Heeringa, W.J. “Is what we have acquired early, less vulnerable to variation? A comparison between data from dialect lexicography and data from first language acquisition”, presentation held at the VIIth Congress of the International Society for Dialectology and Geolinguistics (SIDG), Vienna, July 24, 2012.

Vossen, P. (2012), “Cornetto-LMF-RDF”, presentation held at the Call 3 kick off meeting, Utrecht, March 6, 2012.

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.

Wyatt, S. (2012), “eHumanities within the KNAW”, presentation held at the Call 4 Information Session, Utrecht, August 30, 2012.

2013

Barbiers, S. (2013), MIMORE, presentation at the CLARIN Annual Conference, Prague, Oct 23, 2013, [url]

Blom, J. (2013), ‘A demonstration of PoliMedia’, presentation held at the PoliMedia Symposium, Amsterdam, January 23, 2013.

Bloothooft, G. (2013), MIGMAP, presentation at the CLARIN Annual Conference, Prague, Oct 22, 2013, [url]

De Boer, Victor & Matthias van Rossum (2013) 'DSS', presentation held at the CLARIN-NL Call 4 Kick-off Meeting, Utrecht, February 21, 2013.

Victor de Boer, Johan van Doornik, Lars Buitinck, Maarten Marx, Tim Veken, Kees Ribbens. Verrijkt Koninkrijk: Linking a Historiographical Text to the Web of Data. Poster presentation at the E-Humanities workshop, Soeterbeeck, the Netherlands, 13-14 June 2013. [Abstract] [Poster]

Bouwman, A (2013) 'e-BNM+: Linked data on Middle Dutch sources kept worldwide', presentation held at the CLARIN-NL Kick-off Meeting Call4, Utrecht, February 21, 2013.

Bouwman, A (2013) 'Het project e-BNM+', presentation held at the Tweede Dag van de Medioneerlandistiek, Utrecht, June 13, 2013.

Broeder, D. (2013). CMDI registry. poster at the CLARIN Annual Conference, Prague, Oct 23, 2013.

Broeder, Daan (2013) 'ColTime: Collaboration on Time-Based Resources', presentation held at the CLARIN-NL Call 4 Kick-off Meeting, Utrecht, February 21, 2013

Broeder, Daan (2013) 'AAI SSO', presentation held at the CLARIN-NL Call 4 Kick-off Meeting, Utrecht, February 21, 2013.

Broeder, D., Vis, K., Sanders, T. (2013). *Building a discourse annotated corpus for Dutch.* Presentation at International workshop on discourse annotation, Utrecht, The Netherlands, June 21, 2013. [pdf]

Dijk, Suzan van (2013) 'COBWWWEB: Transnational sharing of data about women's authorship (production, circulation, reception up to the early 20th century)', presentation held at the CLARIN-NL Call 4 Kick-off Meeting, Utrecht, February 21, 2013

Filarski, G. (2013), eLaborate/ePistolarium, poster at the CLARIN Annual Conference, Prague, Oct 23, 2013.

Heuvel, H. van den (2013). The CLARIN-NL Data Curation Service, presentation at the 8th International Digital Curation Conference: "Infrastructure, Intelligence, Innovation: driving the Data Science agenda", 14-17 January 2013, Amsterdam, Netherlands [recording]

Hoogervorst, Tom G. and Marian Klamer. 'Language Archive of Insular SE Asia & W New Guinea (LAISEANG)' . Poster and Power Point demonstration at the *Clarin-NL Call 3: International Advisory Panel Demonstrator Event*, Utrecht, 4 Sept 2013. [pdf]

Hoogervorst, Tom G. 'The Role of Language Archiving in the Preservation of Local Traditions' Institute of East Asian Studies seminars (University of Malaysia Sarawak), 14 June 2013, Kuching, Malaysia. [pdf]

Hoogervorst, Tom G. 'The Role of Language Archiving in the Preservation of Local Traditions' *International Congress on Asia Folklore 2013*, panel Folklore and Folklife in the Modern World: Unity and Multi Cultured Pluralism, 7-9 June 2013, Yogyakarta, Indonesia. [pdf]

Juric, D. (2013) PoliMedia symposium: linking political debates and media. Amsterdam, 23 January 2013. Linking the data sets slideshare .

Kemman, M. (2013) PoliMedia symposium: linking political debates and media. Amsterdam, 23 January 2013. User research for the development of PoliMedia slideshare

Kemman, M. (2013), Building the PoliMedia search system; data- and user-driven. eHumanities Group. Amsterdam, 18 april 2013 slideshare

Kemps-Snijders, M. (2013). CMDI Search, poster at the CLARIN Annual Conference, Prague, Oct 23, 2013.

Kemps-Snijders, M. (2013). TTNWW, poster at the CLARIN Annual Conference, Prague, Oct 23, 2013.

Klamer, Marian. *Minority languages in Indonesia: Challenges and Opportunities*. 7th EuroSEAS, Lisbon, 2-5 July 2013. [pdf]

Klatter, Jetske (2013), 'Vulnerability in Acquisition: Language Impairments in Dutch: Creating a CLARIN-supported VALID Data Archive', presentation held at the CLARIN-NL Call 4 Kick-off Meeting, Utrecht, February 21, 2013

Kleppe, M. (2013). PoliMedia. IAMHIST. Leicester, UK, 18 juli 2013 slideshare

Kleppe, M. (2013) PoliMedia symposium: linking political debates and media. Amsterdam, 23 January 2013. Introduction to PoliMedia slideshare

Kok, W. (2013), BlackLab: A corpus retrieval engine, poster at the CLARIN Annual Conference, Prague, Oct 23, 2013.

Odijk, J. (2013, Feb 21). *CLARIN-NL Call 4 Kick-Off Introduction.* Utrecht, CLARIN-NL Call 4 Kick-Off. [ppt]

Odijk, J. (2013), `Example Queries for Federated Search', use case for the CLARIN Federated Content Search Workshop, Copenhagen, April 24, 2013. [docx]

Odijk, J. (2013), Example Queries for Federated Search', presentation held at the CLARIN Federated Content Search Workshop, Copenhagen, April 24, 2013 [pptx]

Odijk, J. (2013), "CLARIAH", presentation held at the EuroRisNet+ workshop *Research Infrastructures towards 2020,* Lisbon, July 5, 2013. [pptx] [URL]

Odijk, J. (2013), "Comparative Linguistic Research in the CLARIN Infrastructure", presentation held at the Language Diversity Congress, Groningen. 18 July 2013. [pptx]

Odijk, J. (2013) "Interim Evaluation", presentation held at the IAP Demonstrator Event, Utrecht September 4, 2013 [pptx]

Odijk, J. (2013) 'The CLARIN Infrastructure (NL Part)' presentation held at the IAP Demonstrator Event, Utrecht September 4, 2013 [pptx]

Odijk, J. (2013) 'The CLARIN Infrastructure (NL Part): Current State and Near Future', key note presentation held at the Digital Humanities Summer School, Leuven, September 20, 2013. [pptx]

Oostdijk, N. & Heuvel, H. van den (2013), 'The CLARIN-NL Data Curation Service: Bringing Data to the Foreground', presentation held at the International Data Curation Conference (IDCC 2013), Amsterdam, January 16, 2013.

Oostdijk, N. (2013), 'Data Curation Service', presentation held at the Soeterbeeck eHumanities Workshop, June 13 and 14, 2013, Soeterbeeck, Ravenstein, the Netherlands.

Peursen, Wido van (2013) 'SHEBANQ: System for HEBrew Text: ANnotations for Queries and Markup', presentation held at the CLARIN-NL Call 4 Kick-off Meeting, Utrecht, February 21, 2013

Quené, Hugo (2013) 'D-LUCEA', presentation held at the CLARIN-NL Call 4 Kick-off Meeting, Utrecht, February 21, 2013

Reynaert, Martin (2013) '@PhilosTEI/TiCCLing Philosophy', presentation held at the CLARIN-NL Call 4 Kick-off Meeting, Utrecht, February 21, 2013

Rijke, Maarten de 92013) 'QuaMeRDES', presentation held at the CLARIN-NL Call 4 Kick-off Meeting, Utrecht, February 21, 2013

Sanders, T., Scholman, M. (2013). *Categories of coherence relations in discourse annotation.* Presentation at International workshop on discourse annotation, Utrecht, The Netherlands, June 21, 2013. [pdf]

Schuurman, T. & Kemps-Snijder, M. (2013), 'TST Tools voor het Nederlands als Webservices in Workflow (TTNWW), a CLARIN pilot', presentation held at CLIN, Enschede, January 18, 2013.

Sloetjes, Han (2013), 'EXILSEA: Exploiting ISOcat's Language Sections in ELAN and ANNEX', presentation held at the CLARIN-NL Call 4 Kick-off Meeting, Utrecht, February 21, 2013

Trilsbeek, Paul. 'Het LAISEANG project', *CLARIN-NL Kick-off meeting Call 4*, 21 February 2013, Utrecht, the Netherlands. [pdf]

Verberne, Suzan (2013) 'A Digital Workbench for Rembrandt Research', presentation held at the CLARIN-NL Call 4 Kick-off Meeting, Utrecht, February 21, 2013

Westerhout, E. & J. Odijk (2013), 'Metadata for tools: creating a CMDI profile for tools', presentation held at CLIN 2013, Enschede, Jan 18, 2013. [pdf]

Windhouwer, M. (2013), ISOCAT. poster at the CLARIN Annual Conference, Prague, Oct 23, 2013.

2014

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

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

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

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

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

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

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

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

Brugman, Hennie, Nicoline van der Sijs, René van Stipriaan, Erik Tjong Kim Sang (2014), Presentation of Nederlab for the Swedish Språkbanken, Oct 23, 2014, Meertens Institute.

Brugman, Hennie, Nicoline van der Sijs, René van Stipriaan, Antal van den Bosch (2014), Workshop voor adviesraden Nederlab, Nov 17, 2014, Meertens Institute.

Crasborn, O. (2014) Quelques nouvelles sur ELAN et SignBank. Invited lecture, Université de Namur, Belgium.

Crasborn, O. & H. Sloetjes, Improving the Exploitation of Linguistic Annotations in ELAN. Proceedings of LREC 2014, Reykjavik, Iceland. [URL]

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

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

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

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

Hollink, L. (2014). Connecting political data to media data. Presentation at ASCoR Spring Colloquium 'Big Data at the University of Amsterdam', 18 February 2014, Amsterdam, Netherlands.

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

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

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

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

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

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

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

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

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

Odijk, J. `Searching for Constructions with GrETEL, Syntax Interface Meeting, Mar 31, 2014 [pptx]

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

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

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

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

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

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

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

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

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

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

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

Sijs, N. van der (2014), 'Nederlab voor historisch-letterkundigen', at the Workshop Databank literaire oganisaties, Nov 24, 2014, Utrecht.

Sijs, N. van der (2014), 'Nederlab voor archivarissen', KVAN-dagen 2014, Jun 16, 2014, Assen.

Sijs, N. van der (2014), 'De mogelijkheden van Nederlab', Lezingenmiddag voor studenten, Meertens Institute, May 15, 2014.

Sijs, N. van der (2014), 'Het onderzoeksportaal Nederlab', voor de PI-groep Language in Transition Stages van de RUN, Apr 14, 2014.

Sloetjes, H., Somasundaram, A., & Seibert, O. (2014). New developments in the multimedia annotation tool ELAN. Poster presented at the 6th Conference of the International Society For Gesture Studies: Gesture in Interaction (ISGS 2014), San Diego, CA.

Stipriaan, R. van (2014), 'The Diachronic Dimension. Time and Space in Dutch Cultural History', workshop Using Large-scale Text Collections for Research, Apr 1-2, 2014 Wuerzburg.

Tjong Kim Sang, Erik (2014), 'Nederlab: A complete archive of the historical development of Dutch' , University of Götenborg, Mar 11, 2014.

Tjong Kim Sang, Erik (2014), 'Finding Characteristics of Surinamese Dutch', presentatie held at tye conference DH Benelux, Den Haag, Nederland, Jun 12, 2014.

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

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

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

2015

Boer, V. de et al. (2015), 'Dutch Ships and Sailors', demo at the CLARIN-NL Closing Event, Hilversum, March 13, 2015.

Bosch, Antal van den, Hennie Brugman en René van Stipriaan (2015), 'Lancering en demonstratie van Nederlab', March 13, 2015 Beeld en Geluid at the CLARIAH kick-off.

Bouma, G., van Koppen, J. M., Landsbergen, F., Odijk, J. E. J. M., van der Wouden, T. & van de Camp, M. (2015), 'Enriching a Descriptive Grammar with Treebank Queries', Presentation at the *Fourteenth International Workshop on Treebanks and Linguistic Theories (TLT14)*.

Bouwman, A. (2015), 'BMN-I: Linked Data on Middle Dutch Sources Kept Worldwide', demo at the CLARIN-NL Closing Event, Hilversum, March 13, 2015.

Broeder, D. (2015), 'CLAPOP, the CLARIN-NL Portal', demo at the CLARIN-NL Closing Event, Hilversum, March 13, 2015.

Camp, M. van de et al. (2015), 'OpenSONAR', demo at the CLARIN-NL Closing Event, Hilversum, March 13, 2015.

Crasborn, O. (2015) COLTIME: Presentation/demonstration at closing event of CLARIN-NL, Hilversum

Crasborn, O. (2015) EXILSEA: Presentation/demonstration at closing event of CLARIN-NL, Hilversum.

Dijk, S. van (2015), 'COBWWWEB', demo at the CLARIN-NL Closing Event, Hilversum, March 13, 2015.

Goosen, Twan, Thomas Eckart, Matej Durco, Davor Ostojic (2015), 'VLO developments', presentation and demo at the CLARIN Conference, Wrocław, Poland, 16 October 2015.

Gorp, J. van (2015), 'AVResearcherXL', demo at the CLARIN-NL Closing Event, Hilversum, March 13, 2015.

Landsbergen , Frank, Ton van der Wouden, Jorik van Engeland, Erlinde Meertens, en Sylvia Wijsman (2015), 'Enriching a grammatical database with intelligent links to linguistic resources', demo at the CLARIN 2015 Conference, Wrocław, Poland, 16 October 2015.

Mak, I. et al (2015), 'CORNETTO', demo at the CLARIN-NL Closing Event, Hilversum, March 13, 2015.

Niestad, J. et al. (2015), 'AutoSearch', demo at the CLARIN-NL Closing Event, Hilversum, March 13, 2015.

Niestadt, Jan (2015), BlackLab: advantages and disadvantages of using Lucene for a linguistic corpus search system. Presentation at CLIN26, Amsterdam, 18 December 2015.

Noord, G. van (2015), 'PaQu', demo at the CLARIN-NL Closing Event, Hilversum, March 13, 2015.

Odijk, J. (2015), 'CLARIN-NL: Contributions by NLP & Current Status', presentation at CLIN25 2015, Antwerp, Feb 6, 2015 [pptx] [pdf]

Odijk, J (2015), 'Linguistic Research with PaQu', poster at CLIN25 2015, Antwerp, Feb 6, 2015 [pptx] [pdf]

Odijk, J. (2015), 'What CLARIN has to offer to linguists', plenary lecture at the TIN-dag, Utrecht, Feb 7, 2015. [pptx] [pdf]

Odijk, J. (2015), 'Investigating *heel, erg, zeer* with PaQu', lecture at the TIN-dag, Utrecht, Feb 7, 2015. [pdf]

Odijk, J. (2015), 'CLARIAH: A Research Infrastructure for the Humanities in the Netherlands', invited lecture at the Austrian Academy of Sciences, Vienna, 9 April 2015. [pptx]

Odijk, J. (2015), `Infrastructure for Digital Humanities', lecture at the Utrecht Digital Humanities Session, Utrecht, 10 June 2015. [pptx]

Odijk, J. (2015), 'Multiword Expressions and LMF', invited lecture held at the PARSEME Hands-on workshop on the lexical encoding of multiword expressions, Iaşi, Romania, 22 September 2015. presentation: [ppt] [pdf] ; handout: [pdf]

Odijk, J. (2015), 'PARSEME Alpino MWE Encoding', lecture held at the PARSEME Hands-on workshop on the lexical encoding of multiword expressions, Iaşi, Romania, 23 September 2015. [ppt] and dataset

Odijk, J. (2015), 'Zoeken naar Constructies', presentation and poster held at the DRONGO Language Festival, Utrecht, 26 September 2015. [presentation] [poster]

Odijk, J. (2015), 'Taal- en Spraaktechnologie (TST) voor het Nederlands: Status 2015', presentation held at the Taalcongres 'Taal schept Kansen', Flemish Parliament, Brussels, 10 October 2015. [pptx]

Odijk, J. (2015), 'Using PaQu for language acquisition research', presentation at the CLARIN 2015 Conference, Wrocław, Poland, 16 October 2015 [pptx] [pdf]

Odijk, J. (2015), 'Using PaQu for language acquisition research', demo at the CLARIN 2015 Conference, Wrocław, Poland, 16 October 2015 [pdf]

Odijk, J (2015), 'Can we extend Alpino with flexible MWEs? An unexpected possible use of PaQu', presentation and demo at the CLARIN Conference, Wrocław, Poland, 16 October 2015. [pdf] [data] [handout] [analysis]

Odijk, J. (2015), 'Linguistic Research with CLARIN', MA Rotation Lecture, Utrecht, 2015-11-10. [pptx] [pdf]

Odijk, J. (2015), 'Extending Alpino with Flexible MWEs. An unexpected possible use of PaQu', poster and presentation/demo at CLIN26, Amsterdam, 18 December 2015 [poster] [presentation]

Peursen, W. van & D. Roorda (2015), 'SHEBANQ', demo at the CLARIN-NL Closing Event, Hilversum, March 13, 2015.

Peursen, Wido van, Dirk Roorda and Martijn Naaijer (2015), 'SHEBANQ: Annotations based hub of research in the Hebrew Bible'. Presentation at CLIN26, Amsterdam, 18 December 2015.

Reynaert, M. et al. (2015), 'PICCL**:** Philosophical Integrator of Computational and Corpus Libraries**'**, demo at the CLARIN-NL Closing Event, Hilversum, March 13, 2015.

Reynaert, Martin, Maarten Van Gompel, Ko van der Sloot and Antal van Den Bosch (2015), 'PICCL: Philosophical Integrator of Computational and Corpus Libraries'. Presentation at the CLARIN 2015 Conference, Wrocław, Poland, 16 October 2015.

Roorda, Dirk (2015), 'Hebrew Bible: System of Data and Annotations', presentation and demo at the CLARIN Conference, Wrocław, Poland, 16 October 2015.

Schuurman, I, & M. Windhouwer (2015), 'The CLARIN Concept Registry (CCR)', poster and demo at the CLARIN-NL Closing Event, Hilversum, March 13, 2015.

Schuurman, Ineke, Menzo Windhouwer, Oddrun Ohren & Daniel Zeman (2015). 'CLARIN Concept Registry: the new semantic registry', presentation at the CLARIN 2015 Conference, Wrocław, Poland, 17 October 2015. [pptx] [pdf]

Schuurman, Ineke, Menzo Windhouwer and Olha Shkaravska (2015). **'**The CLARIN Concept Registry'. Presentation at CLIN26, Amsterdam, 18 December 2015.

Sloetjes, H., Seibert, O., Somasundaram, A., & Drude, S. (2015). Language identification in ELAN. E-poster presented at the 4th International Conference on Language Documentation & Conservation (ICLDC), Honolulu, HI.

Verberne, S. (2015), 'RemBench', demo at the CLARIN-NL Closing Event, Hilversum, March 13, 2015.

Windhouwer, Menzo (2015), 'CMDI first aid kit', presentation and demo at the CLARIN Conference, Wrocław, Poland, 16 October 2015.

Wouden, Ton van der, Gosse Bouma, Matje van de Kamp, Marjo van Koppen, Frank Landsbergen and Jan Odijk (2015), 'Enriching a grammatical database with intelligent links to linguistic resources', presentation at the CLARIN 2015 Conference, Wrocław, Poland, 16 October 2015 [pdf]

Wynne, Martin and Arjan van Hessen (2015), 'Working with oral history archive data', poster at the CLARIN 2015 Conference, Wrocław, Poland, 16 October 2015.

2016

Komen, Erwin (2016). Corpus Studio Web Version. Demo and poster at the CLARIAH-dag, Amersfoort, January 22.

Noord, Gertjan van & Jan Odijk (2016). `Goed of Fout', demonstration at the CLARIAH-day, Amersfoort, January 22, 2016. [pdf] [pptx]

Noord, Gertjan van & Jan Odijk (2016). `Goed of Fout: Wat gebruikt men feitelijk?', presentation at the `Grote Taaldag' (TIN), Utrecht, 6 February 2016. [handout] [pptx] [pdf]

Reynaert, Martin & Matje van de Camp (2016), OpenSoNaR-CGN. Poster and demo at the CLARIAH-dag, Amersfoort, 22 January.

Sijs, Nicoline van der, Henk van den Heuvel, Roeland van Hout & Eric Sanders (2016), 'CARE: Curation of Dutch Regional Dialect Dictionaries', demo en poster presentation at the CLARIAH-dag , Amersfoort, 22 January.

Wouden, Ton van der and Frank Landsbergen (2016). ''Enriching a grammatical database with intelligent links to linguistic resources'. Poster and demo at the CLARIAH-dag, Amersfoort, 22 January.

Wouden, Ton van der (2016). A boost for the digital infrastructure for linguistics: the Taalportaal. DH Benelux 2016.

Wouden, Ton van der (2016). Taalportaal: A new tool for linguistic research. Submitted & accepted DH 2016 Krakow.

Awards, Prizes

2013

PoliMedia wins the Veni LinkedUp Challenge

Polimedia is finalist of the Semantic Web Challenge.

2014

SHEBANQ wins the Digital Humanities Award 2014 on the category *Best Tool or Suite of Tools*

http://www.clarin.nl/node/2085 and http://dhawards.org/

2015

PoliMedia wins LODLAM Open Data Prize

During the LODLAM Challenge in Sydney in the summer of 2015, CLARIN-NL project PoliMedia has been awarded the Open Data Prize. Out of 39 entries, the jury judged PoliMedia to be the most innovative project because of its optimal use of semantic techniques and transparent process in opening up the generated links to other researchers.

CLARIN Young Scientist Award for Maarten van Gompel

At the annual CLARIN Conference 2015 held in Wrocław, Poland, Maarten van Gompel (RU Nijmegen) has been awarded the CLARIN Young Scientist Award in recognition of his work on the design, development, and support of CLAM (Computational Linguistics Application Mediator, a web service shell) and FoLiA (Format for Linguistic Annotation, an XML standard). “Both tools have had a major impact on language resource infrastructure development in the Netherlands and Flanders in the past four years.” See also http://www.clarin.nl/node/2104 .