

# Overview of a Final Meeting of a Consortium of the "Europeana Newspapers" Project

**DATE OF SUBMISSION:**  
**DATE OF ACCEPTANCE:**

1 April 2015  
14 April 2015

Nataša Dakić

dakic@unilib.bg.ac.rs

Dejana Kavaja Stanišić

kavaja@unilib.bg.ac.rs

*University of Belgrade,  
University Library "Svetozar Marković"*

Unlike books, newspapers have never aspired to be eternal. The main goal of newspapers - to bring about new information every day, without taking care of the quality of the printing paper - resulted in a great number of brittle, fragile original copies which are nowadays stored in library collections. In order to preserve them and enable their use, many libraries first microfilmed many titles, and with the development of technologies they started the process of scanning. As both microfilmed and scanned texts are difficult to read or absolutely unreadable giving only limited browsing possibilities, it is necessary to clear and additionally process the text using new tools and technologies which would provide Optical Character Recognition (OCR), Optical Layout Recognition (OLR) and Named Entity Recognition (NER), with the aim of enabling more efficient search for pictures, articles or certain terms in texts.

Consideration of these needs brought about the start of the project "Europeana Newspapers" financed by the European Commission within the CIP ICT-PSP Best Practice Network 5<sup>th</sup> Call which was successfully finished. At first, there were 18 full partner institutions<sup>1</sup>, among which there were some of the most important national and university European libraries, including the

University Library "Svetozar Marković". In the meantime, the importance of the project was recognized so it has spread and many libraries, organizations and other European projects have become part of it.

Success analysis of achieved goals and results of the project and the possibility for future cooperation among the project partners were some of the topics of the final meeting of the Consortium of the "Europeana Newspapers" project. This two day meeting held at the National Library of Estonia (Eesti Rahvusraamatukogu) in Tallinn on January 22<sup>nd</sup> and 23<sup>rd</sup> was attended only by the full partners in the project. Moreover, the meeting was attended by the leaders of all six work packages who handed in their individual reports.

Clemens Neudecker, coordinator of the project, opened the meeting and in his presentation he referred to the set goals and achieved results. In his analysis he estimated that all set goals were achieved while some of them were even surpassed. For instance, out of the expected 10 000 000 digitized pages which would be fully available and searchable in full text, today there are 10 132 955 available in full text on The European Library Portal (TEL) while the total number of all available digitized pages is 20 014 931, which includes the pages provided by the associated

<sup>1</sup> <http://www.europeana-newspapers.eu/consortium/project-partners/#project>

partners in the project. He also pointed out that there were some products that resulted from the project and which represent added value to the project as they were not planned in the project documentation. Namely, Europeana integrated an application for browsing newspaper collections and enabled browsing newspapers in full text on the Europeana portal. Project partners such as Content Conversion Specialists, Staats- und Universitätsbibliothek Hamburg and University of Salford developed an application for accessing and browsing newspapers on mobile devices. The National Library of Netherlands (Koninklijke Bibliotheek) widened the scope of its work related to the recognition of noun entities by developing a tool which supports the polysemy of entities and their connection with identifiers in DBpedia.

During the meeting it was emphasized that the Consortium of the project should ensure more stable cooperation with Europeana so that the further dissemination of the project would provide information about both access points, i.e. about Europeana and The European Library. The strategy for sustainability of the already existing digital collections should be established and the methods for the enlargement of newspaper collections and attraction of new users should be determined. Furthermore, it is necessary to monitor the quality of delivered metadata and to make sure that all those who deliver content sign an Agreement on Data Exchange.

In his report Alastair Dunning, a TEL representative, pointed out that The European Library collected, in addition to pictures and full text pages, metadata which it had successfully converted into the Europeana Data Mode. He reported that overall 3 342 titles of newspapers were added to the search engine, which is not a final number as the final set of data which should be provided by CCS and Universität Innsbruck should be added and also a collection of newspapers from Turkey. At the end of his presentation he announced the decision of the Board of Europeana and The European Library on merging these two organizations and he pointed out potential issues which might result from this decision as it had been initially planned that these

two organizations would continue their work as two separate entities.

Representatives of partner libraries in short five minute presentations talked about their results, observations, expectations and further plans. They mentioned challenges and difficulties they faced while working on the project and suggested possible ways of integration of successful results into their future work. They presented their plans in the field of newspaper collections' digitization which most often included finding a way to improve the quality of already processed texts and working on a new project.

Within the "Europeana Newspapers" project the University Library digitized 392.323 of 45 selected titles of Serbian newspapers between 1830 and 1944. All the materials are OCR-ed, which makes this digital textual corpus completely searchable. Results of optical character recognition were encouraging bearing in mind that the complete newspaper collection (except for one title) was printed in Cyrillic. However, to provide better readability and search of digitized newspaper collection, the University Library managed to develop a tool which would enable correction of METS/ALTO files by hand and in that way improve the quality of pages in full text. After the initial difficulties, a tool for searching the entire collection was developed and it will also be used for searching other digital material. In order to promote and ensure greater visibility of materials provided by this Project, the Library promoted it in on the National Television of Serbia in a 30 minute programme and in numerous articles published online. At the same time this was the content of the five minute presentation of our results in the final meeting of the Consortium of the "Europeana Newspapers" project.

This two days' meeting created pleasant working atmosphere and it left the impression that all partners are motivated to focus on upcoming tasks and activities in the last phase of the project. Thirty registered participants, project associates, used the time to exchange experience and open important discussions about possible future activities and cooperation within other European projects.