

Launch of the Europeana Newspapers project

The Hague, 26th of June 2012



A group of 17 European partner institutions have joined forces in the “Europeana Newspapers” project to, over the next 3 years, provide more than 18 million newspaper pages to the online service Europeana. Europeana is a single access point to millions of digitised books, paintings, films, museum objects and archival records sourced from throughout Europe.

The Europeana Newspapers project is funded under the Competitiveness and Innovation Framework Program 2007-2013 of the European Commission with the aim of aggregation and refinement of newspaper content through The European Library.

Each library participating in the project will distribute digitised newspapers and full-text via Europeana. The project aims to make the newspaper content directly accessible for users through a special interface within the content browser. This will be integrated into the Europeana portal and will allow queries of phrases or single words within the newspapers’ texts. This goes far beyond the standard libraries catalogue search functions which usually allow the searching by date or title only.

The project addresses challenges linked with digitised newspapers such as Optical Character Recognition (OCR), Optical Layout Recognition (OLR), article segmentation and page class recognition, and named entity recognition (NER). OCR is the electronic conversion of scanned images of handwritten, typewritten or printed text into machine-encoded text. OLR is concerned with the detection and separation of articles on a scanned page with more than one article. NER seeks to locate entities in the full text and to classify them according to standardised names for persons, locations, and organisations.

The project will also evaluate the quality of the refinement technologies and transform the local metadata into the Europeana Data Model standard in close collaboration with stakeholders from the public and private sector.

The Europeana Newspapers project is co-ordinated by the Staatsbibliothek zu Berlin – Preußischer Kulturbesitz. Follow the advancements of the Europeana Newspapers project at www.europeana-newspapers.eu. For any further information please contact Hans-Jörg Lieder or Thorsten Siegmann at Staatsbibliothek zu Berlin via info@europeana-newspapers.eu.

Project Partners:

Berlin State Library
National Library of Estonia
University of Helsinki, National Library of Finland
National Library of France
CCS Content Conversion Specialists GmbH
National Library of Latvia
University of Beograd
Dr. Friedrich Tessmann Library
University of Salford

National Library of the Netherlands
Austrian National Library
Hamburg State and University Library
National Library of Poland
LIBER Foundation
National Library of Turkey
University of Innsbruck
The British Library
The European Library

Europeana in a nutshell

Europeana is a multi-lingual online collection of millions of digitized items from European museums, libraries, archives and audiovisual collections. Currently Europeana gives integrated access to 23 million books, films, paintings, museum objects and archival documents from some 2.200 content providers from across Europe.

www.europeana.eu

The Europeana Newspapers project is partially funded under the ICT Policy Support Programme (ICT PSP, http://ec.europa.eu/ict_psp) as part of the Competitiveness and Innovation Framework Programme by the European Community. This document reflects only authors’ views. The EC is not liable for any use that may be done of the information contained therein.

