## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLARIN-NL Objectives</td>
<td>6</td>
</tr>
<tr>
<td>Success Criteria</td>
<td>7</td>
</tr>
<tr>
<td>Global Budget</td>
<td>9</td>
</tr>
<tr>
<td>Governance</td>
<td>11</td>
</tr>
<tr>
<td>Board</td>
<td>11</td>
</tr>
<tr>
<td>Executive Board</td>
<td>11</td>
</tr>
<tr>
<td>National Advisory Panel</td>
<td>11</td>
</tr>
<tr>
<td>International Advisory Panel</td>
<td>12</td>
</tr>
<tr>
<td>Project office</td>
<td>12</td>
</tr>
<tr>
<td>CLARIN Centres</td>
<td>13</td>
</tr>
<tr>
<td>CLARIN Data Providers</td>
<td>14</td>
</tr>
<tr>
<td>11-027 EdiTE</td>
<td>14</td>
</tr>
<tr>
<td>11-030 Sound and Vision as CLARIN Data provider</td>
<td>14</td>
</tr>
<tr>
<td>11-036 KB as Data Provider</td>
<td>15</td>
</tr>
<tr>
<td>13-008 Utrecht University Library</td>
<td>15</td>
</tr>
<tr>
<td>Calls</td>
<td>16</td>
</tr>
<tr>
<td>Texts (integrally taken from original publications)</td>
<td>16</td>
</tr>
<tr>
<td>Call 1 – Open Call (2009)</td>
<td>16</td>
</tr>
<tr>
<td>Call 2 – Open Call (2010)</td>
<td>26</td>
</tr>
<tr>
<td>Call 3 (2011)</td>
<td>37</td>
</tr>
<tr>
<td>Call 4 (2012)</td>
<td>59</td>
</tr>
<tr>
<td>Call 3 (Open Call)</td>
<td>93</td>
</tr>
<tr>
<td>Call 3 (Closed Call)</td>
<td>107</td>
</tr>
<tr>
<td>Call 4 (Open Call)</td>
<td>120</td>
</tr>
<tr>
<td>Call 4 (Closed Call)</td>
<td>140</td>
</tr>
<tr>
<td>Projects</td>
<td>160</td>
</tr>
<tr>
<td>Call 1</td>
<td>160</td>
</tr>
<tr>
<td>Call 2</td>
<td>166</td>
</tr>
<tr>
<td>Call 3</td>
<td>171</td>
</tr>
<tr>
<td>Call 4</td>
<td>179</td>
</tr>
<tr>
<td>2013 Projects</td>
<td>184</td>
</tr>
<tr>
<td>13-012 Lassy</td>
<td>184</td>
</tr>
</tbody>
</table>
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
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. Interoperability 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
Objectives

- 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:

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

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

<table>
<thead>
<tr>
<th>Centre</th>
<th>Focus on resources</th>
<th>Resource examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meertens Institute</td>
<td>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.</td>
<td>Typological databases for Dutch dialects, descriptions of Dutch songs, databases of Dutch names, etc.</td>
</tr>
<tr>
<td>Max Planck Institute for Psycholinguistics</td>
<td>Related to the study of psychological, social and biological foundations of language.</td>
<td>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.</td>
</tr>
<tr>
<td>Huygens ING</td>
<td>Related to the study of history and literature of the Netherlands.</td>
<td>Historical and literary manuscripts and their annotations, tools to annotate such manuscripts and create scholarly editions, etc.</td>
</tr>
<tr>
<td>Institute for Dutch Lexicology</td>
<td>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.</td>
<td>Lexicons, lexical databases, text corpora, speech corpora, language and speech technology tools, etc.</td>
</tr>
<tr>
<td>Data Archiving and Networking Services</td>
<td>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”.</td>
<td>Text resources, databases, spreadsheets, audio and audio-visual resources.</td>
</tr>
</tbody>
</table>
**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 *Nationale 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))

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dutch Library for Dutch Literature (DBNL)</td>
<td>€30,000</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€30,000</strong></td>
</tr>
</tbody>
</table>

**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

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Netherlands Institute for Sound and Vision</td>
<td>€30,000</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€30,000</strong></td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Library of the Netherlands</td>
<td>€30,000</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€30,000</strong></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utrecht University Library</td>
<td>€12,350</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€12,350</strong></td>
</tr>
</tbody>
</table>

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-
Calls - Texts (integrally taken from original publications)

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
Calls - Texts (integratedly taken from original publications)

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**

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

**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**

<table>
<thead>
<tr>
<th>Activity</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLARIN-NL Second Call Open</td>
<td>2010-07-06</td>
</tr>
<tr>
<td>CLARIN-NL Second Call Information Session</td>
<td>2010-08-26 (afternoon)</td>
</tr>
<tr>
<td>Deadline Proposal Submission</td>
<td>2010-09-29 13:00hrs</td>
</tr>
</tbody>
</table>

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.
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
Calls - Texts (integratedly taken from original publications)

- 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

<table>
<thead>
<tr>
<th>Activity</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLARIN-NL Third Call Open</td>
<td>Sunday July 3, 2011</td>
</tr>
<tr>
<td>CLARIN-NL Third Call Information Session</td>
<td>Thursday August 25, 2011 (afternoon)</td>
</tr>
<tr>
<td>Deadline Proposal Submission</td>
<td>Friday September 30, 2011 13:00hrs CET</td>
</tr>
</tbody>
</table>

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
Calls - Texts (integranly taken from original publications)

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.

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).
Calls - Texts (integralely taken from original publications)

<table>
<thead>
<tr>
<th>Activity</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLARIN-NL Third Call Open</td>
<td>Sunday July 3, 2011</td>
</tr>
<tr>
<td>CLARIN-NL Third Call Information Session</td>
<td>Thursday August 25, 2011 (afternoon)</td>
</tr>
<tr>
<td>Deadline Proposal Submission</td>
<td>Friday September 30, 2011 13:00hrs CET</td>
</tr>
</tbody>
</table>

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
Calls - Texts (integrally taken from original publications)

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

• Contribution to CLARIN-NL as a whole
  o Conformance to the goals of CLARIN-NL in particular and CLARIN in general and the priorities set within them
  o 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
  o 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.
  o 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.
  o 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
  o 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

<table>
<thead>
<tr>
<th>Activity</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLARIN-NL Fourth Call Open</td>
<td>Monday July 2, 2012</td>
</tr>
<tr>
<td>CLARIN-NL Fourth Call Information Session</td>
<td>Thursday August 30, 2012 (afternoon)</td>
</tr>
<tr>
<td>Deadline Proposal Submission</td>
<td>Wednesday September 26, 2012 13:00hrs CET</td>
</tr>
<tr>
<td>Feedback / Questions from IAP</td>
<td>Monday November 12, 2012</td>
</tr>
<tr>
<td>Response to the IAP feedback</td>
<td>one week after the feedback has been sent by e-mail (normally:</td>
</tr>
</tbody>
</table>
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](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](http://www.clarin.nl/node/130)
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.
<table>
<thead>
<tr>
<th>Term</th>
<th>Expansion</th>
<th>Explanation</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>API</td>
<td>Application Programmers Interface</td>
<td>see there</td>
<td>Wikipedia</td>
</tr>
<tr>
<td>Application</td>
<td>Application software</td>
<td>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</td>
<td>Wikipedia</td>
</tr>
<tr>
<td>Application Programmer</td>
<td></td>
<td>software engineer that develops an application and usually accomplishes this by making (partly) use of other software</td>
<td></td>
</tr>
<tr>
<td>Application Programmers Interface</td>
<td></td>
<td>an interface that allows a piece of software to use data, procedures and methods from another piece of software</td>
<td>Wikipedia</td>
</tr>
<tr>
<td>Application Programmers Interface documentation</td>
<td></td>
<td>documentation of the Application Programmers Interface aimed at application programmers</td>
<td></td>
</tr>
<tr>
<td>CLARIN centre</td>
<td></td>
<td>organisation that has the ambition to become a recognized CLARIN centre. See <a href="http://www.clarin.nl/node/130">http://www.clarin.nl/node/130</a> for more details about CLARIN centres in the Netherlands</td>
<td></td>
</tr>
<tr>
<td>CLARIN server</td>
<td></td>
<td>server hosted by or on behalf of a CLARIN centre that runs services accessible for the (CLARIN) research community via the CLARIN infrastructure</td>
<td></td>
</tr>
<tr>
<td>CLARIN-compliant</td>
<td></td>
<td>For a detailed description see <a href="http://trac.clarin.nl/trac/wiki/WikiStart#CLARIN-compatible">http://trac.clarin.nl/trac/wiki/WikiStart#CLARIN-compatible</a>. By curating a resource as described in this call, you make it CLARIN-compliant</td>
<td></td>
</tr>
<tr>
<td>CMDI</td>
<td>Component-based MetaData</td>
<td>the approach to metadata advocated in CLARIN. For details, see</td>
<td></td>
</tr>
<tr>
<td><strong>Infrastructure</strong></td>
<td><strong><a href="http://www.clarin.eu/cmdi">http://www.clarin.eu/cmdi</a></strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td>-------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMDI component</td>
<td>A metadata component defined in accordance with CMDI.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMDI profile</td>
<td>A metadata profile defined in accordance with CMDI.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Component-based MetaData Infrastructure</td>
<td>the approach to metadata advocated in CLARIN. See <a href="http://www.clarin.eu/cmdi">http://www.clarin.eu/cmdi</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>core component</td>
<td>(of a software application) the whole application minus the user interface</td>
<td></td>
<td></td>
</tr>
<tr>
<td>curation</td>
<td>explained in section <em>Fout! Verwijzingsbron niet gevonden.</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>data category</td>
<td>a name and associated information to express a concept</td>
<td></td>
<td></td>
</tr>
<tr>
<td>data category registry</td>
<td>a data-base or directory where data categories can be stored and viewed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>demonstration scenario</td>
<td>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.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>demonstrator</td>
<td>explained in section <em>Fout! Verwijzingsbron niet gevonden.</em>, p. 145, second paragraph</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethical Issues</td>
<td>Issues related to ethics that may arise by making data available to a wider audience, e.g. privacy violations etc.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Formal description</strong></td>
<td>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</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>infrastructure</strong></td>
<td>(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.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Intellectual Property</strong></td>
<td>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</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Intellectual Property Rights</strong></td>
<td>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.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>interoperability</strong></td>
<td>resources (data and software tools) are interoperable when they can work together in a fully automated manner with minimal human intervention</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>IPR</strong></td>
<td>Intellectual Property Rights; see there</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>language resource</strong></td>
<td>data containing language or a software tool operating on or yielding language data</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>metadata</strong></td>
<td>Literally: data about data; in CLARIN usually restricted to descriptions of language resources (in CMDI format)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>metadata component</td>
<td>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.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>metadata harvesting</td>
<td>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)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>metadata profile</td>
<td>A metadata profile is an XML document definition consisting of metadata components and is used to describe the properties of a language resource</td>
<td></td>
<td></td>
</tr>
<tr>
<td>method</td>
<td>(in programming languages) a special type of procedure or function</td>
<td></td>
<td></td>
</tr>
<tr>
<td>method invocation</td>
<td>(in programming languages) calling a procedure or function</td>
<td></td>
<td></td>
</tr>
<tr>
<td>registry</td>
<td>(digital) data-base or directory to store and access information</td>
<td></td>
<td></td>
</tr>
<tr>
<td>resource</td>
<td>used in this document mostly as equivalent to language resource</td>
<td></td>
<td></td>
</tr>
<tr>
<td>resource curation</td>
<td>see: curation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Service</td>
<td>(here used in a software context) a software program that provides certain functionality and communicates its input and output with other software programs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Service Oriented Architecture</td>
<td>Software architecture that is designed around (distributed) services interacting with each other to provide certain functionality</td>
<td></td>
<td></td>
</tr>
<tr>
<td>web application</td>
<td>an application with a web-based user interface</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Web service</td>
<td>service that communicates with other programs over a network (in particular, the internet)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>wrapper (function)</td>
<td>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</td>
<td>Wikipedia</td>
<td></td>
</tr>
<tr>
<td>XML</td>
<td>eXtensible Markup Language</td>
<td>is a markup language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable.</td>
<td>Wikipedia</td>
</tr>
<tr>
<td>XML element</td>
<td>a particular constituent in a document with XML markup</td>
<td>W3CSchools</td>
<td></td>
</tr>
<tr>
<td>XML Schema</td>
<td>A specific way of defining the structure, content and semantics of XML documents</td>
<td>W3C</td>
<td></td>
</tr>
</tbody>
</table>
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.
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

<table>
<thead>
<tr>
<th>Activity</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLARIN-NL Fourth Call Open</td>
<td>Monday July 2, 2012</td>
</tr>
<tr>
<td>CLARIN-NL Fourth Call</td>
<td>Thursday August 30, 2012</td>
</tr>
</tbody>
</table>
Information Session

<table>
<thead>
<tr>
<th>Event</th>
<th>Date/Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deadline Proposal Submission</td>
<td>Wednesday September 26, 2012 13:00hrs CET</td>
</tr>
<tr>
<td>Feedback / Questions from IAP</td>
<td>Monday November 12, 2012</td>
</tr>
<tr>
<td>Response to the IAP feedback/questions</td>
<td>one week after the feedback has been sent by e-mail (normally: Monday November 19, 2012)</td>
</tr>
<tr>
<td>Decision by the Board</td>
<td>Mid December 2012</td>
</tr>
</tbody>
</table>

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](http://trac.clarin.nl/trac)
e-mail: helpdesk@clarin.nl
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
<table>
<thead>
<tr>
<th>Term</th>
<th>Expansion</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>API</td>
<td>Application Programmers Interface</td>
<td>see there</td>
</tr>
<tr>
<td>Application</td>
<td></td>
<td>a piece of software developed to for a specific task and a specified user interface specifically designed for the targeted user group</td>
</tr>
<tr>
<td>Application software</td>
<td></td>
<td>software engineer that develops an application and usually also making (partly) use of other software</td>
</tr>
<tr>
<td>Application Programmer</td>
<td></td>
<td>an interface that allows a piece of software to use data, procedures from another piece of software</td>
</tr>
<tr>
<td>Interface</td>
<td></td>
<td>documentation of the Application Programmers Interface for the benefit of programmers</td>
</tr>
<tr>
<td>CLARIN centre</td>
<td></td>
<td>organisation that has the ambition to become a recognized CLARIN centre. <a href="http://www.clarin.nl/node/130">http://www.clarin.nl/node/130</a> for more details about CLARIN in the Netherlands</td>
</tr>
<tr>
<td>CLARIN server</td>
<td></td>
<td>server hosted by or on behalf of a CLARIN centre that runs services for the (CLARIN) research community via the CLARIN infrastructure</td>
</tr>
<tr>
<td>CLARIN-compliant</td>
<td></td>
<td>For a detailed description see <a href="http://trac.clarin.nl/trac/wiki/">http://trac.clarin.nl/trac/wiki/</a> WiP compatible. By curating a resource as described in this call, you make it CLARIN-compliant</td>
</tr>
<tr>
<td>CMDI</td>
<td>Component-based MetaData Infrastructure</td>
<td>the approach to metadata advocated in CLARIN. For details, see <a href="http://www.clarin.eu/cmdi">http://www.clarin.eu/cmdi</a></td>
</tr>
<tr>
<td>CMDI component</td>
<td></td>
<td>A metadata component defined in accordance with CMDI.</td>
</tr>
<tr>
<td>CMDI profile</td>
<td></td>
<td>A metadata profile defined in accordance with CMDI</td>
</tr>
<tr>
<td>Component-based MetaData</td>
<td></td>
<td>the approach to metadata advocated in CLARIN. See <a href="http://www.clarin.eu/cmdi">http://www.clarin.eu/cmdi</a></td>
</tr>
<tr>
<td>Infrastructure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>core component</td>
<td></td>
<td>(of a software application) the whole application minus the user interface</td>
</tr>
<tr>
<td>curation</td>
<td></td>
<td>explained in section Fout! Verwijzingsbron niet gevonden.</td>
</tr>
<tr>
<td>data category</td>
<td></td>
<td>a name and associated information to express a concept</td>
</tr>
<tr>
<td>data category registry</td>
<td></td>
<td>a data-base or directory where data categories can be stored and managed</td>
</tr>
<tr>
<td>demonstration scenario</td>
<td>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 offered demonstration scenario should serve as an example of how to and/or solve a particular type of problem.</td>
<td></td>
</tr>
<tr>
<td>Demonstrator</td>
<td>explained in section Fout! Verwijzingsbron niet gevonden., p. 1, paragraph</td>
<td></td>
</tr>
<tr>
<td>Ethical Issues</td>
<td>Issues related to ethics that may arise by making data available to the audience, e.g. privacy violations etc.</td>
<td></td>
</tr>
<tr>
<td>Formal description</td>
<td>a formal description is a description in a form defined by well make it possible to completely and unambiguously determine description from the form</td>
<td></td>
</tr>
<tr>
<td>Infrastructure</td>
<td>(in the CLARIN context) a combination of hardware and software for researcher or research team to find and use language resources, and store own data and metadata to make the research community.</td>
<td></td>
</tr>
<tr>
<td>Intellectual Property</td>
<td>a variety of intangible assets produced by intellectual labor, such as literary, and artistic works; discoveries and inventions; and works symbols, and designs</td>
<td></td>
</tr>
<tr>
<td>Intellectual Property Rights</td>
<td>All the rights and restrictions that are associated with intellectual property the owner, is the resource licensed, what is one allowed to do with it etc. Common types of intellectual property rights include copyrights, patents, etc.</td>
<td></td>
</tr>
<tr>
<td>Interoperability</td>
<td>resources (data and software tools) are interoperable when they can be used together in a fully automated manner with minimal human intervention</td>
<td></td>
</tr>
<tr>
<td>IPR</td>
<td>Intellectual Property Rights</td>
<td>see there</td>
</tr>
<tr>
<td>Language resource</td>
<td>data containing language or a software tool operating on or yielding language resources</td>
<td></td>
</tr>
<tr>
<td>Metadata</td>
<td>Literally: data about data; in CLARIN usually restricted to describe resources (in CMDI format)</td>
<td></td>
</tr>
<tr>
<td>Metadata component</td>
<td>A metadata component is an XML structure containing metadata (recursively) other metadata components; it is used to describe properties of a language resource.</td>
<td></td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>------------</td>
<td></td>
</tr>
<tr>
<td>metadata harvesting</td>
<td>the process of obtaining metadata from a different server (met a network (e.g. to get these metadata available in central data</td>
<td></td>
</tr>
<tr>
<td>metadata profile</td>
<td>A metadata profile is an XML document definition consisting of components and is used to describe the properties of a language</td>
<td></td>
</tr>
<tr>
<td>method</td>
<td>(in programming languages) a special type of procedure or function</td>
<td></td>
</tr>
<tr>
<td>method invocation</td>
<td>(in programming languages) calling a procedure or function</td>
<td></td>
</tr>
<tr>
<td>registry</td>
<td>(digital) data-base or directory to store and access information</td>
<td></td>
</tr>
<tr>
<td>resource</td>
<td>used in this document mostly as equivalent to language resource</td>
<td></td>
</tr>
<tr>
<td>resource curation</td>
<td>see: curation</td>
<td></td>
</tr>
<tr>
<td>Service</td>
<td>(here used in a software context) a software program that provides functionality and communicates its input and output with other</td>
<td></td>
</tr>
<tr>
<td>Service Oriented Architecture</td>
<td>Software architecture that is designed around (distributed) services, each other to provide certain functionality</td>
<td></td>
</tr>
<tr>
<td>web application</td>
<td>an application with a web-based user interface</td>
<td></td>
</tr>
<tr>
<td>Web service</td>
<td>service that communicates with other programs over a network (internet)</td>
<td></td>
</tr>
<tr>
<td>wrapper (function)</td>
<td>is a software component whose main purpose is to make the further software component available and does little or no additional o</td>
<td></td>
</tr>
<tr>
<td>XML</td>
<td>eXtensible Markup Language</td>
<td>is a markup language that defines a set of rules for encoding data that is both human-readable and machine-readable.</td>
</tr>
<tr>
<td>XML element</td>
<td>a particular constituent in a document with XML markup</td>
<td></td>
</tr>
<tr>
<td>XML Schema</td>
<td>A specific way of defining the structure, content and semantics</td>
<td></td>
</tr>
</tbody>
</table>
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. 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](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 ISOCAT4 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. 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). 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

---

6 [http://www.clarin.eu/cmdi](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/#](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](http://www.nwo.nl/nwohome.nsf/pages/NWOP_67QK4E) more specifically
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

<table>
<thead>
<tr>
<th>Activity</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLARIN-NL Third Call Open</td>
<td>Sunday July 3, 2011</td>
</tr>
<tr>
<td>CLARIN-NL Third Call Information Session</td>
<td>Thursday August 25, 2011 (afternoon)</td>
</tr>
<tr>
<td>Deadline Proposal Submission</td>
<td>Friday September 30, 2011 13:00hrs CET</td>
</tr>
</tbody>
</table>

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
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€.

**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.

**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](http://www.clarin.eu/recommendations)
¹¹ [helpdesk@clarin.nl](mailto: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\textsuperscript{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

\textsuperscript{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. 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).\(^\text{14}\) 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.\(^\text{15}\) Any required or

\(^{14}\) [http://www.clarin.eu/cmdi](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/#](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
  o Competence of the participating partners (including their past performance);
  o Balanced cooperation and task assignments within the project. Justification of the composition of the team.
  o Availability of the infrastructure required for the project to be successful
  o 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
  o Does the project address needs of the targeted infrastructure users (linguists and humanities researchers)?
  o 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.
  o Is there cooperation with or support from the targeted (future) infrastructure users?
  o Is the resulting tool/service user-friendly, i.e. will non-technical linguistic and humanities researchers be able to use it?
  o Is dissemination of the results to the targeted users and (where appropriate) training of them planned?

• Contribution to CLARIN-NL as a whole
  o Conformance to the goals of CLARIN-NL in particular and CLARIN in general and the priorities set within them
  o 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
  o 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.
  o 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.
  o 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
  o 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.

U.pdf . See also http://www.nwo.nl/files.nsf/pages/SPES_5VEADDR/$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](http://www.clarin.nl/node/72)).

**Timetable**

<table>
<thead>
<tr>
<th>Activity</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLARIN-NL Third Call Open</td>
<td>Sunday July 3, 2011</td>
</tr>
<tr>
<td>CLARIN-NL Third Call Information Session</td>
<td>Thursday August 25, 2011 (afternoon)</td>
</tr>
<tr>
<td>Deadline Proposal Submission</td>
<td>Friday September 30, 2011 13:00hrs CET</td>
</tr>
</tbody>
</table>

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](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](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. 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.

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:

- The resource needs to be stored at computers that are accessible via the internet.
- The resource needs to be identified in a persistent manner.
- 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, 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. 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).
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 ISOCAT21), 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.

---

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. 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](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). 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.

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


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

\(^{25}\) [link to more information]
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

<table>
<thead>
<tr>
<th>Activity</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLARIN-NL Fourth Call Open</td>
<td>Monday July 2, 2012</td>
</tr>
<tr>
<td>CLARIN-NL Fourth Call</td>
<td></td>
</tr>
<tr>
<td>Information Session</td>
<td>Thursday August 30, 2012 (afternoon)</td>
</tr>
<tr>
<td>Deadline Proposal Submission</td>
<td>Wednesday September 26, 2012 13:00hrs CET</td>
</tr>
</tbody>
</table>
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.
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
<table>
<thead>
<tr>
<th>Term</th>
<th>Expansion</th>
<th>Explanation</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>API</td>
<td>Application Programmers Interface</td>
<td>see there</td>
<td>Wikipedia</td>
</tr>
<tr>
<td>Application software</td>
<td></td>
<td>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</td>
<td></td>
</tr>
<tr>
<td>Application Programmer</td>
<td></td>
<td>software engineer that develops an application and usually accomplishes this by making (partly) use of other software</td>
<td></td>
</tr>
<tr>
<td>Application Programmers Interface</td>
<td></td>
<td>an interface that allows a piece of software to use data, procedures and methods from another piece of software</td>
<td>Wikipedia</td>
</tr>
<tr>
<td>Application Programmers Interface documentation</td>
<td></td>
<td>documentation of the Application Programmers Interface aimed at application programmers</td>
<td></td>
</tr>
<tr>
<td>CLARIN centre</td>
<td></td>
<td>organisation that has the ambition to become a recognized CLARIN centre. See <a href="http://www.clarin.nl/node/130">http://www.clarin.nl/node/130</a> for more details about CLARIN centres in the Netherlands</td>
<td></td>
</tr>
<tr>
<td>CLARIN server</td>
<td></td>
<td>server hosted by or on behalf of a CLARIN centre that runs services accessible for the (CLARIN) research community via the CLARIN infrastructure</td>
<td></td>
</tr>
<tr>
<td>CLARIN-compliant</td>
<td></td>
<td>For a detailed description see <a href="http://trac.clarin.nl/trac/wiki/WikiStart#CLARIN-compatible">http://trac.clarin.nl/trac/wiki/WikiStart#CLARIN-compatible</a>. By curating a resource as described in this call, you make it CLARIN-</td>
<td></td>
</tr>
<tr>
<td>CMDI component</td>
<td>A metadata component defined in accordance with CMDI.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>--------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMDI profile</td>
<td>A metadata profile defined in accordance with CMDI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Component-based</td>
<td>the approach to metadata advocated in CLARIN. See <a href="http://www.clarin.eu/cmdi">http://www.clarin.eu/cmdi</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MetaData Infrastructure</td>
<td>The demonstration scenario should serve as an example of how to use the application and/or solve a particular type of problem.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>core component</td>
<td>(of a software application) the whole application minus the user interface</td>
<td></td>
<td></td>
</tr>
<tr>
<td>curation</td>
<td>explained in section 5.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>data category</td>
<td>a name and associated information to express a concept</td>
<td></td>
<td></td>
</tr>
<tr>
<td>data category registry</td>
<td>a data-base or directory where data categories can be stored and viewed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>demonstration scenario</td>
<td>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).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>demonstrator</td>
<td>explained in section 5.2, p. 145, second paragraph</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethical Issues</td>
<td>Issues related to ethics that may arise by making data available to a wider audience, e.g. privacy violations etc.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------</td>
<td>-------------------------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formal description</td>
<td>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</td>
<td></td>
<td></td>
</tr>
<tr>
<td>infrastructure</td>
<td>(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.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intellectual Property</td>
<td>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 <a href="http://en.wikipedia.org/wiki/Intellectual_property">Wikipedia</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intellectual Property Rights</td>
<td>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. Common types of intellectual property rights include copyrights, trademarks, patents, etc. <a href="http://en.wikipedia.org/wiki/Intellectual_property_rights">Wikipedia</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>interoperability</td>
<td>Resources (data and software tools) are interoperable when they can work together in a fully automated manner with minimal human intervention</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IPR</td>
<td>Intellectual Property Rights see there</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Language resource</strong></td>
<td><strong>Data containing language or a software tool operating on or yielding language data</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td>-----------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Metadata</strong></td>
<td>Literally: data about data; in CLARIN usually restricted to descriptions of language resources (in CMDI format)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Metadata component</strong></td>
<td>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.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Metadata harvesting</strong></td>
<td>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)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Metadata profile</strong></td>
<td>A metadata profile is an XML document definition consisting of metadata components and is used to describe the properties of a language resource</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Method</strong></td>
<td>(in programming languages) a special type of procedure or function</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Method invocation</strong></td>
<td>(in programming languages) calling a procedure or function</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Registry</strong></td>
<td>(digital) data-base or directory to store and access information</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Resource</strong></td>
<td>Used in this document mostly as equivalent to language resource</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Resource curation</strong></td>
<td>See: curation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Service</td>
<td>(here used in a software context) a software program that provides certain functionality and communicates its input and output with other software programs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Service Oriented Architecture</td>
<td>Software architecture that is designed around (distributed) services interacting with each other to provide certain functionality</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Web application</td>
<td>an application with a web-based user interface</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Web service</td>
<td>service that communicates with other programs over a network (in particular, the internet)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>wrapper (function)</td>
<td>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</td>
<td></td>
<td></td>
</tr>
<tr>
<td>XML</td>
<td>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.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>XML element</td>
<td>a particular constituent in a document with XML markup</td>
<td></td>
<td></td>
</tr>
<tr>
<td>XML Schema</td>
<td>A specific way of defining the structure, content and semantics of XML documents</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Call 4 (Closed Call)

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 date. 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.

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:

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.

**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,\(^2\) 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

\(^2\) [http://www.clarin.eu/recommendations](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. 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/](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
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.

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). 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.

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.

---


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

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](http://www.clarin.nl/node/72), which can be found on the CLARIN-NL website.

**Timetable**

<table>
<thead>
<tr>
<th>Activity</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLARIN-NL Fourth Call Open</td>
<td>Monday July 2, 2012</td>
</tr>
<tr>
<td>CLARIN-NL Fourth Call Information Session</td>
<td>Thursday August 30, 2012 (afternoon)</td>
</tr>
<tr>
<td>Deadline Proposal Submission</td>
<td>Wednesday September 26, 2012 13:00hrs CET</td>
</tr>
<tr>
<td>Feedback / Questions from IAP</td>
<td>Monday November 12, 2012</td>
</tr>
<tr>
<td>Response to the IAP feedback /questions</td>
<td>one week after the feedback has been sent by e-mail (normally: Monday November 19, 2012)</td>
</tr>
<tr>
<td>Decision by the Board</td>
<td>Mid December 2012</td>
</tr>
</tbody>
</table>

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.
<table>
<thead>
<tr>
<th>Term</th>
<th>Expansion</th>
<th>Explanation</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>API</td>
<td>Application Programmers Interface</td>
<td>see there</td>
<td>Wikipedia</td>
</tr>
<tr>
<td>Application</td>
<td>Application software</td>
<td>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</td>
<td>Wikipedia</td>
</tr>
<tr>
<td>Application Programmer</td>
<td></td>
<td>software engineer that develops an application and usually accomplishes this by making (partly) use of other software</td>
<td></td>
</tr>
<tr>
<td>Application Programmers Interface</td>
<td></td>
<td>an interface that allows a piece of software to use data, procedures and methods from another piece of software</td>
<td>Wikipedia</td>
</tr>
<tr>
<td>Application Programmers Interface documentation</td>
<td></td>
<td>documentation of the Application Programmers Interface aimed at application programmers</td>
<td></td>
</tr>
<tr>
<td>CLARIN centre</td>
<td></td>
<td>organisation that has the ambition to become a recognized CLARIN centre. See <a href="http://www.clarin.nl/node/130">http://www.clarin.nl/node/130</a> for more details about CLARIN centres in the Netherlands</td>
<td></td>
</tr>
<tr>
<td>CLARIN server</td>
<td></td>
<td>server hosted by or on behalf of a CLARIN centre that runs services accessible for the (CLARIN) research community via the CLARIN infrastructure</td>
<td></td>
</tr>
<tr>
<td>CLARIN-compliant</td>
<td></td>
<td>For a detailed description see <a href="http://trac.clarin.nl/trac/wiki/WikiStart#CLARIN-compatible">http://trac.clarin.nl/trac/wiki/WikiStart#CLARIN-compatible</a>. By curating a resource as described in this call, you make it CLARIN-compliant</td>
<td></td>
</tr>
<tr>
<td>CMDI</td>
<td>Component-based MetaData</td>
<td>the approach to metadata advocated in CLARIN. For details, see</td>
<td></td>
</tr>
<tr>
<td><strong>Infrastructure</strong></td>
<td><strong><a href="http://www.clarin.eu/cmdi">http://www.clarin.eu/cmdi</a></strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------</td>
<td>--------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMDI component</td>
<td>A metadata component defined in accordance with CMDI.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMDI profile</td>
<td>A metadata profile defined in accordance with CMDI.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Component-based MetaData Infrastructure</td>
<td>the approach to metadata advocated in CLARIN. See <a href="http://www.clarin.eu/cmdi">http://www.clarin.eu/cmdi</a>.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>core component</td>
<td>(of a software application) the whole application minus the user interface.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>curation</td>
<td>explained in section 5.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>data category</td>
<td>a name and associated information to express a concept.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>data category registry</td>
<td>a data-base or directory where data categories can be stored and viewed.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>demonstration scenario</td>
<td>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.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>demonstrator</td>
<td>explained in section 5.2, p. 145, second paragraph</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethical Issues</td>
<td>Issues related to ethics that may arise by making data available to a wider audience, e.g. privacy violations etc.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formal description</td>
<td>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</td>
<td></td>
<td></td>
</tr>
<tr>
<td>description from the form</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>infrastructure</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(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.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Intellectual Property</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>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</td>
<td><a href="https://en.wikipedia.org/wiki/Intellectual_property">Wikipedia</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Intellectual Property Rights</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>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. Common types of intellectual property rights include copyrights, trademarks, patents, etc.</td>
<td><a href="https://en.wikipedia.org/wiki/Intellectual_property_rights">Wikipedia</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>interoperability</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>resources (data and software tools) are interoperable when they can work together in a fully automated manner with minimal human intervention</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>IPR</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intellectual Property Rights</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>see there</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>language resource</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>data containing language or a software tool operating on or yielding language data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>metadata</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Literally: data about data; in CLARIN usually restricted to descriptions of language resources (in CMDI format)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>metadata component</td>
<td>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.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>metadata harvesting</td>
<td>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)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>metadata profile</td>
<td>A metadata profile is an XML document definition consisting of metadata components and is used to describe the properties of a language resource</td>
<td></td>
<td></td>
</tr>
<tr>
<td>method</td>
<td>(in programming languages) a special type of procedure or function</td>
<td></td>
<td></td>
</tr>
<tr>
<td>method invocation</td>
<td>(in programming languages) calling a procedure or function</td>
<td></td>
<td></td>
</tr>
<tr>
<td>registry</td>
<td>(digital) data-base or directory to store and access information</td>
<td></td>
<td></td>
</tr>
<tr>
<td>resource</td>
<td>used in this document mostly as equivalent to language resource</td>
<td></td>
<td></td>
</tr>
<tr>
<td>resource curation</td>
<td>see: curation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Service</td>
<td>(here used in a software context) a software program that provides certain functionality and communicates its input and output with other software programs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Service Oriented Architecture</td>
<td>Software architecture that is designed around (distributed) services interacting with each other to provide certain functionality</td>
<td></td>
<td></td>
</tr>
<tr>
<td>web application</td>
<td>an application with a web-based user interface</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Term</td>
<td>Description</td>
<td>Source</td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-------------------------</td>
<td></td>
</tr>
<tr>
<td>Web service</td>
<td>A service that communicates with other programs over a network (in particular, the internet)</td>
<td>Wikipedia</td>
<td></td>
</tr>
<tr>
<td>wrapper (function)</td>
<td>A software component whose main purpose is to make the functions of a second software component available and does little or no additional computation</td>
<td>Wikipedia</td>
<td></td>
</tr>
<tr>
<td>XML</td>
<td>eXtensible Markup Language, a markup language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable.</td>
<td>Wikipedia, W3CSchools</td>
<td></td>
</tr>
<tr>
<td>XML element</td>
<td>A particular constituent in a document with XML markup</td>
<td>W3CSchools</td>
<td></td>
</tr>
<tr>
<td>XML Schema</td>
<td>A specific way of defining the structure, content and semantics of XML documents</td>
<td>W3C</td>
<td></td>
</tr>
</tbody>
</table>
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/

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Huygens ING</td>
<td>€59,250</td>
</tr>
<tr>
<td>Descartes Centre</td>
<td>€0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€59,250</strong></td>
</tr>
</tbody>
</table>

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/

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utrecht University (UiL-OTS)</td>
<td>€0</td>
</tr>
</tbody>
</table>
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 chart 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)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nederlands Gebarencentrum</td>
<td>€10,318</td>
</tr>
<tr>
<td>Radboud University</td>
<td>€27,188</td>
</tr>
<tr>
<td>Max Planck Institute for Psycholinguistics</td>
<td>€18,825</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€56,331</strong></td>
</tr>
</tbody>
</table>

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
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)

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 TICCllops
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
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](http://hdl.handle.net/1839/00-SERV-0000-0000-004F-F)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radboud University Nijmegen</td>
<td>€26,839</td>
</tr>
<tr>
<td>Max Planck Institute for Psycholinguistics</td>
<td>€8,812</td>
</tr>
<tr>
<td>Leiden University</td>
<td>€20,635</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€56,286</strong></td>
</tr>
</tbody>
</table>

Abstract

The AAM-LR project aims at building a demonstrator of a web service that will help field 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/](http://www.gabmap.nl/)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>University of Groningen</td>
<td>€46,367</td>
</tr>
<tr>
<td>Meertens Institute</td>
<td>€10,500</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€56,867</strong></td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Archiving and Networking Services</td>
<td>€19,826</td>
</tr>
<tr>
<td>Veteran Institute</td>
<td>€8,261</td>
</tr>
<tr>
<td>Max Planck Institute for Psycholinguistics</td>
<td>€2,754</td>
</tr>
<tr>
<td>Radboud University Nijmegen</td>
<td>€27,188</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€58,029</strong></td>
</tr>
</tbody>
</table>

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
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 TEI annotation. 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/

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](http://tds2.dans.knaw.nl/TDS.html)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Archiving and Networking Services</td>
<td>€27,329</td>
</tr>
<tr>
<td>Max Planck Institute for Psycholinguistics</td>
<td>€7,613</td>
</tr>
<tr>
<td>Meertens Institute</td>
<td>€17,400</td>
</tr>
<tr>
<td>Utrecht University</td>
<td>€5,159</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€57,501</strong></td>
</tr>
</tbody>
</table>

**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](http://www.meertens.knaw.nl/coavasite/?page_id=8)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tilburg University</td>
<td>€8,946</td>
</tr>
<tr>
<td>Meertens Institute</td>
<td>€57,449</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€66,395</strong></td>
</tr>
</tbody>
</table>

**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.
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
2011-01-01 – 2011-12-31
Projectcoördinator: Dr. Hinke Piersma (NIOD)
http://www.clarin.nl/node/410
http://data.politicalmashup.nl/wip/

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>University of Amsterdam</td>
<td>€72,114</td>
</tr>
<tr>
<td>Data Archiving and Networking Services</td>
<td>€12,878</td>
</tr>
<tr>
<td>NIOD</td>
<td>€33,482</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€118,473</strong></td>
</tr>
</tbody>
</table>

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)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radboud University Nijmegen</td>
<td>€16,248</td>
</tr>
<tr>
<td>Max Planck Institute for Psycholinguistics</td>
<td>€26,644</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€42,892</strong></td>
</tr>
</tbody>
</table>

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/

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descartes Centre</td>
<td>€52,493</td>
</tr>
<tr>
<td>National Library of the Netherlands</td>
<td>€3,675</td>
</tr>
<tr>
<td>University of Amsterdam</td>
<td>€52,493</td>
</tr>
<tr>
<td>Huygens ING</td>
<td>€7,874</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€116,535</strong></td>
</tr>
</tbody>
</table>

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)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meertens Institute</td>
<td>€46,244</td>
</tr>
<tr>
<td>Radboud University Nijmegen</td>
<td>€23,180</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€69,424</strong></td>
</tr>
</tbody>
</table>

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/

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max Planck Institute for Psycholinguistics</td>
<td>€0</td>
</tr>
<tr>
<td>Huygens ING</td>
<td>€39,432</td>
</tr>
<tr>
<td>Utrecht University</td>
<td>€7,727</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€47,158</strong></td>
</tr>
</tbody>
</table>

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/

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>VU University Amsterdam</td>
<td>€38,971</td>
</tr>
<tr>
<td>Institute for Dutch Lexicology</td>
<td>€2,266</td>
</tr>
<tr>
<td>Tilburg University</td>
<td>€17,646</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€58,883</strong></td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utrecht University</td>
<td>€5,151</td>
</tr>
<tr>
<td>Meertens Institute</td>
<td>€24,747</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€29,898</strong></td>
</tr>
</tbody>
</table>

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)
### 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)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>University of Amsterdam</td>
<td>€45,715</td>
</tr>
<tr>
<td>Meertens Institute</td>
<td>€33,114</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€78,829</strong></td>
</tr>
</tbody>
</table>

### 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)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tilburg University</td>
<td>€46,372</td>
</tr>
<tr>
<td>Meertens Institute</td>
<td>€33,114</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€79,486</strong></td>
</tr>
</tbody>
</table>
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)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leiden University</td>
<td>€24,665</td>
</tr>
<tr>
<td>Amsterdam University Press</td>
<td>€6,068</td>
</tr>
<tr>
<td>University of Amsterdam</td>
<td>€42,055</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€72,788</strong></td>
</tr>
</tbody>
</table>

Abstract

This project combines a resource curation and demonstrator project. It concerns the new dictionary (ancient) 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)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Huygens ING</td>
<td>€17,486</td>
</tr>
<tr>
<td>Duke August Library Wolfenbüttel</td>
<td>€0</td>
</tr>
<tr>
<td>University of Illinois at Urbana-Champaign</td>
<td>€0</td>
</tr>
<tr>
<td>Utrecht University</td>
<td>€2,585</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€20,071</strong></td>
</tr>
</tbody>
</table>
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)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utrecht University</td>
<td>€36,190</td>
</tr>
<tr>
<td>University College Utrecht</td>
<td>€10,340</td>
</tr>
<tr>
<td>Radboud University Nijmegen</td>
<td>€5,170</td>
</tr>
<tr>
<td>Max Planck Institute for Psycholinguistics</td>
<td>€16,557</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€68,257</strong></td>
</tr>
</tbody>
</table>

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 issuing 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/

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIDI</td>
<td>€5,170</td>
</tr>
<tr>
<td>Meertens Institute</td>
<td>€33,114</td>
</tr>
</tbody>
</table>
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)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>University of Amsterdam</td>
<td>€36,190</td>
</tr>
<tr>
<td>VU University Amsterdam</td>
<td>€28,435</td>
</tr>
<tr>
<td>Data Archiving and networking Services</td>
<td>€5,519</td>
</tr>
<tr>
<td>Meertens Institute</td>
<td>€5,519</td>
</tr>
<tr>
<td>NIOD</td>
<td>€35,508</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€68,257</strong></td>
</tr>
</tbody>
</table>

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
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)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIA</td>
<td>€1,104</td>
</tr>
<tr>
<td>NIOD</td>
<td>€1,104</td>
</tr>
<tr>
<td>Descartes Centre</td>
<td>€52,431</td>
</tr>
<tr>
<td>National Library of the Netherlands</td>
<td>€1,656</td>
</tr>
<tr>
<td>Staatsbibliothek zu Berlin</td>
<td>€0</td>
</tr>
<tr>
<td>Deutsche Nationalbibliothek</td>
<td>€0</td>
</tr>
<tr>
<td>University of Amsterdam</td>
<td>€57,950</td>
</tr>
<tr>
<td>Huygens ING</td>
<td>€5,519</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€119,762</strong></td>
</tr>
</tbody>
</table>

Abstract
This project aims at using and populating the basic CLARIN infrastructure to enable bilingual and bисcriptural 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)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max Planck Institute for Psycholinguistics</td>
<td>€49,671</td>
</tr>
<tr>
<td>Radboud University Nijmegen</td>
<td>€5,170</td>
</tr>
</tbody>
</table>
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)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Institute for Dutch Lexicology</td>
<td>€55,190</td>
</tr>
<tr>
<td>VU University Amsterdam (LET)</td>
<td>€51,700</td>
</tr>
<tr>
<td>VU University Amsterdam (FEW)</td>
<td>€10,340</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€117,230</strong></td>
</tr>
</tbody>
</table>

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)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delft University of Technology</td>
<td>€41,533</td>
</tr>
<tr>
<td>Erasmus University Rotterdam</td>
<td>€38,775</td>
</tr>
<tr>
<td>Netherlands Institute for Sound and Vision</td>
<td>€39,185</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€119,493</strong></td>
</tr>
</tbody>
</table>
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)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>University of Amsterdam</td>
<td>€44,402</td>
</tr>
<tr>
<td>Institute for Dutch Lexicology</td>
<td>€54,310</td>
</tr>
<tr>
<td>Huygens ING</td>
<td>€20,930</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€119,642</strong></td>
</tr>
</tbody>
</table>

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)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utrecht University</td>
<td>€56,266</td>
</tr>
<tr>
<td>Max Planck Institute for Psycholinguistics</td>
<td>€16,527</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€72,793</strong></td>
</tr>
</tbody>
</table>
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)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utrecht University</td>
<td>€15,637</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>€15,637</td>
</tr>
</tbody>
</table>

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)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leiden University</td>
<td>€67,760</td>
</tr>
<tr>
<td>Max Planck Institute for Psycholinguistics</td>
<td>€5,727</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>€73,487</td>
</tr>
</tbody>
</table>

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)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>VU University Amsterdam</td>
<td>€ t.b.a.</td>
</tr>
<tr>
<td>Huygens ING</td>
<td>€ t.b.a.</td>
</tr>
<tr>
<td>Tilburg University</td>
<td>€ t.b.a.</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€ t.b.a.</strong></td>
</tr>
</tbody>
</table>

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)
Project coordinator: Dr. Jasmijn van Gorp (Utrecht University)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utrecht University</td>
<td>€31,218</td>
</tr>
<tr>
<td>Netherlands Institute for Sound and Vision</td>
<td>€11,713</td>
</tr>
<tr>
<td>University of Amsterdam</td>
<td>€75,977</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€75,977</strong></td>
</tr>
</tbody>
</table>
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
Project coordinator: Dr. André Bouwman (Leiden University)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utrecht University</td>
<td>€5,203</td>
</tr>
<tr>
<td>Leiden University</td>
<td>€14,048</td>
</tr>
<tr>
<td>Huygens ING</td>
<td>€48,884</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€68,135</strong></td>
</tr>
</tbody>
</table>

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)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radboud University Nijmegen</td>
<td>€41,519</td>
</tr>
<tr>
<td>Max Planck Institute for Psycholinguistics</td>
<td>€8,279</td>
</tr>
<tr>
<td>University of Amsterdam</td>
<td>€5,170</td>
</tr>
<tr>
<td>Utrecht University</td>
<td>€9,306</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€64,274</strong></td>
</tr>
</tbody>
</table>
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:

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Huygens ING</td>
<td>€95,490</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€95,490</strong></td>
</tr>
</tbody>
</table>

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 Knjiženstvo; 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
Project coordinator: Prof. dr. Wido van Peursen (VU University Amsterdam)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>VU University Amsterdam</td>
<td>€60,910</td>
</tr>
<tr>
<td>Data Archiving and Networking Services</td>
<td>€40,274</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€101,184</strong></td>
</tr>
</tbody>
</table>

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)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>VU University Amsterdam</td>
<td>€42,925</td>
</tr>
<tr>
<td>Huygens ING</td>
<td>€55,512</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€98,436</strong></td>
</tr>
</tbody>
</table>

**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)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radboud University Nijmegen</td>
<td>€53,794</td>
</tr>
<tr>
<td>Max Planck Institute for Psycholinguistics</td>
<td>€11,038</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€64,832</strong></td>
</tr>
</tbody>
</table>

**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

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radboud University Nijmegen</td>
<td>€10,340</td>
</tr>
<tr>
<td>Max Planck Institute for Psycholinguistics</td>
<td>€52,431</td>
</tr>
<tr>
<td>University of Amsterdam</td>
<td>€5,170</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€67,941</strong></td>
</tr>
</tbody>
</table>

**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)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radboud University Nijmegen</td>
<td>€63,717</td>
</tr>
<tr>
<td>Huygens ING</td>
<td>€50,500</td>
</tr>
</tbody>
</table>
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)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tilburg University TiCC</td>
<td>€98,582</td>
</tr>
<tr>
<td>Tilburg University TSH</td>
<td>€5,203</td>
</tr>
<tr>
<td>Institute for Dutch Lexicology</td>
<td>€41,107</td>
</tr>
<tr>
<td>Huygens ING</td>
<td>€2,601.50</td>
</tr>
<tr>
<td>Meertens Institute</td>
<td>€2,601.50</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€150,095</strong></td>
</tr>
</tbody>
</table>

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)
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 CMD12RDF
Converting CMDI to RDF
Project Coordinator: dr. Menzo Windhouwer (DANS)
2014-05-01 – 2014-12-01

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
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

<table>
<thead>
<tr>
<th>Organization</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>INL</td>
<td>€ 11,246</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Organization</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>INL</td>
<td>€ 10,039</td>
</tr>
</tbody>
</table>

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
Calls - Projects

Project Coordinator: prof.dr. Piek Vossen (Free University Amsterdam)

2014-09-01 – 2015-03-31

<table>
<thead>
<tr>
<th>Organisation</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>INL</td>
<td>€ 29,659.50</td>
</tr>
<tr>
<td>Free University</td>
<td>€ 5,094.00</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€ 34,753.50</strong></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Organisation</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>INL</td>
<td>€ 16,371</td>
</tr>
<tr>
<td>Meertens Institute</td>
<td>€ 2,547</td>
</tr>
<tr>
<td>Utrecht University</td>
<td>€ 16,709</td>
</tr>
<tr>
<td>RU Groningen</td>
<td>€ 16,709</td>
</tr>
<tr>
<td>Taalmonsters</td>
<td>€ 5,457</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€ 57,792</strong></td>
</tr>
</tbody>
</table>

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)
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

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
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

<table>
<thead>
<tr>
<th>Organisation</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>RUN</td>
<td>€ 39,865</td>
</tr>
<tr>
<td>Total</td>
<td>€ 39,865</td>
</tr>
</tbody>
</table>

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)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Huygens ING</td>
<td>€10,000</td>
</tr>
<tr>
<td>Max Planck Institute for Psycholinguistics</td>
<td>€16,000</td>
</tr>
<tr>
<td>Meertens Institute</td>
<td>€74,000</td>
</tr>
<tr>
<td>University of Groningen</td>
<td>€70,000</td>
</tr>
<tr>
<td>Utrecht University</td>
<td>€30,000</td>
</tr>
<tr>
<td>Tilburg University</td>
<td>€75,000</td>
</tr>
<tr>
<td>Institute for Dutch Lexicology</td>
<td>€30,000</td>
</tr>
<tr>
<td>Radboud University Nijmegen</td>
<td>€25,000</td>
</tr>
<tr>
<td>University of Twente</td>
<td>€60,000</td>
</tr>
<tr>
<td>Aletta Institute for women’s history</td>
<td>€5,000</td>
</tr>
<tr>
<td>Catholic Documentation Centre</td>
<td>€5,000</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€400,000</strong></td>
</tr>
</tbody>
</table>

**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)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meertens Institute</td>
<td>€30,000</td>
</tr>
<tr>
<td>Max Planck Institute for Psycholinguistics</td>
<td>€15,000</td>
</tr>
<tr>
<td>Institute for Dutch Lexicology</td>
<td>€30,000</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€75,000</strong></td>
</tr>
</tbody>
</table>

**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
Infrastructure Implementation Projects
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
*Infrastructuurimplementatieplan*
2010-05-01 – 2013-05-01
Project coordinator: Ir. Daan Broeder (Max Planck Institute)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meertens Institute</td>
<td>€192,000</td>
</tr>
<tr>
<td>Max Planck Institute for Psycholinguistics</td>
<td>€384,000</td>
</tr>
<tr>
<td>Institute for Dutch Lexicology</td>
<td>€192,000</td>
</tr>
<tr>
<td>Data Archiving and Networking Services</td>
<td>€30,000</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€798,000</strong></td>
</tr>
</tbody>
</table>

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)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Huygens ING</td>
<td>€52,974</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€52,974</strong></td>
</tr>
</tbody>
</table>

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.
Infrastructure Implementation Projects

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)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meertens Institute</td>
<td>€180,000</td>
</tr>
<tr>
<td>Max Planck Institute for Psycholinguistics</td>
<td>€216,000</td>
</tr>
<tr>
<td>Institute for Dutch Lexicology</td>
<td>€126,000</td>
</tr>
<tr>
<td>Data Archiving and Networking Services</td>
<td>€54,000</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€576,000</strong></td>
</tr>
</tbody>
</table>

**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)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utrecht University (UiL-OTS)</td>
<td>€18,000</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€18,000</strong></td>
</tr>
</tbody>
</table>

**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)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radboud University Nijmegen</td>
<td>€18,400</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€18,400</strong></td>
</tr>
</tbody>
</table>
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)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utrecht University (UiL-OTS)</td>
<td>€48,000</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€48,000</strong></td>
</tr>
</tbody>
</table>

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)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meertens Institute</td>
<td>€30,000</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€30,000</strong></td>
</tr>
</tbody>
</table>

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)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meertens Institute</td>
<td>€17,000</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€17,000</strong></td>
</tr>
</tbody>
</table>

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)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meertens Institute</td>
<td>€35,635</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€35,635</strong></td>
</tr>
</tbody>
</table>

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)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radboud University Nijmegen</td>
<td>€174,807</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€174,807</strong></td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meertens Institute</td>
<td>€150,000</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>€150,000</strong></td>
</tr>
</tbody>
</table>
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)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utrecht University</td>
<td>€30,000</td>
</tr>
</tbody>
</table>

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 background 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**
Events

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 CLARIN and GrETEL: Lecture for the UU Digital Groningen
14 Oct  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
<table>
<thead>
<tr>
<th>Date</th>
<th>Description</th>
<th>Location</th>
<th>Form of Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>25-26 Jan</td>
<td>CLARIN Demonstrator Workshop</td>
<td>Amsterdam, Netherlands</td>
<td>Infrastructure and Microvariation Financial support for invited speakers</td>
</tr>
<tr>
<td>5 Feb</td>
<td>CLIN20</td>
<td>Utrecht, Netherlands</td>
<td>10k euro + plenary presentation on CLARIN-NL</td>
</tr>
<tr>
<td>4-5 Aug</td>
<td><a href="#">Relish Workshop</a></td>
<td>Nijmegen, Netherlands</td>
<td>Funding travel and subsistence of some participants</td>
</tr>
</tbody>
</table>

**2011**

<table>
<thead>
<tr>
<th>Date</th>
<th>Description</th>
<th>Location</th>
<th>Form of Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>17 Jan</td>
<td>CMDI developers meeting</td>
<td>Nijmegen, Netherlands</td>
<td>Subsidy of €228.50</td>
</tr>
<tr>
<td>9 May</td>
<td>Federated Search &amp; EDC Workshop</td>
<td>Nijmegen, Netherlands</td>
<td>Subsidy of €3,000</td>
</tr>
<tr>
<td>19 May</td>
<td>Kick-off meeting VU-DNC</td>
<td>Amsterdam, Netherlands</td>
<td>Subsidy of €116.15</td>
</tr>
<tr>
<td>23 Jun</td>
<td>Workshop E-History</td>
<td>Amsterdam, Netherlands</td>
<td>Subsidy of €423.85</td>
</tr>
</tbody>
</table>

**2012**

<table>
<thead>
<tr>
<th>Date</th>
<th>Description</th>
<th>Location</th>
<th>Form of Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>29 oct</td>
<td>CLARIN’s Turn to the literary text</td>
<td>Den Haag, Netherlands</td>
<td>Subsidy of €2,761</td>
</tr>
<tr>
<td>31 Jan</td>
<td>Lustrumcongres Alfainformatica en oratie Gertjan van Noord</td>
<td>Groningen, Netherlands</td>
<td>Subsidy of €1,500</td>
</tr>
<tr>
<td>20 Jan</td>
<td>CLIN</td>
<td>Tilburg, Netherlands</td>
<td>Subsidy of €2500</td>
</tr>
<tr>
<td>30 Nov</td>
<td>A Shared LMF Core for Sign Language Lexicons</td>
<td>Ravenstein, Netherlands</td>
<td>Subsidy of €6840</td>
</tr>
</tbody>
</table>

**2013**

<table>
<thead>
<tr>
<th>Date</th>
<th>Description</th>
<th>Location</th>
<th>Form of Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>12 Apr</td>
<td>CLARIN Literary Workshop</td>
<td>Amsterdam</td>
<td>financial</td>
</tr>
<tr>
<td>17 Jan</td>
<td>CLIN</td>
<td>Hengelo</td>
<td>financial</td>
</tr>
</tbody>
</table>

**2014**

<table>
<thead>
<tr>
<th>Date</th>
<th>Description</th>
<th>Location</th>
<th>Form of Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>14 Jan</td>
<td>THATCamp</td>
<td>Den Haag</td>
<td>financial</td>
</tr>
<tr>
<td>17 Jan</td>
<td>CLIN</td>
<td>Leiden</td>
<td>financial</td>
</tr>
<tr>
<td>Date</td>
<td>Description</td>
<td>Location</td>
<td>Form of Support</td>
</tr>
<tr>
<td>------------</td>
<td>-------------------------------------------</td>
<td>-----------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>5 Feb</td>
<td>CLIN25</td>
<td>Antwerpen</td>
<td>financial</td>
</tr>
<tr>
<td>13 Apr</td>
<td>Workshop STCN</td>
<td>Den Haag</td>
<td>financial</td>
</tr>
<tr>
<td>25 Sep</td>
<td>DRONGO Festival</td>
<td>Utrecht</td>
<td>financial (via UU)</td>
</tr>
<tr>
<td>16 Nov</td>
<td>Workshop Morphosyntactische verrijking van 17e eeuwse teksten</td>
<td>Utrecht</td>
<td>Website, announcements, active participation</td>
</tr>
<tr>
<td>18 Dec</td>
<td>CLIN26</td>
<td>Amsterdam</td>
<td>financial</td>
</tr>
</tbody>
</table>
## Events

### 2009

<table>
<thead>
<tr>
<th>Date</th>
<th>Event</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>30 Jan</td>
<td>DANS Digital Preservation Workshop</td>
<td>The Hague, Netherlands</td>
</tr>
<tr>
<td>4-5 Apr</td>
<td>CLARIN/FLaReNet Usage Scenario Meeting</td>
<td>Athens, Greece</td>
</tr>
<tr>
<td>11-12 May</td>
<td>CLARIN Consortium Meeting</td>
<td>Barcelona, Spain</td>
</tr>
<tr>
<td>4 Jun</td>
<td>KNAW Symposium Computational Humanities: Bridging the gap between the</td>
<td>Amsterdam, Netherlands</td>
</tr>
<tr>
<td></td>
<td>Humanities and the Computational Sciences</td>
<td></td>
</tr>
<tr>
<td>15-16 Jun</td>
<td>Workshop Small Tools for Cross-Linguistic Research</td>
<td>Utrecht, Netherlands</td>
</tr>
<tr>
<td>30 Sep</td>
<td>CLARIN - FLaReNet workshop Standards in the LRT field</td>
<td>Helsinki, Finland</td>
</tr>
<tr>
<td>1-2 Oct</td>
<td>NEERI09 – interaction between research infrastructures on common</td>
<td>Helsinki, Finland</td>
</tr>
<tr>
<td></td>
<td>topics</td>
<td></td>
</tr>
<tr>
<td>21-23 Oct</td>
<td>Working together to strengthen Research in Europe</td>
<td>Brussels, Belgium</td>
</tr>
<tr>
<td>2 Nov</td>
<td>SILT-FLaReNet Workshop Toward an Operational Definition of</td>
<td>Waltham, US</td>
</tr>
<tr>
<td></td>
<td>Interoperability for Language Technology</td>
<td></td>
</tr>
<tr>
<td>5 Nov</td>
<td>CLARIN centres workshop</td>
<td>Prague, Czech Republic</td>
</tr>
<tr>
<td>2 Dec</td>
<td>Symposium - Door Data Gedreven, nieuwe inzichten in de wetenschap</td>
<td>The Hague, Netherlands</td>
</tr>
<tr>
<td></td>
<td>door innovatief gegevensgebruik</td>
<td></td>
</tr>
</tbody>
</table>

### 2010

<table>
<thead>
<tr>
<th>Date</th>
<th>Event</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>25-26 Jan</td>
<td>Workshop on CLARIN EDC (European Demonstrator Case)</td>
<td>Amsterdam, Netherlands</td>
</tr>
<tr>
<td>5 Feb</td>
<td>CLIN 20</td>
<td>Utrecht, Netherlands</td>
</tr>
<tr>
<td>16-23 May</td>
<td>LREC 20</td>
<td>Valletta, Malta</td>
</tr>
<tr>
<td>11 Jun</td>
<td>Symposium “Toekomst voor Digitale Geesteswetenschappen”</td>
<td>Utrecht, Netherlands</td>
</tr>
<tr>
<td>17-18 Jun</td>
<td>NO-CLARIN</td>
<td>Oslo, Norway</td>
</tr>
<tr>
<td>28 Jun</td>
<td>Ed Hovy Symposium</td>
<td>Utrecht, Netherlands</td>
</tr>
<tr>
<td>1 Jul</td>
<td>DigiHist: Studiedag Digitalisering Historische Handschriften</td>
<td>Leuven, Belgium</td>
</tr>
<tr>
<td>7-8 Jul</td>
<td>Digital Humanities 2010</td>
<td>London, UK</td>
</tr>
<tr>
<td>13 Sep</td>
<td>BIG Grid User event</td>
<td>Amsterdam, Netherlands</td>
</tr>
<tr>
<td>18-22 Oct</td>
<td>SDH2010 + NEERI</td>
<td>Vienna, Austria</td>
</tr>
<tr>
<td>2 Dec</td>
<td>Conferentie Grootschalige onderzoeksfaciliteiten</td>
<td>Leiden, Netherlands</td>
</tr>
</tbody>
</table>

### 2011

<table>
<thead>
<tr>
<th>Date</th>
<th>Event</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>17-18 Jan</td>
<td>CMDI workshop</td>
<td>Nijmegen, Netherlands</td>
</tr>
<tr>
<td>21 Jan</td>
<td>Oratie Paul Wouters</td>
<td>Leiden, Netherlands</td>
</tr>
<tr>
<td>11 Feb</td>
<td>CLIN</td>
<td>Gent, Belgium</td>
</tr>
<tr>
<td>20 Feb</td>
<td>DARIAH workshop</td>
<td>London, UK</td>
</tr>
<tr>
<td>08 March</td>
<td>Virtual Organization Workshop</td>
<td>Mannheim, Germany</td>
</tr>
<tr>
<td>28 March</td>
<td>EHumanities Symposium</td>
<td>Wassenaar Netherlands</td>
</tr>
<tr>
<td>30 March</td>
<td>Diachroon Corpus</td>
<td>The Hague, Netherlands</td>
</tr>
<tr>
<td>Date</td>
<td>Event</td>
<td>Location</td>
</tr>
<tr>
<td>------------</td>
<td>----------------------------------------------------------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>01 April</td>
<td>TTNWW workshop</td>
<td>Tilburg, Netherlands</td>
</tr>
<tr>
<td>07 April</td>
<td>SURFnet event about data management</td>
<td>Utrecht, Netherlands</td>
</tr>
<tr>
<td>12-13 April</td>
<td>EPIC User Forum</td>
<td>Amsterdam, Netherlands</td>
</tr>
<tr>
<td>09 May</td>
<td>EU workshop Federated Search</td>
<td>Nijmegen, Netherlands</td>
</tr>
<tr>
<td>18 May</td>
<td>SURFshare Open OnderzoeksData Dag</td>
<td>Amsterdam, Netherlands</td>
</tr>
<tr>
<td>26-27 May</td>
<td>FlareNet workshop</td>
<td>Venice, Italy</td>
</tr>
<tr>
<td>08 Jun</td>
<td>Dealing with data</td>
<td>The Hague, Netherlands</td>
</tr>
<tr>
<td>9-10 Jun</td>
<td>Federated Identity Management Workshop</td>
<td>Genève, Swiss</td>
</tr>
<tr>
<td>2-3 Sept</td>
<td>Peter Wittenburg International workshop</td>
<td>Kleva, Germany</td>
</tr>
<tr>
<td>7-9 Sept</td>
<td>Panel Digging into Data</td>
<td>London, UK</td>
</tr>
<tr>
<td>12-13 Sept</td>
<td>IRISC workshop</td>
<td>Helsinki, Finland</td>
</tr>
<tr>
<td>29 Sept</td>
<td>Alfalab Symposium</td>
<td>Amsterdam, Netherlands</td>
</tr>
<tr>
<td>10 Oct</td>
<td>Creatieve industrie en wetenschappelijk onderzoek</td>
<td>Amsterdam, Netherlands</td>
</tr>
<tr>
<td>1 Nov</td>
<td>CATCH meeting at HITIME</td>
<td>Amsterdam, Netherlands</td>
</tr>
<tr>
<td>5-19 Nov</td>
<td>IJCNLP Chiang Mai</td>
<td>Chiang Mai, Thailand</td>
</tr>
<tr>
<td>9 Dec</td>
<td>SARA 40</td>
<td>Amsterdam, Netherlands</td>
</tr>
<tr>
<td>2012</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 Dec</td>
<td>Oratie Rens Bod</td>
<td>Amsterdam, Netherlands</td>
</tr>
<tr>
<td>11 Dec</td>
<td>Humanities Lecture Frans Wieringa</td>
<td>Utrecht, Netherlands</td>
</tr>
<tr>
<td>3-4 Dec</td>
<td>CATCH+ / DEN Digital Heritage Conference</td>
<td>Rotterdam, Netherlands</td>
</tr>
<tr>
<td>29 Oct</td>
<td>CLARIN's Turn to the literary text</td>
<td>The Hague, Netherlands</td>
</tr>
<tr>
<td>24-28 Oct</td>
<td>CLARIN Integration Conference</td>
<td>Sofia, Bulgaria</td>
</tr>
<tr>
<td>23 Oct</td>
<td>Kick-off CLICK-NL</td>
<td>Eindhoven, Netherlands</td>
</tr>
<tr>
<td>23 Oct</td>
<td>Digital Humanities Lecture Linguistic Research and The CLARIN Infrastructure</td>
<td>Utrecht, Netherlands</td>
</tr>
<tr>
<td>22-23 Oct</td>
<td>International Workshop on Open Access</td>
<td>Rotterdam, Netherlands</td>
</tr>
<tr>
<td>19 Oct</td>
<td>Kick-off Digital Humanities Amsterdam, KN AW</td>
<td>Amsterdam, Netherlands</td>
</tr>
<tr>
<td>18 Oct</td>
<td>PhD Defense Folkert de Vriend</td>
<td>Nijmegen, Netherlands</td>
</tr>
<tr>
<td>8 Oct</td>
<td>Presentation Project ‘Gekapte Brieven’</td>
<td>Rotterdam, Netherlands</td>
</tr>
<tr>
<td>3-4 Oct</td>
<td>SURF Relatiedagen</td>
<td>Noordwijkerhout, Netherlands</td>
</tr>
<tr>
<td>7 Sept</td>
<td>Afscheidscollege Geert Booij</td>
<td>Leiden, Netherlands</td>
</tr>
<tr>
<td>7 Sept</td>
<td>Symposium Geert Booij</td>
<td>Leiden, Netherlands</td>
</tr>
<tr>
<td>3 Sept</td>
<td>Opening Academisch jaar</td>
<td>Utrecht, Netherlands</td>
</tr>
<tr>
<td>3 Sept</td>
<td>Opening letterenbibliotheek</td>
<td>Utrecht, Netherlands</td>
</tr>
<tr>
<td>31 August</td>
<td>Presentation Network Cultural Heritage / CLICK-NL</td>
<td>Hilversum, Netherlands</td>
</tr>
<tr>
<td>30 August</td>
<td>CLARIN informatiesessie Call 4</td>
<td>Amsterdam, Netherlands</td>
</tr>
<tr>
<td>20-21 Jun</td>
<td>META-FORUM 2012</td>
<td>Brussels, Belgium</td>
</tr>
<tr>
<td>4 Jun</td>
<td>Workshop Open Access, Museon</td>
<td>The Hague, Netherlands</td>
</tr>
<tr>
<td>20-27 May</td>
<td>LREC</td>
<td>Istanbul, Turkey</td>
</tr>
<tr>
<td>23 March</td>
<td>Meeting with CLARIN-CZ/LINDAT (Jan Hajić, Eva)</td>
<td>Prague, Czech Republic</td>
</tr>
<tr>
<td>Date</td>
<td>Event</td>
<td>Location</td>
</tr>
<tr>
<td>--------</td>
<td>----------------------------------------------------------------------</td>
<td>------------------------------</td>
</tr>
<tr>
<td>22 Mar</td>
<td>Lecture in Prague</td>
<td>Prague, Czech Republic</td>
</tr>
<tr>
<td>6 Mar</td>
<td>CLARIN Kick-off</td>
<td>Hilversum, Netherlands</td>
</tr>
<tr>
<td>31 Jan</td>
<td>Lustrumcongres Alfa-Informatika en oratie Gertjan van Noord en Johan Bos</td>
<td>Groningen, Netherlands</td>
</tr>
<tr>
<td>20 Jan</td>
<td>CLIN</td>
<td>Tilburg, Netherlands</td>
</tr>
<tr>
<td></td>
<td><strong>2013</strong></td>
<td></td>
</tr>
<tr>
<td>13 Jan</td>
<td>8th International Digital Curation Conference 2013</td>
<td>Amsterdam, Netherlands</td>
</tr>
<tr>
<td>17-18 Jan</td>
<td>CLIN</td>
<td>Enschede, Netherlands</td>
</tr>
<tr>
<td>23 Jan</td>
<td>The PoliMedia symposium: linking political debates and media</td>
<td>Amsterdam, Netherlands</td>
</tr>
<tr>
<td>21 Feb</td>
<td>CLARIN Kick-off</td>
<td>Utrecht, Netherlands</td>
</tr>
<tr>
<td>28 Feb</td>
<td>SURF Research and Innovation Event</td>
<td>The Hague, Netherlands</td>
</tr>
<tr>
<td>6 Mar</td>
<td>Route 2020: the Future for Large-Scale Research Infrastructures</td>
<td>Amsterdam, Netherlands</td>
</tr>
<tr>
<td>18 Mar</td>
<td>Research Data Alliance Launch and First Plenary</td>
<td>Götenborg, Sweden</td>
</tr>
<tr>
<td>21 Mar</td>
<td>NIAS DH Workshop</td>
<td>Wassenaar, NL</td>
</tr>
<tr>
<td>10 Apr</td>
<td>eHumanities Address Software and Tools</td>
<td>Amsterdam, NL</td>
</tr>
<tr>
<td>12 Apr</td>
<td>CLARIN Literary Workshop</td>
<td>Amsterdam, NL</td>
</tr>
<tr>
<td>24 Apr</td>
<td>CLARIN Federated Content Search workshop</td>
<td>Copenhagen, DK</td>
</tr>
<tr>
<td>25 Apr</td>
<td><a href="#">CLARIN Centre Assessment Workshop</a></td>
<td>Copenhagen, DK</td>
</tr>
<tr>
<td>17 May</td>
<td>OpenSKOS and controlled vocabularies meeting</td>
<td>Utrecht, NL</td>
</tr>
<tr>
<td>18 Jul</td>
<td>Language Diversity Congress</td>
<td>Groningen, NL</td>
</tr>
<tr>
<td>10 Sep</td>
<td><a href="#">Open Research Data on the Horizon ... and beyond - RDA Europe Workshop prior to the RDA Plenary in Washington</a></td>
<td>Garching, US</td>
</tr>
<tr>
<td>20 Sep</td>
<td>Digital Humanities Summerschool</td>
<td>Leuven, BE</td>
</tr>
<tr>
<td>30 Sep</td>
<td><a href="#">Workshop: CLARIN, Standards and the Text Encoding Initiative</a></td>
<td>Rome, IT</td>
</tr>
<tr>
<td>17-18 Oct</td>
<td><a href="#">DARIAH/DASISH Access and Authentication Infrastructures Workshop</a></td>
<td>Köln, DE</td>
</tr>
<tr>
<td>20-22 Oct</td>
<td>CLARIN ANNUAL Meeting</td>
<td>Prague, CZ</td>
</tr>
<tr>
<td>28-30 Oct</td>
<td><a href="#">2nd EUDAT Conference (October 28-30, Rome)</a></td>
<td>Rome, IT</td>
</tr>
<tr>
<td>26-27 Nov</td>
<td><a href="#">Annual Meeting General Assembly</a></td>
<td>Wroclaw, PO</td>
</tr>
</tbody>
</table>
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 Meeting35 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.
### Events

<table>
<thead>
<tr>
<th>Date</th>
<th>Event</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 Feb</td>
<td>CLIN25</td>
<td>Antwerpen, NL</td>
</tr>
<tr>
<td>7 Feb</td>
<td>TIN-dag</td>
<td>Utrecht, NL</td>
</tr>
<tr>
<td>24 Mar</td>
<td>Digitale Historische Kranten als Big Data</td>
<td>Den Haag, NL</td>
</tr>
<tr>
<td>26 Mar</td>
<td>CLARIN-TCC Creative Camp</td>
<td>Amsterdam, NL</td>
</tr>
<tr>
<td>1 Apr</td>
<td>Symposium <em>Luid zingend op een ijsschots de zomer tegemoet: de toekomst van de geesteswetenschappen</em></td>
<td>Amsterdam, NL</td>
</tr>
<tr>
<td>9 Apr</td>
<td>Invited Lecture at the Austrian Centre for Digital Humanities</td>
<td>Vienna, AT</td>
</tr>
<tr>
<td>28 Apr</td>
<td>Digital Humanities Spring School (Keynote)</td>
<td>Leuven, BE</td>
</tr>
<tr>
<td>20-24 Sep</td>
<td>PARSEME Workshop</td>
<td>Iași, RO</td>
</tr>
<tr>
<td>25-26 Sep</td>
<td>DRONGO-festival</td>
<td>Utrecht, NL</td>
</tr>
<tr>
<td>15-17 Oct</td>
<td>CLARIN Conference</td>
<td>Wrocław, PO</td>
</tr>
<tr>
<td>30 Oct</td>
<td>CLARIN-TCC Creative Camp</td>
<td>Den Haag, NL</td>
</tr>
<tr>
<td>16 Nov</td>
<td>Workshop Morphosyntactische verrijking van 17e eeuwse teksten</td>
<td>Utrecht, NL</td>
</tr>
<tr>
<td>19-20 Nov</td>
<td>CLARIN General Assembly</td>
<td>Copenhagen, DK</td>
</tr>
<tr>
<td>18 Dec</td>
<td>CLIN26</td>
<td>Amsterdam, NL</td>
</tr>
</tbody>
</table>

### Meetings

<table>
<thead>
<tr>
<th>Date</th>
<th>Event</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>19 Jan</td>
<td>CLARIN-NL preparatory committee</td>
<td>Utrecht, Netherlands</td>
</tr>
<tr>
<td>21 Jan</td>
<td>NGT</td>
<td>Utrecht, Netherlands</td>
</tr>
<tr>
<td>28 Jan</td>
<td>NWO</td>
<td>The Hague, Netherlands</td>
</tr>
<tr>
<td>Mar</td>
<td>Institut für Deutsche Sprache</td>
<td>Mannheim, Germany</td>
</tr>
<tr>
<td>11 Mar</td>
<td>MPI</td>
<td>Nijmegen, Netherlands</td>
</tr>
<tr>
<td>15 Apr</td>
<td>DANS</td>
<td>Utrecht, Netherlands</td>
</tr>
<tr>
<td>19 May</td>
<td>Alfalab</td>
<td>Utrecht, Netherlands</td>
</tr>
<tr>
<td>17 Sep</td>
<td>NWO</td>
<td>The Hague, Netherlands</td>
</tr>
<tr>
<td>6 Oct</td>
<td>Geleerdenbrievenproject (CKCC)</td>
<td>Utrecht, Netherlands</td>
</tr>
<tr>
<td>9 Oct</td>
<td>EWI</td>
<td>Brussels, Belgium</td>
</tr>
<tr>
<td>9 Dec</td>
<td>Ministerie van OCW on ERIC</td>
<td>The Hague, Netherlands</td>
</tr>
</tbody>
</table>
Events

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 Kruijf      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
08 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 interview            Wassenaar, Netherlands
06 Apr   TAB Taalportaal                  Leiden, Netherlands
15 Apr   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 Bontouridis  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.
The movies on DSS and OpenSONAR are planned to be available as of Mid 2016.

See [http://www.clarin.nl/node/403](http://www.clarin.nl/node/403)

Several subprojects have made movies of their own:

<table>
<thead>
<tr>
<th>Project</th>
<th>Link to Movie / Screen capture</th>
</tr>
</thead>
<tbody>
<tr>
<td>GabMap</td>
<td><a href="http://www.youtube.com/watch?v=anElatt5ffM&amp;feature=plcp">http://www.youtube.com/watch?v=anElatt5ffM&amp;feature=plcp</a></td>
</tr>
<tr>
<td>Interviews / Veteran Tapes</td>
<td><a href="http://www.youtube.com/watch?v=VpCXh_NVVsO">http://www.youtube.com/watch?v=VpCXh_NVVsO</a></td>
</tr>
<tr>
<td>WFT-GTB</td>
<td><a href="http://www.youtube.com/watch?v=fevAPknCaDo&amp;feature=player_embedded">http://www.youtube.com/watch?v=fevAPknCaDo&amp;feature=player_embedded</a></td>
</tr>
<tr>
<td>RemBench</td>
<td><a href="http://rembench.huygens.knaw.nl/help">http://rembench.huygens.knaw.nl/help</a></td>
</tr>
</tbody>
</table>

**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 ]“Woordenboek der Friese taal online”, article at Kennislink, July 16, 2010.


[WFT-GTB ]“40.000 kaartjes voor het woordje ‘de’ staan online”, article about WFT-GTB in Friesch Dagblad, July 6, 2010.

'Naar een betere toegang tot tekst- en spraaktools'. Artikel door Marc Kemps-Snijders en Ineke Schuurman in E-Data & Research 5.1, 2010. [URL]

**2011**

[GABMAP] "Het beeld en de streektaal", artikel over het onderzoek project GABMAP door Erica Renckens in E-Data & Research 6(2), p.3.

[CLARIN-NL] "CLARIN-NL - Common Language Resources and Technology Infrastructure in Nederland", focus artikel door Inge Angevaare in E-Data & Research, 6(1), juli 2011, p.6.

[TICCLops] "Oude teksten bruikbaar voor nieuw onderzoek", artikel over de subprojecten TICCLops en Adelheid door Erica Renckens in E-Data & Research 6(1), juni 2011, p.3.

[WIP] "Dat doet denken aan de oorlog", artikel over het onderzoek project WIP door Erica Renckens in E-Data & Research 6(2), oktober 2011, p.1 [PDF].


[CLARIN-NL] "CLARIN-NL - Common Language Resources and Technology Infrastructure in Nederland", focus artikel door Inge Angevaare in E-Data & Research, 6(1), juni 2011, p.6. [URL] [PDF].


[CLARIN-NL] "Negen nieuwe CLARIN projecten van start", artikel door Erica Renckens in E-Data & Research 5(4), maart 2011, p.7 [URL] [PDF].

[CLARIN-NL] "CLARIN-NL deelnemer in eData & Research", artikel door Erica Renckens in E-Data & Research 5(4), maart 2011, p.1 [URL] [PDF].

[WIP] "Waarom politici graag over de oorlog praten", artikel over het onderzoek project WIP door Bart Funnekotter in NRC, februari 25, 2011 [URL] [PDF] [WIP website].

**2012**

[CLARIN-NL] "De vraag van morgen moet je vandaag kunnen beantwoorden", interview met Peter Wittenburg door Heidi Berkhout en Erica Renckens in E-Data & Research 7(1), oktober 2012, p.5.


[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] "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


[CLARIN-NL] ‘We moeten nog leren werken met Big Data’. Article by Inge Angevaere in eData & Research, 8.3, p. 5-5 Jun 2014


[CLARIN-NL] ‘Eerste Nederlandse THATCamp blijkt leerzame unconference. Article by Peter Boot in eData & Research, 8.2, p. 6-6 Feb 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

[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 Nicole 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].

[Polimedi] "Eerste prijs voor online tool met linked data", article by Erica Renckens in E-Data & Research 8(2), February 2014, p.8 [URL] [PDF].
**Media**

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


2011


2012


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]


Rooij, O. de, Kenter, T., & de Rijke, M. xTAS and ThemeStreams. *proceedings dir 2013*, 2.


**2014**


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]


http://dspace.library.uu.nl/bitstream/handle/1874/303426/histoinformatics2013_wahsp.pdf?sequence=1


https://staff.science.uva.nl/m.derijke/Publications/ecir2014-fp-verticals.pdf


https://staff.science.uva.nl/m.derijke/Publications/sigir2014-fp-fusion.pdf


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]


Reynaert, M. Synergy of Nederlab and @PhilosTEI: Diachronic and Multilingual Text-induced Corpus Clean-up. Proceedings of LREC 2014, Reykjavik, Iceland. [URL]


Sanders, E. et al. The Dutch LESLLA Corpus, Proceedings of LREC 2014, Reykjavik, Iceland. [URL]


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). 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).


2015


Odijk, J. (2015) 'Linguistic Research with PaQU' *Computational Linguistics in The Netherlands journal* 5, p. 3-14 [pdf] [url]


Odijk, J. (2015) 'Linguistic Research with PaQU' *Computational Linguistics in The Netherlands journal* 5, p. 3-14 [pdf] [url]


2016


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.


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.


**Presentations**

**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) "CLARIN: wat is dat?", presentation held at the Dag van de Fonetiek, Utrecht, December 18, 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, November 1, 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 CLARIN-NL 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


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


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


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


Heuvel, H. van den (2010), “'INTER-VIEWS'. Curation of Interview Data". Presentation held at the first CLARIN-NL meeting, Feb 19, 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.


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/]


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: What is going on?", presentation held at the CLARIN-NL Second Open Call Info Session, Amsterdam, August 26, 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.


Vriend, F. de (2010), "MIMORE", presentation held at the CLARIN-NL Second Open Call Info Session, Amsterdam, August 26, 2010.


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.


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.


Fikkert, P. (2011), "Integrating and publishing resources on sign language acquisition (IPROSLA), presentation held at the CLARIN-NL Kick-off meeting Call 2, Utrecht, February 9, 2011.


2012


2013

Barbiers, S. (2013), MIMORE, presentation at the CLARIN Annual Conference, Prague, Oct 23, 2013, [url]


Bloothooft, G. (2013), MIGMAP, presentation at the CLARIN Annual Conference, Prague, Oct 22, 2013, [url]


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]


Broeder, Daan (2013) 'CoILTime: Collaboration on Time-Based Resources', presentation held at the CLARIN-NL Call 4 Kick-off Meeting, Utrecht, February 21, 2013


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


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]


Kemman, M. (2013), Building the PoliMedia search system; data- and user-driven. eHumanities Group. Amsterdam, 18 april 2013 slideshare


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


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. [ptx] [URL]


Odijk, J. (2013) 'The CLARIN Infrastructure (NL Part)' presentation held at the IAP Demonstrator Event, Utrecht September 4, 2013 [pptx]


Oostdijk, N. (2013), 'Data Curation Service’, presentation held at the Soeterbeeck eHumanities Workshop, June 13 and 14, 2013, Soeterbeeck, Ravenstein, the Netherlands.


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


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]


Westerhout, E. & J. Odijk (2013), 'Metadata for tools: creating a CMDI profile for tools', presentation held at CLIN 2013, Enschede, Jan 18, 2013. [pdf]


2014


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


Media


Hoogervorst, Tom G. ‘Heritage through digitisation’, Talk presented at the ASEASUK 14th update, 12-14 September 2014, Brighton, United Kingdom [pdf]


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]

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


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


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]


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


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


2015


Niestadt, Jan (2015), BlackLab: advantages and disadvantages of using Lucene for a linguistic corpus search system. Presentation at CLIN26, Amsterdam, 18 December 2015.


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), '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), 'Using PaQu for language acquisition research', presentation at the CLARIN 2015 Conference, Wroclaw, Poland, 16 October 2015 [pptx] [pdf]

Odijk, J. (2015), 'Using PaQu for language acquisition research', demo at the CLARIN 2015 Conference, Wroclaw, 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, Wroclaw, Poland, 16 October 2015. [pdf] [data] [handout] [analysis]


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]


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]


2016


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.


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.