CLARIN-NL
Call 3

Jan Odijk
CLARIN-NL Call 3 Info-session
Utrecht, 25 Aug 2011
Overview

- Call 3 Major Characteristics
- Analysis
- Closed Call
- Open Call
- CLARIN-Centres
- Evaluation
- Practical Matters
Call 3

- Open Call and Closed Call
- Continue population with data and tools
- Continue testing CLARIN against real data and tools
- Resource Curation and Demonstrator Projects
- Total budget 1M euro
- Projects duration: max 12 months
Project Types

• **Resource Curation Projects**
  – Data or tools
  – Curation of digital language resources

• **Demonstrator Projects**
  – Web applications
    • User interface
    • Core component (possibly web-service)
  – Demonstrator

• **Mix of the two**
Analysis

• Analysis of discipline coverage (by EB)
• Input
  – Submissions previous calls
  – Results User Survey http://www.clarin.nl/node/64
  – Data Overview Data Curation Service
    • http://www.clarin.nl/node/147
• Results
  – Priorities Open Call
  – Priorities Closed Call
  – Selected research groups to be invited for the Closed Call
Closed Call

• Priorities: 5 disciplines prioritized
  – Relatively ill-covered in CLARIN-NL so-far
• Submissions
  – Only on invitation
  – 5 invited, all accepted
• No competition among projects
• Budget
  – Maximum per project 80k euro
  – Total budget 400k euro
• [http://www.clarin.nl/node/193](http://www.clarin.nl/node/193)
Closed Call

- Prioritized Disciplines
  - Theology / Religion Sciences
  - Philology / Classical Studies
  - Language documentation / Descriptive linguistics
  - Language acquisition / impairment / attrition
  - Discourse
Open Call

• Every employee of a CLARIN partner can submit a proposal
  – If your organisation is not yet a CLARIN-partner, contact the CLARIN Office
• Competition among the projects
• Max Budget per project: 120k euro
• Total Budget: 600k euro
Open Call: Priorities

1. Literary Studies
2. Media Studies and Communication Sciences, Esp. the new media (sms, chats, blogs, twitter, etc.)
3. History
4. Linguistics
   a. Syntax
   b. Phonology
   c. Semantics
   d. Descriptive linguistics / language documentation
   e. Discourse
Project Goals (1)

• Resource Curation Projects
  – Adapt existing resources to make them visible, uniquely referable and accessible via the web, CLARIN-compatible and properly documented

• Demonstrator Projects
  – Create a Web application
    • Starting from an existing tool / application
  – Demonstration scenario
Project Goals (2)

• For all projects:
  – establish requirements and desiderata for the CLARIN infrastructure
    • By applying standards and best practices
    • By making use of CLARIN architecture and agreements
    So that we understand their limitations
CLARIN centres

• No recognized CLARIN centers exist yet
• The following institutes are working towards it:
  – MPI, Nijmegen (‘typological data’)
  – INL, Leiden (‘lexical and textual data’)
  – MI, Amsterdam (‘dialect data’)
  – DANS, The Hague
  – Huygens ING, The Hague (literary data)
Evaluation Procedure

• Evaluation by IAP
• Evaluation by NAP, with IAPs input
  – Focus on national considerations
• For Open Call: Ranking by IAP, possibly adapted by NAP
• Recommendation by EB to Board
• Decision by the Board
Evaluation Procedure

• Next to approval, possible outcomes are
  – Conditionally approval, or
  – Request for submission of improved version

• Evaluation Criteria described in the Call
  – For the Open call, the highest ranked proposals will receive funding
  – For the Closed Call, the approved projects will receive funding
Practical Matters

• **Who:**
  – researchers employed by CLARIN-NL participants that have signed the CLARIN-NL consortium agreement.

• **Eligible Costs**
  – Personnel in accordance with NWO-VSNU agreement
  – Fee for travel and subsistence
    • Max 3000 Euro / FTE / year
Practical Matters

• Use Proposal Template
• Proposal language: English
• Submit Proposal via the web form
  – http://www.clarin.nl/node/135
• As a single file (.pdf or .zip)
## Time Line

<table>
<thead>
<tr>
<th>Activity</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLARIN-NL Call 3 Open</td>
<td>Sunday July 3, 2011</td>
</tr>
<tr>
<td>Call 3 Info Session</td>
<td>Thursday August 25, 2011 (today)</td>
</tr>
<tr>
<td>Deadline Proposal Submission</td>
<td><strong>Friday September 30, 2011 13:00 hours CET</strong></td>
</tr>
<tr>
<td>Assessment IAP</td>
<td>End October 2011</td>
</tr>
<tr>
<td>Assessment NAP</td>
<td>End November 2011</td>
</tr>
<tr>
<td>Decision on Funding by Board</td>
<td><strong>Mid December 2011</strong></td>
</tr>
</tbody>
</table>
Official Documents

• Can be found on the CLARIN-NL website
• Since July 3, 2011

http://www.clarin.nl/

http://www.clarin.nl/node/183
Questions

• For any technical questions:
  – Contact the CLARIN-NL Helpdesk!
  – helpdesk.clarin@uu.nl

• For any questions related to this proposal, or for partner search:
  – Contact the CLARIN-NL office!
  – clarinnl@uu.nl
Thanks for your attention!
Background

• CLARIN-NL Phase 1: Specification and Design
• Incorporate existing data and tools
• Standards and Interoperability
  – http://www.clarin.eu/recommendations
• Conversions and Adaptations
  – Standardized data and metadata
  – Web applications structured to become web services
  – Demonstrators
  – Requirements and Desiderata
Roles

- User
- Data Provider
- Technology Provider
- Infrastructure Specialist
User

• Is a researcher from
  – Linguistics, or
  – Humanities more broadly

• Wants to investigate research question(s)

• Some data and tools can facilitate this

• Must be the project leader
Data provider

- Has digital language data
- Can be used to address the user’s research question
- Has the rights to use these data and to make them available on CLARIN-servers
Technology provider

• Has technology and expertise on this technology
  – E.g. language or speech technology
• Can be used to address the user’s research question
  – Can form the basis for a web application and demonstrator, or
  – Can be used for resource curation
• Has the rights to use this technology and (for web applications) to make them available on CLARIN-servers
Infrastructure Specialist

• Has a deep understanding of
  – CLARIN service-oriented architecture and its requirements, and/or
  – Data, metadata or tool format standards and best practices supported in CLARIN

• Will assist a project and its members if needed

• Always via the CLARIN-NL HelpDesk
Roles and People

- Often roles will be played by different persons (possibly of different organisations), but
- Several roles may coincide in one person
  - E.g. user & data provider
  - Data provider & technology provider
Resource Curation Project (1)

- Resource must be brought in CLARIN-standards compatible format
- Metadata descriptions must be created and made available, including persistent identifiers, using CMDI
- Linguistic encoding must be related to the data category registry as implemented in ISOCAT: http://www.isocat.org
Resource Curation Project (2)

• Resource must be properly documented
• Resource must be made available on a server of a designated CLARIN centre
• All aspects must be tested, e.g.
  – Metadata harvesting test
  – Testing against XML schemata
  – …
Demonstrator Project (1)

• Adapt existing tool / application into
  – A web-based application with
    • Clearly separated user interface
    • Core component for which an API is defined
      – Can be turned into a web-service if desired/required

• Provide Data for the demonstrator
Demonstrator Project (2)

• Provide Demonstration scenario
  – Optionally movie or screen capture sequence
• Provide proper documentation
• Metadata description
• Make it available on a server of a designated CLARIN center
Requirements and Desiderata

• Data and Metadata
  – Requirements for data formats and encoding standards
  – IPR / restricted use / ethical issues requirements
  – Metadata elements and formats

• Tools / Applications
  – Processing requirements
  – Memory requirements
  – Network Bandwidth requirements
  – API requirements (e.g. Calling conventions)

• Infrastructure
  – Web-service wrappers
  – User workspace requirements
  – Repository Requirements
  – Requirements for registering and resolving PIDs

• Semantic interoperability
  – Data Category Registry
  – ISOcat