Working towards a Metadata Federation of CLARIN and DARIAH-DE

Thomas Eckart
Natural Language Processing Group
University of Leipzig
teckart@informatik.uni-leipzig.de

Tobias Gradl
Applied Computer Science Group
University of Bamberg
tobias.gradl@uni-bamberg.de
Agenda

- **Metadata Infrastructure in CLARIN**
  (CMDI, Component Registry, OAI-PMH, VLO, Metadata Curation ...)

- Metadata Infrastructure in DARIAH-DE
- General Thoughts
- Steps to a MD Federation
- Conclusion
DARIAH-DE MD Infrastructure: The big picture
DARIAH-DE MD Infrastructure: Differences and Analogies

- No central integration model, but “semantic clusters”
- Integration models are generated according to needs of particular research questions
- Larger perspective: mapping to more generic schemata

<table>
<thead>
<tr>
<th>DARIAH-DE</th>
<th>CLARIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collection</td>
<td>~ Centre</td>
</tr>
<tr>
<td>Collection Registry</td>
<td>Centre Registry</td>
</tr>
<tr>
<td>Schema Registry</td>
<td>CMDI Component Registry</td>
</tr>
<tr>
<td>Mapping Registry</td>
<td>None, implicit via ConceptLinks</td>
</tr>
<tr>
<td>Generic Search</td>
<td>VLO</td>
</tr>
</tbody>
</table>
DARIAH-DE MD Infrastructure: Mapping Registry

URL: http://schereg.de.dariah.eu/
General Thoughts

- Design decision reflect characteristics of the targeted research communities
- As a consequence, implemented solutions differ in their architecture and functionality
- Basic assumption: both metadata infrastructures are *optimized* for their specific use cases and user groups
  - No need for another interoperability layer, reusing and combining functionality instead
Steps towards a MD Federation: DARIAH-DE → CLARIN

- Central interface: Dublin Core via OAI-PMH
- CLARIN MD Harvester already capable of transformation to CMDI; test harvests
- Acceptable (but still suboptimal) results in VLO
- Minor changes in VLO vocabulary mappings
- Problematic:
  - (Missing) License information
  - Resource links usually “Landing pages”
Steps towards a MD Federation: DARIAH-DE → CLARIN

URL: http://clariah.informatik.uni-leipzig.de/vlo/?fq=nationalProject:DARIAH-DE
Steps towards a MD Federation: CLARIN → DARIAH-DE

- Import of CMDI schema in Schema Registry
- Manual creation of mappings CMDI ↔ Dublin Core
- Creation of test collections in Collection Registry for CLARIN OAI-PMH endpoints
- Import in Generic Search
Steps towards a MD Federation: CLARIN → DARIAH-DE

URL: https://search.de.dariah.eu/
Conclusion

• No major obstacles (but open work on details); still prototypical implementation

• Potential of mutual re-use of infrastructure components
  - Mapping registry: making relations between (CMDI) schemas explicit?
  - (Semi-)automatic import of schemas between CR and SR?
  - Metadata curation workflow for DARIAH-DE resources?

• Main problem: outstanding decision about relevant resources that should be imported in production systems
Thank you!

Questions? Remarks?

More information about MD in DARIAH-DE?