

Title Europeana DSI-3 task 3.2.7 work plan
Version 1
Author(s) Twan Goosen, Dieter Van Uytvanck
(CLARIN ERIC), Nienke van Schaverbeke
(Europeana Foundation)
Date 2017-12-06
Status Draft
Distribution Public
ID CE-2017-1147



Task description

From the DSI-3 Tender¹:

Subtask 3.2.7. Connect to Research Infrastructures (EF , CLARIN)

Building on the work in Europeana DSI-2, EF will continue to work with CLARIN on incorporating digitised cultural heritage data into the European Research Infrastructure for Language Resources and Technology as well as (linguistic) processing of the data. Through its infrastructure CLARIN makes digital language resources available to scholars, researchers and students, primarily in the humanities and social sciences. CLARIN will undertake a systematic assessment of data complementarity and functionality on a selected number of Europeana datasets/collections. It will also investigate exploiting new data such as the Europeana Newspaper's collections, technical metadata and direct links.

This task does not produce a deliverable.

Subtasks

Subtask A. Task details & work plan

- Create and sign DSI-3 partners agreement
- Create work plan (this document)

Subtask B. Harvesting and importing workflow

1. Establish a durable, scalable pipeline for regular metadata harvesting.
2. Set up schedule for regular (monthly?) harvesting and import.
3. Assess performance, provide feedback to EF
 - a. Investigate if the OAI viewer/curation module can be used for this

Subtask C. Metadata conversion extensions

1. Add support metadata hierarchies

¹ Deployment and Maintenance of Europeana DSI core services - SMART 2016/1019

2. Improve mapping to collection name (see [GitHub issue](#))

Subtask D. Assessment of data complementarity

Undertake a systematic assessment of data complementarity and functionality on a selected number of Europeana datasets/collections.

1. Quantify coverage in terms of the discovery & processing pipeline
 - Asses the current state
 - How many resources (in how many data sets) for how many languages can be discovered in the vLO and processed with how many tools using the LRS?
 - Asses the potential
 - Which high potential data set can be identified that at present do not suffice in terms of (meta)data quality? Take into account which tools are available in the LRS and which languages are supported.
2. Evaluate impact in terms of the Europeana Research strategy
 - “[By 2020] Digital Humanities Researchers use Europeana as a trusted source of cultural heritage data to develop new knowledge and insights from our past or create new research methodologies. They can use the data directly from the Europeana platform or they can find it in research-related web services and networks. By nurturing partnerships that place the data in the workflow of the researcher, and allow for download of high-quality re-usable cultural heritage data we reduce their transaction costs for finding and mining primary resource materials for their research.” (Business Plan 2017)
3. Assess (potential) uptake of Europeana resources in CLARIN (towards end of DSI-3)
 - See what conclusions we can draw on basis of VLO/LRS access statistics.
 - Ask the User Involvement committee and/or a representation of CLARIN community members to evaluate a Europeana data sets available through the VLO or directly via Europeana and their potential for research usage.
4. Share intermediate outcomes with Europeana, in particular information on data (sets) that could be improved.
 - Europeana reports to CLARIN about changes in the situation (see also [Subtask E. Evaluate novel content](#))
 - See what means of sharing this information is most suitable

Subtask E. Evaluate novel content

1. Extend the set of harvested and imported data sets
 - Based on input from Europeana, and their regular reporting of data (quality) updates (see [Subtask D. Assessment of data complementarity](#))
2. Harvest and import newly added records to previously included data sets

Subtask F. Newspaper API pilot

If available (expected 2018 Q1 or early Q2) and it fits the schedule/time available, investigate if the Newspaper API provides added value compared to OAI-PMH and collections API and adapting the retrieval workflow to support it is worth the required effort.

Subtask G. Dissemination

Promote and disseminate the output of the work and value of cultural heritage for researchers through blog posts, conferences and social media.

- Blog post to present new content and demonstrate metadata improvements
- Screencast of processing workflow with Europeana data?
- Organise thematic workshops with researchers from CLARIN community?
 - <http://www.theeuropeanlibrary.org/tel4/newspapers/gallery?view=discover>
- DH conferences presence

Planning

There is a total of 3 PMs available from CLARIN for this task. Reserving .5PM for meetings, reporting and dissemination work (subtask G), the remaining 2.5 PMs are expected to be distributed over the tasks according to the following (provisional) schedule:

M1-M3

September - November 2017

Subtask A. Task details & work plan	.25 PM
Subtask B. Harvesting and importing workflow	.25 PM

M4-M6

December 2017 - February 2018

Subtask C. Metadata conversion extensions	.25
Subtask D. Assessment of data complementarity (1/2)	.5

M7-M9

March 2018 - May 2018

Subtask D. Assessment of data complementarity (2/2)	.5
Subtask E. Evaluate novel content	.5

M10-M12

June - August 2018

Subtask F. Newspaper API pilot	.25
--------------------------------	-----