

DELIVERABLE D12: REQUIREMENTS AND DESIDERATA FOR THE CLARIN INFRASTRUCTURE

Author:

Victor de Boer (VU Amsterdam)

Deliverable type: DOCUMENT

Version	Date	Author	Description
0.1	19-4-2014	Victor de Boer	First version

This deliverable describes the requirements and desiderata for the CLARIN infrastructure

CLARIN AND LINKED DATA

As part of the project, a triple store was set up at Huygens ING. In that sense, triple stores and Linked Data now are part of the Clarin infrastructure. However, it would be advisable if Clarin or any continuation of Clarin would explicitly develop a policy to include Linked Data / Semantic Web technologies as part of their data sharing methods. Semantic Web technologies including RDF, resolvable URIs and SPARQL are quickly becoming the W3C and de-facto standard for sharing data across the Web. However, the Clarin centers at the moment do not explicitly support hosting of RDF data. It is our suggestion that triple stores would be hosted at one of the Clarin centers (preferably DANS).

Furthermore, the relation between the Clarin infrastructure/ CMDI / ISOCAT and the Linked Data set of best practices remains unclear. For example, Linked Data uses unique identifiers (URIs) and the Resource Description Framework for individual resources (instances), classes and properties (metadata fields). In that sense there is no clear distinction between data and metadata. The Clarin requirements for using CMDI metadata is orthogonal to that of Linked Data. Efforts exist to find mappings between the two worlds but to individual data providers it is not easy to submit pure data to the Clarin infrastructure that complies to the Clarin requirements.