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

- Requirements and Desiderata (RDs)
- Overview of RDs
- Examples and Actions Taken
- Conclusions
RDs: come from Project Goal (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
RDs: Examples

• 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)
RDs: Examples

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

• Semantic interoperability
  – Data Category Registry
  – ISOcat
RDs: Actuals (5 projects)

<table>
<thead>
<tr>
<th>Row Labels</th>
<th>Count of Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISOCAT</td>
<td>8</td>
</tr>
<tr>
<td>Web Services /Tools</td>
<td>7</td>
</tr>
<tr>
<td>Metadata</td>
<td>5</td>
</tr>
<tr>
<td>Project Management</td>
<td>3</td>
</tr>
<tr>
<td>IPR &amp; Ethical Issues</td>
<td>2</td>
</tr>
<tr>
<td>Standards</td>
<td>2</td>
</tr>
<tr>
<td>PIDs</td>
<td>1</td>
</tr>
<tr>
<td>Desired Functionality</td>
<td>1</td>
</tr>
<tr>
<td>Search</td>
<td>1</td>
</tr>
<tr>
<td><strong>Grand Total</strong></td>
<td><strong>30</strong></td>
</tr>
</tbody>
</table>
RDs: Examples

• **ISOCAT**
  – Semantics of DCs ill-defined
  – Too many closely related DCs
    • Which of these is going to be standardized?
  – I need the same mappings as another project, but probably more: which ones?

• **Action:**
  – Several workshops to discuss problems and possible solutions
  – Several ISOCAT tutorials held
  – Appointment of ISOCAT coordinator Ineke Schuurman
RDs: Examples

• Tools / Web services
  – Authentication not implemented—everyone can see everyone’s files
  – Problems with performance in case of large outputs
  – CLAM documentation can be improved

• Action:
  – Support for web services and CLAM by Maarten van Gompel (now Uvt soon RUN)
  – To start as soon as possible
RDs: Examples

• Metadata
  – Metadata profile for tools is needed
  – Improvements for Metadata Profile / Component Editor desired

• Action:
  – Regular Metadata tutorials
  – Desired improvements discussed
  – Budget reserved for creating a metadata profile for tools / applications and web services
    • Plan under development
    • Looking for specialist
RDs: Examples

• Project Management
  – Coordinate latter tasks with initial tasks at the beginning of the project!
  – Organize the project in such a way that one person leaving is not disastrous

• Action:
  – We mention these here so that you can take them into account
RDs: Examples

• Standards
  – My data do not fit any standard supported by CLARIN
  – (in such a case: First ask a recommendation from the HelpDesk before continuing!!!)

• Action:
  – We are working on making the supported standard list a dynamic registry
  – We are working on a policy for approving additions to the CLARIN Supported Standards
RDs: Examples

• IPR & Ethical Restrictions
  – My data are partially publicly accessible but
  – Full accessibility only if a special license is signed and approved

• Action:
  – We recommended to work with two sets of metadata
  – We take this into account in designing procedures / protocols for handling IPR & Ethical Issues in the CLARIN-infrastructure
RDs: Conclusions

- We all learn a lot from actually carrying out resource curation and demonstrator projects
- Make sure your Req and Des document contains all problems / wishes/ requirements you encountered
- The EB already has taken action and will continue to take action to address matters raised in these documents
- Using these documents will help us achieve compatibility of the CLARIN infrastructure with YOUR data/tools
Thanks for your attention!