

A Gateway to European Newspapers Online: building common history and identity around digital materials

Vesna Vuksan

vuksan@unilib.bg.ac.rs

Senior Librarian, Department of Development and System Administration, Digitization and Cultural Activities

University of Belgrade, University Library "Svetozar Markovic", Belgrade, Serbia

Adam Sofronijevic

sofronijevic@unilib.bg.ac.rs

Head, Department of Development and System Administration, Digitization and Cultural Activities

University of Belgrade, University Library "Svetozar Markovic", Belgrade, Serbia

**INFORUM 2012: 18th Conference on Professional Information Resources
Prague, May 22-24, 2012**

Abstract

"Europeana Newspapers" project is a CIP ICT-PSP project funded by European Commission which aims at bringing newspapers into Europeana. Seventeen partnering institutions, including some of the oldest and most famed libraries of the continent will be aggregating newspaper content for Europeana. Over 10 million newspaper pages will be added over the period of three years making Europeana truly a comprehensive source of shared European history and identity.

Keywords: newspapers, online, Europeana, digitization, libraries, CIP ICT-PCP, European Commission

Introduction

Free exchange of information plays a vital role for researchers and general public as it provides an opportunity for them to participate more fully in the research community or general information spectrum without having to face financial barriers. University Library in Belgrade has been involved in several projects and initiatives during the last decade that offer free data sharing, thus positioning itself amongst the most vocal advocates for open access in Serbia. One of those projects is “Europeana Newspapers”.



The project started on February 1st 2012 and by 2015 usage of advanced technologies including OCR, OLR/article segmentation, and NER - named entity recognition will yield millions of full text pages for the European portal of digital objects. The project will also add to developing of Europeana data model - EDM, standardization of metadata in newspapers and will be providing general guidelines for further digitization projects of newspaper materials. Digital editions of newspapers published during the First World War will provide important addition for EU-funded project “Europeana Collections 1914-1918” that started in May 2011. Contemplating reuse of digital objects is an important issue for all future digitization projects and “Europeana Newspapers” project sets a shining example in this area.

Project's objectives

Newspapers represent a great value in terms of history, culture and identity; they are constantly in demand by researchers, on one hand, and the general public on the other hand. Librarians have been busy digitizing newspaper collections to meet this demand; however the access to these collections is still often set to local access points which limits collections' usability and accessibility. One of the main project goals is to increase accessibility to digitized newspaper collections. Also, the project will bring stakeholders together and will make the process of digitization cost-efficient in areas such as image refinement and the development of newspaper metadata.

Most importantly, it will enable the users to explore the rich past of Europe through a single point of access: Europeana.

In addition, the project addresses challenges particularly connected with digitized newspapers:

- use of refinement methods for OCR, OLR/article segmentation, and named entity recognition (NER), and page class recognition to enhance search and presentation functionalities for Europeana customers,
- quality evaluation for automatic refinement technologies,
- transformation of local metadata to the Europeana Data Model (EDM)
- metadata standardization in close collaboration with stakeholders from the public and private sector.

Participants and Roles

The project consortium is composed of main stakeholders from European Union member states and countries associated with CIP-ICT PSP 5th Call:

1. Staatsbibliothek zu Berlin – Preußischer Kulturbesitz (Germany, Project Coordinator and Manager, Content Provider, WP1 Lead)
2. National Library of the Netherlands – Koninklijke Bibliotheek (Netherlands, WP2+4 Lead, Content Provider, partner in TEL, IMPACT, Europeana)
3. National Library of Estonia (Estonia, Content Provider)
4. Österreichische Nationalbibliothek (Austria, Content Provider)
5. National Library of Finland (Finland, Content Provider)
6. Staats- und Universitätsbibliothek Hamburg (Germany, Content Provider)
7. Bibliothèque nationale de France (France, Content Provider)
8. National Library of Poland (Poland, Content Provider)
9. University of Salford (United Kingdom, Technical Partner, WP3 Lead)
10. CCS Content Conversion Specialists GmbH (Germany, Technical Partner)
11. Stichting LIBER (EU, WP6 Lead)
12. National Library of Latvia (Latvia, Content Provider)
13. National Library of Turkey (Turkey, Content Provider)
14. University Library of Belgrade (Serbia, Content Provider)
15. University of Innsbruck – Department for Digitization and Digital Preservation (Austria, Technical Manager, Technical Partner, WP5 Lead)
16. Landesbibliothek Dr. Friedrich Tessmann (Italy, Content Provider)
17. The British Library (United Kingdom, liaison to private publishers)

All libraries participating in the project will distribute digitized newspapers and full-texts free of any legal restrictions to Europeana.

University library “Svetozar Markovic” is a partner in “Europeana Newspapers” project and will be involved in each segment of the project. As a data provider University library will add 380.000 pages of newspapers published in Serbia before 1941. Librarians from Belgrade will participate in developing EDM and finding new ways to make Europeana content more usable and more used. By using advantages of high-end technologies available in the project University library aims at creating attractive and innovative digital objects that will catch eye and attention of users and bring back into focus the quality historical content.

Impact

The most obvious result of the “Europeana Newspapers” project will be the provision of a critical mass of European newspaper content via Europeana. However, sustainability of the project will be visible through these outcomes:

- Registry of digital newspaper holdings in major European public institutions
- Support for libraries in making newspaper data available to The European Library
- Best practice recommendations for metadata formats
- Best practice recommendations for refinement procedures
- Quality assurance and quality prediction tools
- Provision of data to Europeana
- Increasing the attractiveness of Europeana content
- New type of user experience within Europeana.

Fostering digitization of cultural heritage with a European value has been in the focus of the European digital library initiative designed by European Union. This resulted in the launch of Europeana in 2008 and several projects have contributed to Europeana but many of these resources were mainly relevant to researchers.

Newspapers, on the other hand, offer the political and cultural affairs of cities, regions, and countries on a daily basis. They cover segments of life relevant to virtually all citizens of Europe and attract huge numbers of users. The “Europeana Newspapers” project will help move the Europeana service to a new level by making searchable full text versions of newspaper articles available to Europeana, and increasing the attractiveness of the service significantly.

Conclusion

“Europeana Newspapers” project’s consortium of 17 partners from 12 countries will deliver not only digitized newspapers collections in the following three years, but will also provide best practice recommendations based on objective and measurable factors for digitization, refinement, workflows, metadata and evaluation tools. This will include novel planning and

quality estimation tools to aid decision making processes for future digitization projects. It will also increase Europeana usability by making it the largest provider of pan-European newspaper collections and a comprehensive source of shared European history and identity. Customers, researchers and stakeholders of the newspaper community will be constantly updated about the latest efforts regarding “Europeana Newspapers” project through a specially designed project’s website www.europeana-newspapers.eu, social networking websites and many different workshops and activities planned until 2015.

References

Europeana Newspapers, <http://www.europeana-newspapers.eu/>, retrieved April 27, 2012.

Europeana Libraries, <http://www.europeana-libraries.eu/>, retrieved March 12, 2012.

The European Library, <http://www.theeuropeanlibrary.org/>, retrieved April 10, 2012.

University Library in Belgrade, <http://www.unilib.rs/>, retrieved April 28, 2012.

Europeana Collections, <http://www.europeana-collections-1914-1918.eu/>, retrieved March 28, 2012.

ICT Policy Support Programme (ICT PSP), http://ec.europa.eu/ict_psp/, retrieved April 22, 2012.