Contribution ID: **14** Type: **Presentations**

Bringing Europeana and CLARIN together: Dissemination and exploitation of cultural heritage data in a research infrastructure

Thursday, 30 November 2017 11:00 (15 minutes)

We present the joint work by Europeana (http://www.europeana.eu), a European cultural heritage (CH) infrastructure, with CLARIN (http://www.clarin.eu), a European research infrastructure, to make promptly available for research use the vast data resources that Europeana has aggregated in the past years.

Europeana provides access to digitised cultural resources from a wide range of institutions all across Europe. It seeks to enable users to search and access knowledge in all the languages of Europe, either directly via its web portals, or indirectly via third-party applications leveraging its data services. The Europeana service is based on the aggregation and exploitation of (meta)data about digitised objects from very different contexts. The Europeana Network has defined the Europeana Data Model (EDM) to be used as its model for interoperability of metadata, in line with the vision of linked open vocabularies. One of the lines of action of Europeana, is to facilitate research on the digitised content of Europe's galleries, libraries, archives and museums, with a particular emphasis on digital humanities.

CLARIN (Common Language Resources and Technology Infrastructure) is a networked federation of language data repositories, service centres and centres of expertise.
CLARIN aggregates metadata from resource providers (CLARIN centres and selected "external"parties), and makes the underlying resources discoverable through the Virtual Language Observatory (VLO) to provide a uniform experience and consistent workflow.
The VLO can also serve as a springboard to carry out natural language processing tasks via the Language Resource Switchboard (LRS), allowing researchers to invoke tools with the selected resources directly from its user interface. The potential inclusion of many new CH resources by 'harvesting'metadata from Europeana, opens up new applications for CLARIN's processing tools.

CLARIN and Europeana do not share a common metadata model, and therefore a semantic and structural mapping had to be defined, and a conversion implemented on basis of this. CLARIN's ingestion pipeline was then extended to retrieve a set of selected collections from Europeana and apply this conversion in the process. Several infrastructure components had to be adapted to accommodate the significant increase in the amount of data to be handled and stored. Currently about 775 thousand Europeana records can be found in the VLO, with several times more records expected in the foreseeable future. Currently, about 10 thousand are technically suitable for processing via the LRS. Relatively straightforward improvements to the metadata on the side of Europeana and/or its data providers could substantially increase this number. CLARIN is working with Europeana to implement such improvements. More tools are also expected to be connected to the LRS in the short to mid-term, which is also expected to lead to an increased 'coverage'.

As a next step, CLARIN can extend and refine the selection of included resources, and Europeana can adapt their data and metadata to optimally serve the research community. CLARIN's experience and potentially part of its implementation work can be applied to integrate Europeana with other resource infrastructures.

Topic Area

Interoperability

Type of abstract

Presentation (15 minutes)

Primary authors: NEUDECKER, Clemens (Berlin State Library / Europeana Newspapers); Dr ESKEVICH, Maria (CLARIN ERIC); Dr FREIRE, Nuno (Europeana/INESC-ID); GOOSEN, Twan (CLARIN ERIC)

Presenter: Dr ESKEVICH, Maria (CLARIN ERIC)

Session Classification: Interoperability presentations