



Towards Integration of the CLARIN Technical Infrastructure, Content, and Services

Erhard Hinrichs

Chair of National Coordinators Forum,
Member of CLARIN ERIC Board of Directors

[Reference Document: Towards Integration, ID: CE-2012-0018](#)

Overall Objective



- Overall Objective:

Integration of the technical infrastructure and of CLARIN services (language resources, tools, and knowledge sharing)

- Method:

Layered integration approach

Layers of Integration



- Purely infrastructural integration
- Basic services
- Integration of object data and tools

Purely infrastructural integration



Establishing a CLARIN Trust Domain:

- Using SAML 2.0 standard for secure information exchange and AA middleware (Shibboleth, SimpleSAMLPHP)
- Respecting license access restrictions of service providers
- Open issues to be solved:
 - Attribute sharing among service providers
 - Delegation of user identities

Basic services



- Metadata provisioning using the Component MetaData Infrastructure (CMDI)
- Metadata harvesting using the OAI-PMH standard
- Metadata browsing and searching using the Virtual Language Observatory VLO
- Working out a CLARIN access policy to metadata

Integration of object data and tools



Interoperability of resources and tools

- Requires standardized data formats
- Major Challenges: Legacy data and tools using diverse data formats
- Possible solutions: data conversion (time consuming and expensive) or use of automatic conversion scripts

Milestones for AA



- M1-1 (October 01, 2012)
Compile a list of national identity federations for all CLARIN ERIC members (input to be provided by the National Coordinators) and compile information about the setup and configuration of individual IDPs of the members of the national CLARIN consortia.
- M1-2 (October 22, 2012)
Review the lists under M1-1 during the CLARIN Conference in Sofia.
- M1-3 (December 31, 2012)
Integration of additional national identity federations into the CLARIN Service Provider Federation (SPF) maintained by MPI Nijmegen.

Milestones for Basic Services



- M2-1 (October 1, 2012)
Compile a list of different types of resources and tools that are included as offers by each member country in the Annexes of the CLARIN Agreements and compile a list of existing metadata for these resources and tools.
- M2-2 (October 22, 2012)
Review the existing metadata in M2-1 during the CLARIN Conference in Sofia and begin to convert them into CMDI format in hands-on sessions.
- M2-3 (December 31, 2012)
Establish a repository for resources and tools in at least one national Type A/B centre in each CLARIN ERIC member country and implement an OAI-PMH provider endpoint.

Milestones for Interoperability



- M3-1 (October 1, 2012)
Nominate resource types and tools as candidates for standardization efforts.

- M3-2 (October 22, 2012)
During the CLARIN Conference in Sofia:
 - Presentation & discussion of the proposals obtained in M3-1
 - Presentation of the federated search specification and discussion of ways to extend this functionality beyond CLARIN-AT, CLARIN-D, and CLARIN-NL.



Thank you
for your attention