
Europeana Newspapers – A Gateway to European Newspapers Online

Final Report

Revision: 1.0

Authors: Clemens Neudecker (SBB)
Hans-Jörg Lieder (SBB)
Sandra Kobel (SBB)

Contributions: All project partners

Table of contents

1. Introduction	4
2. Project Objectives	6
3. WP1 – Coordination and management.....	7
4. WP2 – Refinement of digitised newspapers	8
5. WP3 – Evaluation and quality assessment.....	10
6. WP4 – Aggregation and presentation of digitised newspapers	13
7. WP5 – Metadata best practice recommendations.....	17
8. WP6 - Dissemination and exploitation	19
9. Concluding remarks	22



Dive into the details of Europe's historic newspapers.

 www.europeana-newspapers.eu

1. Introduction

The Europeana Newspapers project (<http://www.europeana-newspapers.eu/>) is an EU-funded Best Practice Network under the theme CIP-ICT-PSP.2011.2.1 - Aggregating content in Europeana.

The project lasted from 1 February 2012 until 31 March 2015 and brought together 18 project partners, 11 associated partners and 35 networking partners to achieve the ambitious goal of making vast amounts of European digital historic newspapers available via two prominent cultural heritage websites, [Europeana](#) and [The European Library](#).



Fig. 1: Project network – red: full partners, blue: associated partners, green: networking partners

Full project partners:

- [Berlin State Library \(SBB\)](#)
- [National Library of the Netherlands \(KB\)](#)
- [National Library of Estonia \(NLE\)](#)
- [Austrian National Library \(ONB\)](#)
- [University of Helsinki, National Library of Finland \(NLF\)](#)

- [Hamburg State and University Library \(SUB HH\)](#)
- [National Library of France \(BnF\)](#)
- [National Library of Poland \(NLP\)](#)
- [CCS Content Conversion Specialists GmbH \(CCS\)](#)
- [LIBER Foundation \(LIBER\)](#)
- [National Library of Latvia \(NLL\)](#)
- [National Library of Turkey \(NLT\)](#)
- [University of Belgrade \(UB\)](#)
- [University of Innsbruck \(UIBK\)](#)
- [Dr. Friedrich Tessmann Library \(LFT\)](#)
- [The British Library \(BL\)](#)
- [University of Salford \(USAL\)](#)
- [Europeana Foundation \(EF\)](#)

Associated project partners:

- [National and University Library of Slovenia](#)
- [National and University Library of Iceland](#)
- [The Royal Library of Belgium](#)
- [National Library of Croatia](#)
- [National Library of Bulgaria](#)
- [National Library of Romania](#)
- [National Library of Luxemburg](#)
- [National Library of the Czech Republic](#)
- [National Library of Spain](#)
- [National Library of of Portugal](#)
- [National Library of of Wales](#)

Networking partners:

- [Europeana Collections 1914-10918](#)
- [APARSEN](#)
- [Europeana Cloud](#)
- [Europeana Regia](#)
- [Europeana Sounds](#)
- [EUROCLIO](#)
- [Historiana](#)
- [Gallica](#)
- [Cendari](#)
- [Delpher](#)
- [Trove](#)
- [Open Glam](#)
- [IFLA Newspaper Section](#)
- [European Newspaper Publishers Association](#)

- [BCU-Lausanne](#)
- [National Library of Switzerland](#)
- [Heidelberg University Library](#)
- [Moravian Library](#)
- [University of Nis](#)
- [National Central Library Rome](#)
- [National Central Library of Florence](#)
- [National Library of Venice](#)
- [Institute for East and Southeast European Studies](#)
- [National Library of Serbia](#)
- [Lucian Blaga – Central University Library Romania](#)
- [National Library of Israel](#)
- [National Library of Ireland](#)
- [Aristotle University of Thessaloniki Library](#)
- [National Library of Sweden](#)
- [National Library of Norway](#)
- [National Library of Georgia](#)
- [Zentralbibliothek Zürich](#)
- [Leibniz Information Centre for Economics](#)
- [Infobiro Sarajevo](#)
- [Erfgoedcel CO7](#)

2. Project Objectives

1. Make Digital Historic Newspapers Easier To Search

When historic newspapers are digitised, the resulting electronic version is often just a scanned image of the newspaper. Based on this is not yet possible to effectively search for specific articles or keywords within the text. Europeana Newspapers aimed to change that. It created electronic full-text versions of about 10 million digital newspaper pages using state-of-the-art Optical Character Recognition and Optical Layout Recognition technology. This will dramatically improve the experience of users, compared to earlier digital newspaper collections.

2. Put Digital Historic Newspapers Within Everyone's Reach

Most of the historic newspaper pages assembled by Europeana Newspapers are in the public domain. All titles are freely searchable through an interface created by [The European Library](#) and the entire collection has also been ingested into [Europeana](#). Researchers and end users are thus encouraged to make use of this vast amount of free digital resources.

3. Create Tools That Help Experts To Assess Quality

The Europeana Newspapers project developed an evaluation and quality-assessment infrastructure for newspaper digitisation. This will guide experts in making informed decisions on the planning and cost-efficient commissioning of newspaper digitisation projects. It also allows predicting with reasonable accuracy the use scenarios that will be possible based on the quality of the materials.

4. Assemble An Overview Of Newspaper Digitisation In Europe

A 2012 survey ([PDF](#)) identified all newspaper collections digitised by national, research and public libraries in Europe. It revealed the problem of making 20th century content available, and the fact that while millions of pages have been digitised, this is still only a fraction of the total physical holdings in libraries and newspaper archives.

5. Create Best-Practice Recommendations For Newspaper Metadata

Europeana Newspapers designed and released a comprehensive metadata and structure model specifically for newspapers which is based on established standards like METS and ALTO. This can serve as a best practice for those who are just starting with newspaper digitisation.

6. Raise Awareness Through Workshops and Information Days

Anyone interested in the digitisation and use of newspaper content can learn more through the project workshops and information days. Two kinds of events were organised by the project: three workshops focused on key topics of newspaper digitization and a series of information days which embedded the project learning in the local newspaper digitisation communities.

To reach these goals, the project was divided into six different work packages which were targeted at the particular objectives.

3. WP1 – Coordination and management

The main objective of WP1 was to provide the project consortium with reliable overall coordination and to ensure timely and accurate handling of all the administrative and financial tasks connected with the activities of the consortium. It managed the communication with the European Commission and monitored and reported on the progress of the various Work Packages at periodic intervals.

4. WP2 – Refinement of digitised newspapers

The main objective of WP2 was to coordinate the refinement of digitised newspaper pages with Optical Character Recognition (OCR) and Optical Layout Recognition (OLR) in order to produce electronic and fully searchable text from scanned newspaper images. The ambitious goal of the project was to apply these techniques to approximately 10 million pages from 14 libraries, and accordingly technical project partners with great experience in such refinement were selected for the processing of the newspaper pages.

The University of Innsbruck, who have gained extensive experience with OCR in various European projects about text digitisation such as [MetaE](#), [IMPACT](#) and others and who are also currently providing the technical and administrative backbone of the European eBooks on demand service [EOD](#) which includes OCR amongst its services, were responsible for the production of full-text for 8 million pages.

Content Conversion Specialists GmbH, a company specialising in newspaper digitisation and having ample experience in large-scale newspaper digitisation projects were responsible for the refinement of another 2 million pages. Using their in-house docWorks software technology for the project, it was possible to perform not only OCR, but also OLR which includes advanced features such as the separation of articles and the classification of pages (e.g. into advertisements, titles pages, etc.).

The National Library of the Netherlands was not only coordinating this Work Package, but also contributed itself through the technical development of open source software and resources for Named Entity Recognition for three languages, Dutch, German and French¹. Named Entity Recognition allows the identification of the names of persons, places and organisations within electronic text and the subsequent mark-up of these, which in turn improves search results and is an important stepping stone for the subsequent disambiguation and linking of recognised entities to other online information resources². This is particularly relevant for newspapers where persons and locations are important features of the news that are being reported.

Organising the refinement of over 10 million newspaper pages required a very clear and highly standardised workflow to ensure that all parties are aware of their tasks at all times and can always track the data at any point in the defined workflow.

In the first year of the project, available digital newspaper collections at the partners were analysed and suitable materials for refinement selected. The process and outcomes has been described in the deliverable [D2.1 Dataset for Refinement](#).

¹ <https://github.com/KBNLresearch/europeanap-ner>

² <https://github.com/KBNLresearch/europeanap-dbpedia-disambiguation>



Fig. 2: A sample of title pages from the newspapers selected for refinement

The technical requirements and a minimum quality of the newspapers were specified and a master list was created which combined all required information for each newspaper on the title level. To facilitate the necessary pre-processing of local data to meet the requirements set out in the specification, a number of supporting software tools have been created by the University of Innsbruck which have since also been released as open source³. All this is described in more detail in the deliverable [D2.2 Specification of requirements of OCR and structural refinement services for digitised newspapers in Europeana](#).

Thanks to highly efficient workflows, the University of Innsbruck was able to process significantly more content than foreseen. Overall, UIBK refined a total 8.101.864 newspaper pages until the end of the project. Next to the 8 million pages of full text delivered to The European Library, another 3.012.906 pages from the Austrian National Library and 286.662 pages from the Berlin State Library for the Europeana Collections 1914 - 1918 project were processed at no extra costs, so that overall 11.401.432 pages of full-text were produced. CCS produced OLR processing results for another 2.027.967 refined pages. The results of the automated processing were also verified manually and wrong metadata was corrected.

Thus, in total, **13.429.399** pages are now fully searchable. A detailed overview of the distribution of the upgraded content and the activities of the three technical partners is given in deliverable [D2.3 Upgraded content](#).

³ <https://github.com/dea-uibk>

During this process, decisions were made that were of impact to the workflow and outcome of the refinement and naturally, a great deal was learned. To ensure that other institutions wanting to refine digitised newspapers for their collection and that of Europeana can use the knowledge that was gathered in Europeana Newspapers, the experiences made are summarised in deliverable [D2.4 Recommendations on best practices for refinement of digitised newspapers in Europeana](#).

5. WP3 – Evaluation and quality assessment

The main objective of WP3 was to create a fully-featured evaluation infrastructure for digitised newspapers with particular focus on automated text recognition and layout analysis.

In the first year, high-level use scenarios for digitised newspapers (for instance “scanned images for browsing”, “keyword search”, “text to speech/braille output”, “full text and layout for eBook production”, “document structure for article segmentation”, “layout for repurposing for mobile devices” etc.) were analysed and further specified. An overview of common use scenarios was compiled based on the experience from previous digitisation projects. In order to identify important use scenarios for digitised newspapers, a questionnaire was created and distributed among all technical and library partners.

Another major activity was the setup of an Image and Ground Truth Repository as the central infrastructure for reference and all evaluation related purposes. The Europeana Newspapers Image and Ground Truth Repository was created and made available to interested parties. This representative and detailed dataset used for evaluating the results of the main Europeana Newspapers production workflow also represents a unique resource for future research and development activities related to training and optimisation of OCR engines, tracking the state-of-the-art in the field and may be used for other research in computer science. It has to be emphasised that most libraries agreed to make the Ground Truth set freely available which will increase the usability of the data set drastically.

The data set is available for the long-term use at the University of Salford, PRIMA research group at <http://www.primaresearch.org/datasets/ENP>.



Fig. 3: Screenshot of the browsing interface of the Image and Ground Truth Repository

Further on, to measure the quality that can actually be achieved on representative examples from the Europeana Newspapers production workflow, as well as to put these results into the context of actual use scenarios, it was necessary to implement and extend a number of software tools for evaluation. For the creation of the evaluation tools, a particular focus was on tools for the production of Ground Truth - which is a major prerequisite for any objective evaluation.

The following tools were developed and/or extended during the project:

- *Aletheia*
Aletheia is specialist software for highly accurate and yet efficient ground-truthing of large amounts of documents. Besides a complete redesign of the User Interface, better following the natural order of tasks related to ground truth production, numerous automated and semi-automated tools were integrated in order to aid the user.
- *PAGE converter and validator*
In order to ensure consistent quality and compliance with the project-internal ground truth production guidelines, a tool which was already used in IMPACT in the context of historical books was extended and developed further in order to satisfy the more complex requirements of newspapers. Again, this meant implementing support for the new PAGE XML format but also to create more complex rules against which ground truth files were checked as part of the automated quality assurance workflow. Another feature which was integrated into this tool is the ability to convert (and save) files from any of the previous PAGE XML formats to the latest version.

- *FineReader and Tesseract integration*
Parameterised command line tools were implemented which allowed running different versions of the ABBYY FineReader OCR-engine, returning recognition results in the latest PAGE XML format, as well as Tesseract, the state-of-the-art Open Source OCR engine.
- *OCR Evaluation*
Due to the limitations of existing software, it was necessary to implement a completely new command line tool for evaluating the text accuracy of OCR engines. The implemented approach follows the ISRI OCR Evaluation Tools, originally developed at the University of Nevada, but adds a number of extra features such as the “Bag of words” performance measure (text accuracy disregarding the order of words), evaluation of both word and character accuracy (with the option to ignore stop-words) and support of all latest file formats (PAGE XML, ALTO, ABBYY XML, hOCR).
- *Layout Evaluation*
The layout evaluation tool used in Europeana Newspapers is an improved version of the software that was originally developed in the IMPACT project. Major extensions were the Introduction of new weights corresponding to the special needs of newspapers (for instance to define the importance of correctly recognising adverts), performance improvements in order to cope with potentially very big ground truth and result files of newspapers (compared to digitised books) and again support of all latest file formats (PAGE XML, ALTO, ABBYY XML, hOCR). The layout evaluation tool comes in two versions – a command line executable for workflow integration and batch processing as well as a stand-alone GUI software for inspection of individual results and in-depth analysis of specific problems.

To avoid gaps in the general evaluation infrastructure and to provide the greatest possible flexibility in terms of how the created resources can be used, a number of smaller auxiliary tools were implemented and further improved. The outcome is a comprehensive collection of software with extensive documentation covering all aspects of performance evaluation related to OCR-workflows for newspapers in particular, which is described in full in deliverable [D3.3 Evaluation Tools](#). All tools are available from the website of the PRIMA research group⁴, with several tools also being released as open source⁵.

The above infrastructure was then used in a detailed evaluation of the results which were obtained from the main Europeana Newspapers OCR production workflow based on a representative dataset collected from all partner libraries in the project. In general it can be concluded that the produced results, especially with regard to the overall text accuracy, are of good quality and fit for use in a number of use scenarios. Moreover, technical decisions that were made during the setup of the production workflow could be confirmed. A number of observations (e.g. on the recognition performance for certain languages and particular layout

⁴ <http://primaresearch.org/tools>

⁵ <https://github.com/PRIMA-Research-Lab>

problems) show mainly the limitations of current state-of-the-art methods rather than issues with the implemented workflow for the project. In particular in terms of layout analysis capabilities there is still room for improvement and any progress in this area could have a great impact on the usefulness of OCR results for more sophisticated use scenarios. A detailed account of the evaluation is provided in deliverable [D3.5 Performance Evaluation Report](#).

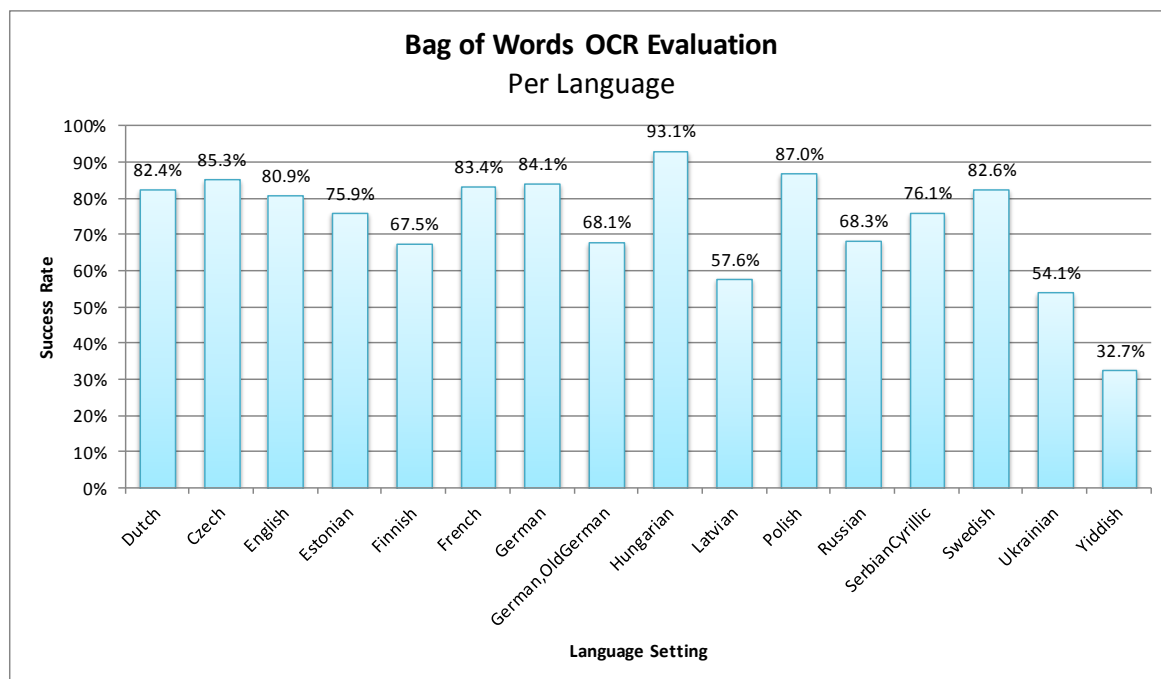


Fig. 4: Overview of the “bag of words” evaluation of OCR results in the project per language

In addition, an approach (complemented with software tools) for estimating the performance of digitisation pipelines was produced and described in deliverable [D3.6 Planning resources and quality estimation toolkit](#). As experimental evidence showed, the quality of digitisation results can be predicted reasonably accurately, using concepts from machine learning, based on only the scanned pages (and with minimal training). Quality estimation can complement pilot projects and be used for purposes of triage (selection of documents to be included in digitisation projects).

6. WP4 – Aggregation and presentation of digitised newspapers

The main objectives of WP4 were the aggregation and indexing of the full-text produced by the refinement, the aggregation of newspaper metadata and its transformation into the Europeana Data Format for ingestion to the Europeana platform and the creation of a feature-rich web interface for searching and exploring the newspaper collection assembled in the project.

To get started on the aggregation and presentation of the digitised newspapers, the project designed a survey on digital newspaper collections in Europe with assistance from LIBER and the KB. It was undertaken until July 2012 and data from more than 40 libraries outside the project consortium was collected. The analysis of the resulting data is completed and the findings contain information on the extent of digitisation of newspapers in major European libraries, as reported in deliverable [D4.1 Europeana Newspapers Survey Report](#).

For the newspaper aggregation and indexing planning, WP4 leader TEL established a realistic schedule for the aggregation of data to ensure smooth workflows for the transferral of images, full-texts and metadata from the numerous data providers. A document “Europeana Newspapers Browser Options” was developed and distributed which gathered information on how content providers in the project wanted to make their data available (images plus full-texts, snippets etc.). The plan was updated and revised at periodic intervals and the final version is available in deliverable [D4.2 Aggregation and Indexing Plan](#).

To create a European registry for digitised newspaper collections, data from the libraries were collected and prepared for ingest into SBB’s Union Catalogue of Serials (Zeitschriftendatenbank/ZDB⁶) which is described in deliverable [D4.3 Report on European Registry for Digital Newspapers Online](#).

In order to serve the highly complex structural demands for digital newspapers, several adaptations of the Europeana Data Model (EDM) were considered and discussed in deliverable [D4.4 Report on EDM for newspapers](#).

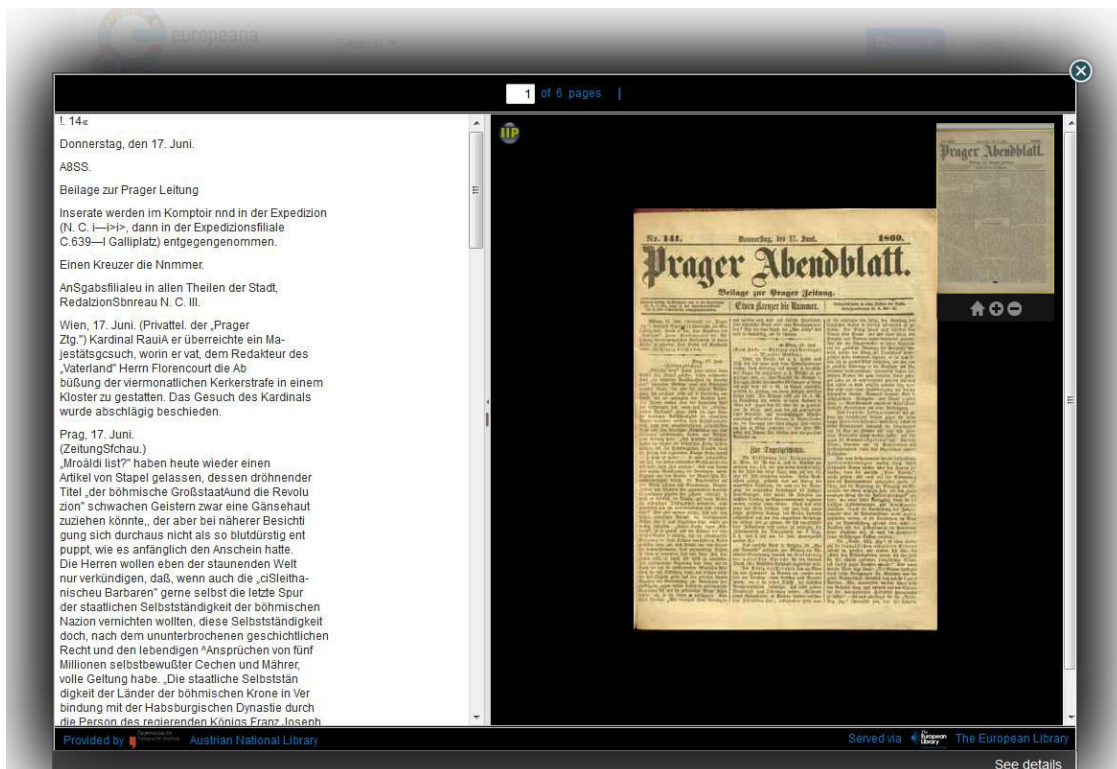
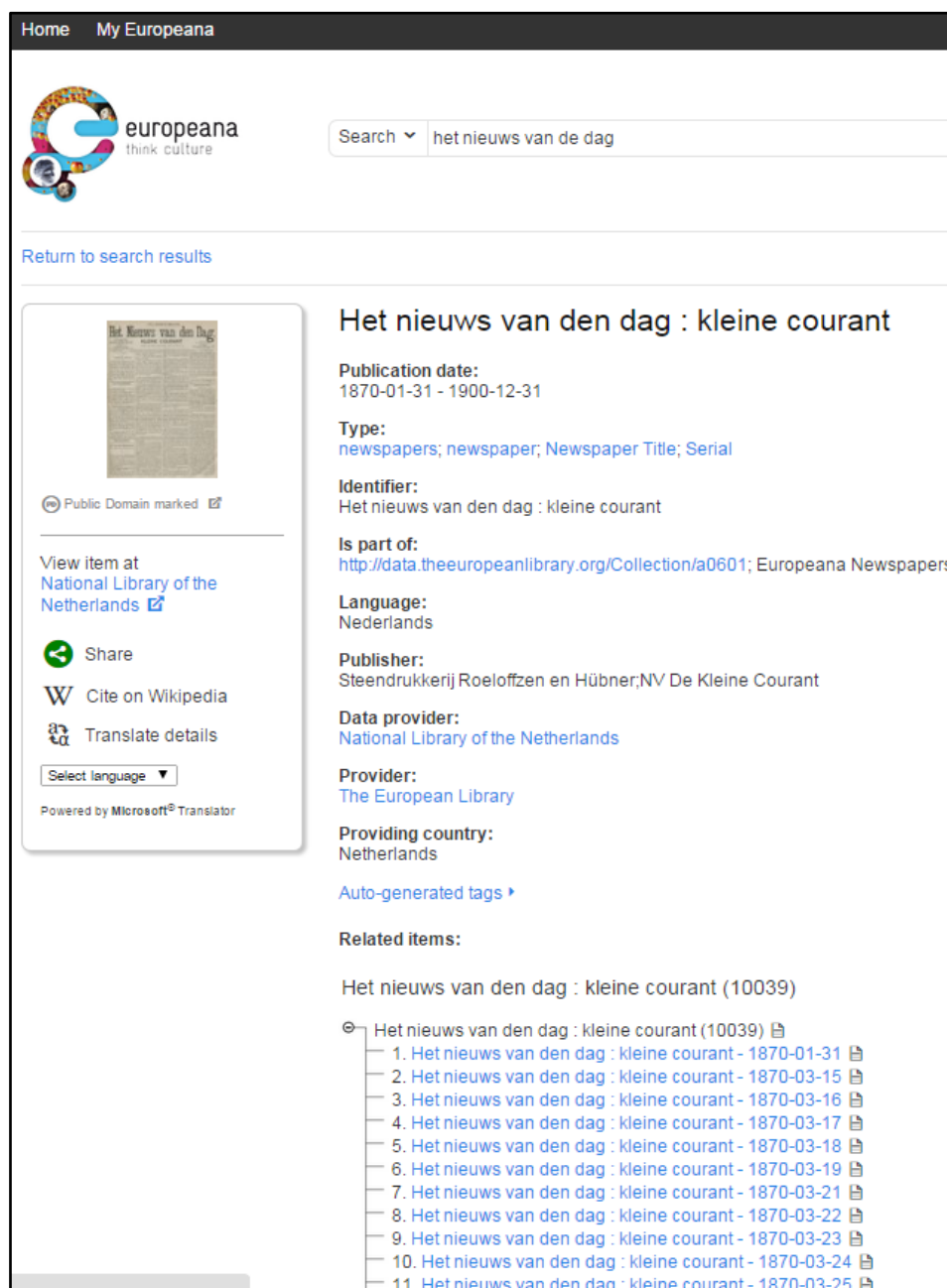


Fig. 5: Presentation of image and full-text directly on Europeana

⁶ <http://www.zeitschriftendatenbank.de/>



The screenshot shows the Europeana website interface. At the top, there are navigation links for 'Home' and 'My Europeana'. The main header features the Europeana logo and a search bar containing the text 'het nieuws van de dag'. Below the search bar, there is a link to 'Return to search results'. The main content area displays a search result for 'Het nieuws van den dag : kleine courant'. On the left side of the result, there is a thumbnail image of a newspaper page, a 'Public Domain marked' icon, and a list of sharing options: 'View item at National Library of the Netherlands', 'Share', 'Cite on Wikipedia', and 'Translate details'. A language selection dropdown is also present. The right side of the result provides detailed metadata: 'Publication date: 1870-01-31 - 1900-12-31', 'Type: newspapers; newspaper; Newspaper Title; Serial', 'Identifier: Het nieuws van den dag : kleine courant', 'Is part of: http://data.theeuropeanlibrary.org/Collection/a0601; Europeana Newspapers', 'Language: Nederlands', 'Publisher: Steendrukkerij Roeloffzen en Hübner; NV De Kleine Courant', 'Data provider: National Library of the Netherlands', 'Provider: The European Library', and 'Providing country: Netherlands'. Below the metadata, there is a section for 'Auto-generated tags' and 'Related items', which lists 'Het nieuws van den dag : kleine courant (10039)' and a hierarchical list of 11 specific newspaper issues with their dates.

Fig. 6: Presentation of newspapers as hierarchical objects on Europeana

Throughout year 2 and year 3 of the project, TEL continued with the aggregation of metadata, full-text and images according to the ingestion plan. Aggregating content was organised either via the technical service providers or directly from content providing libraries. Metadata, full texts and images were aggregated and made accessible via the TEL interface and Europeana. In collecting and distributing over 3.5 million metadata records and over 13 million pages of full text, the project has significantly surpassed the high targets it set for itself. However, it was a complex process with metadata, full text and images all travelling in different directions and at different speeds. The final outcomes are presented in deliverable [D4.5 Report on newspapers aggregated by TEL](#).

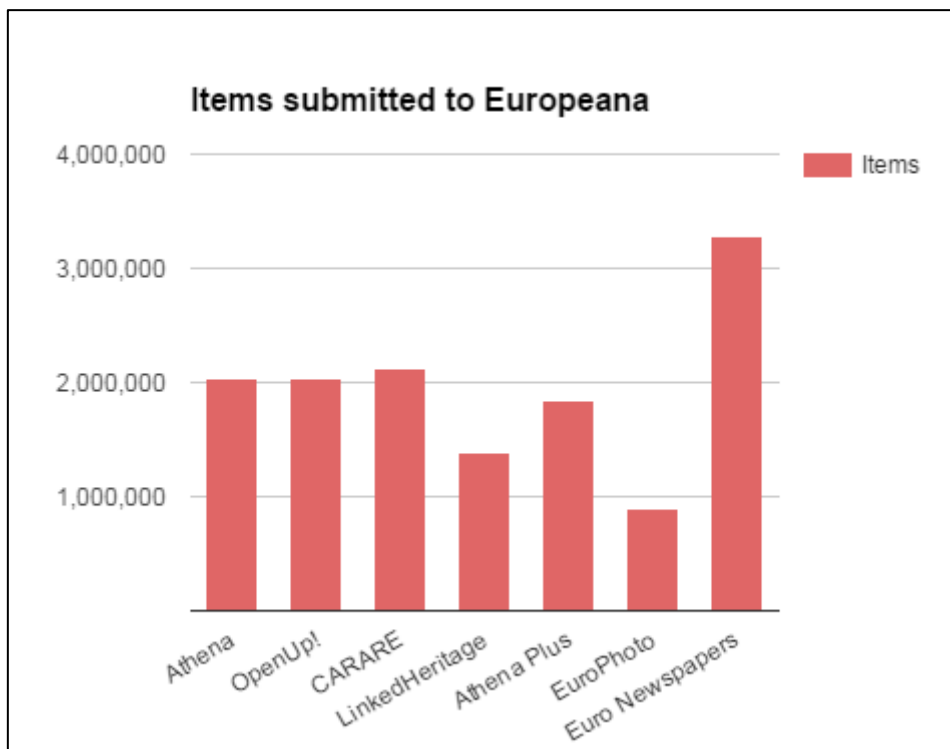


Fig. 7: Largest contribution of items to Europeana for a single project

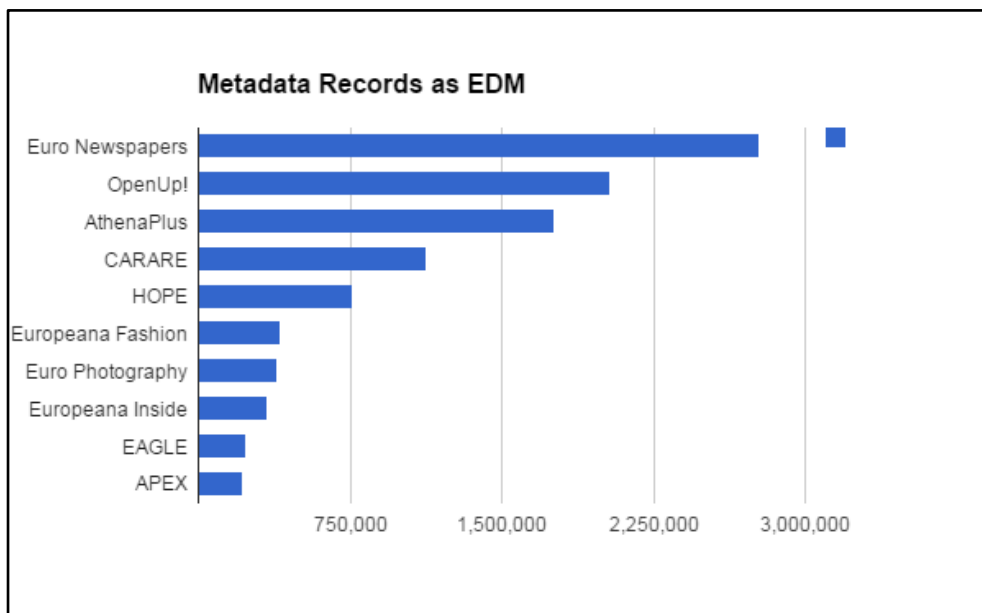
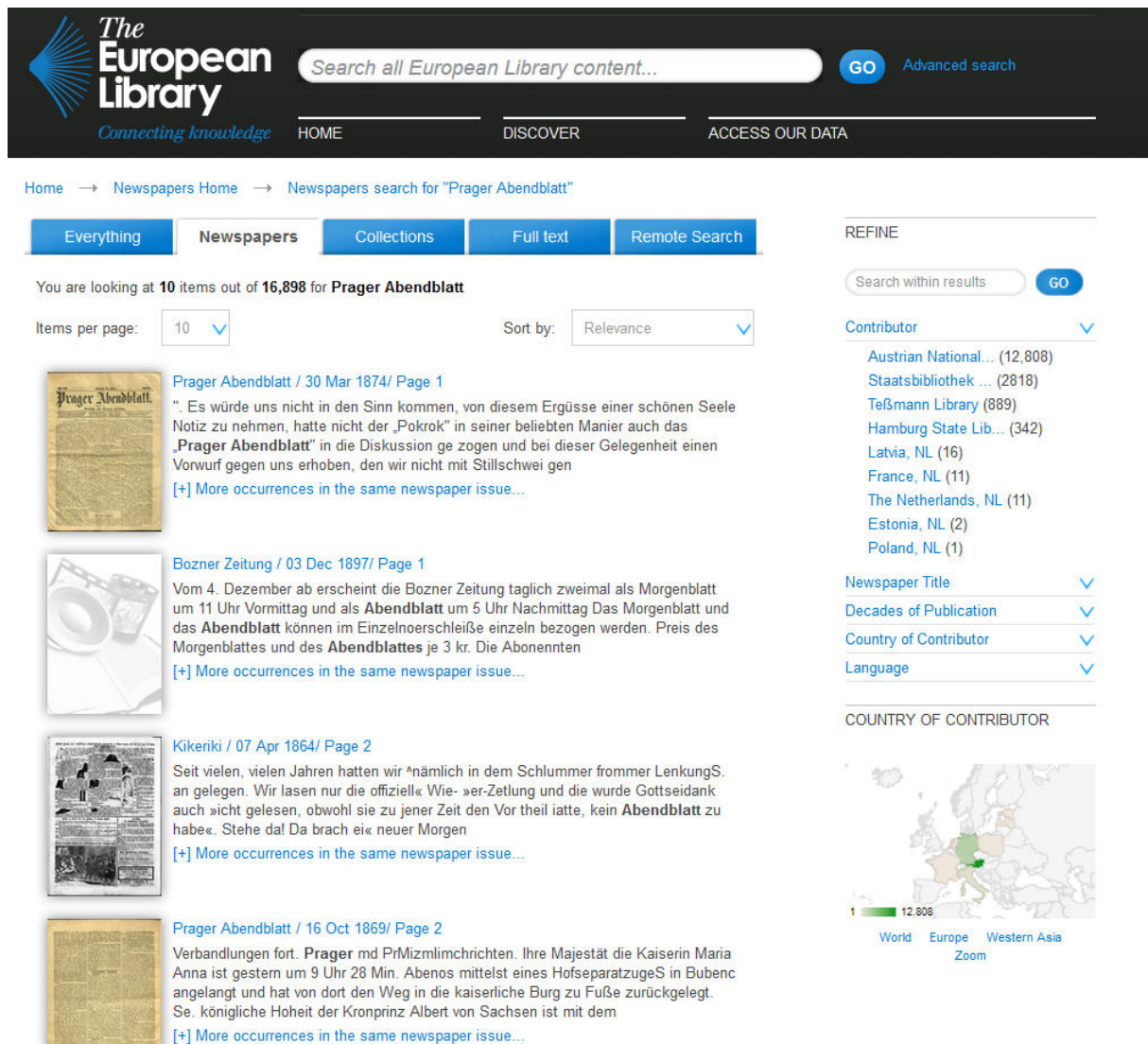


Fig. 8: Largest contribution of metadata records in EDM to Europeana for a single project

With more and more of the content being refined and subsequently added to the web interface, usability testing was performed and changes to the functionality and usability implemented. The final results and usage statistics are available in deliverable [D4.6-7 Report on Indexing, Interface Development and Usability](#).



The screenshot shows the 'The European Library' search interface. At the top, there is a search bar with the text 'Search all European Library content...' and a 'GO' button. Below the search bar are navigation links: HOME, DISCOVER, and ACCESS OUR DATA. The main content area shows search results for 'Prager Abendblatt'. There are tabs for 'Everything', 'Newspapers', 'Collections', 'Full text', and 'Remote Search'. The 'Newspapers' tab is selected. Below the tabs, it says 'You are looking at 10 items out of 16,898 for Prager Abendblatt'. There are filters for 'Items per page' (set to 10) and 'Sort by' (set to Relevance). On the right side, there is a 'REFINE' section with a 'Search within results' button and a 'GO' button. Below this, there are dropdown menus for 'Contributor', 'Newspaper Title', 'Decades of Publication', 'Country of Contributor', and 'Language'. A 'COUNTRY OF CONTRIBUTOR' map is also visible, showing a heatmap of Europe with a legend for 'World', 'Europe', and 'Western Asia'. The search results list several items, each with a thumbnail, title, date, and page number, followed by a brief description and a link to 'More occurrences in the same newspaper issue...'.

Fig. 9: Screenshot of the TEL interface for searching the Europeana Newspapers collection

7. WP5 – Metadata best practice recommendations

In order to serve the specific needs of newspapers and the technical workflows developed within the Europeana Newspapers project, the main objective of WP5 was the design and documentation of a best practice metadata model for digital newspapers.

WP5 leader University of Innsbruck investigated variants of METS/ALTO schemas in use and contacted experts in METS/ALTO (e.g. the British Library, the Library of Congress and the Australian National Library). Several design decisions were based on the results of these investigations and the Europeana Newspapers METS/ALTO Profile (ENMAP) was conceived.

During the second project year the University of Innsbruck integrated the ENMAP schema into the export functions of the OCR production processes of the project. Moreover, first specifications for structural metadata were drafted and examples of such metadata created. Several partners provided feedback on the accuracy of metadata and suggested types and elements for structural metadata on various occasions.

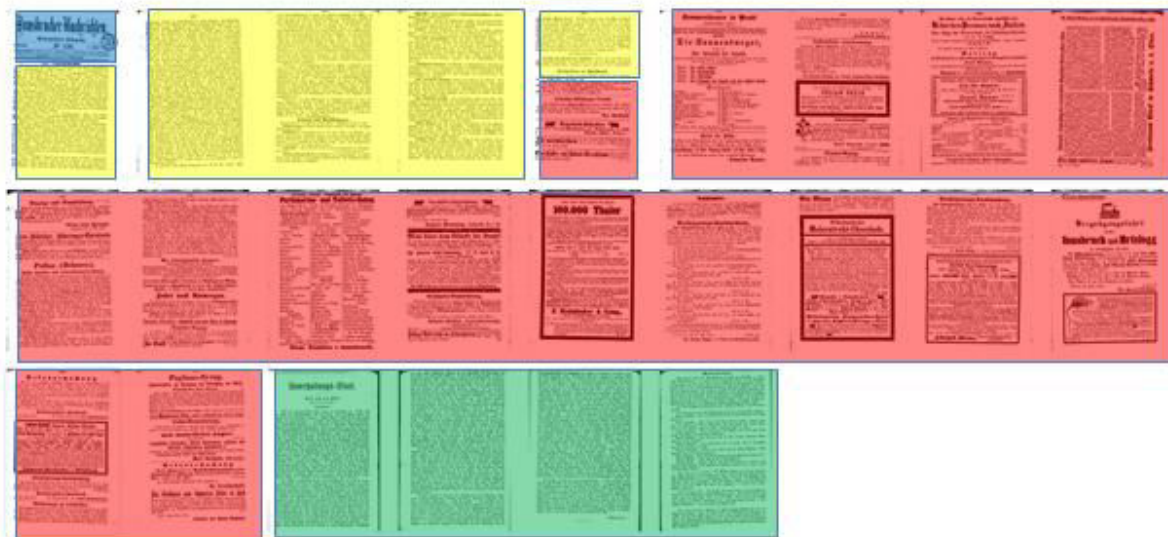


Fig. 10: Newspaper structure disassembled - Metadata (blue), Information (yellow), Advertisements (red) and Entertainment (green)

The newspaper metadata model was then enriched with detailed structural metadata, since this seemed to be one of the most important strategic contributions that can be delivered in this field. The model was introduced to all project partners on several occasions and a feedback cycle with libraries established. A software for exemplifying the structural metadata schema on the ENMAP output format was set up. Partners contributed to the definition of the vocabulary for structural entities and conducted research on existing controlled vocabularies used in published METS profiles and standard formats. As a member of the ALTO Board, partner CCS was successful in convincing the ALTO Board members to support the requirements of the Europeana Newspapers project in the upcoming ALTO version 2.1.

In the third and final project year, ENMAP was finalised and a software tool produced that not only allows the presentation and production of ENMAP, but also provides a set of rich features for a deep structuring of digital newspapers. This platform independent tool, STRUCTIFY⁷, was also made available as open source and continuously updated.

⁷ <http://dbis-halvar.uibk.ac.at/dokuwiki/doku.php?id=main:structify>

The final ENMAP model including major conceptual advances in the deep structuring of digital newspapers and according examples has been released in the deliverable [D5.3 Final Public Release of ENMAP with updated online resource for documentation](#).

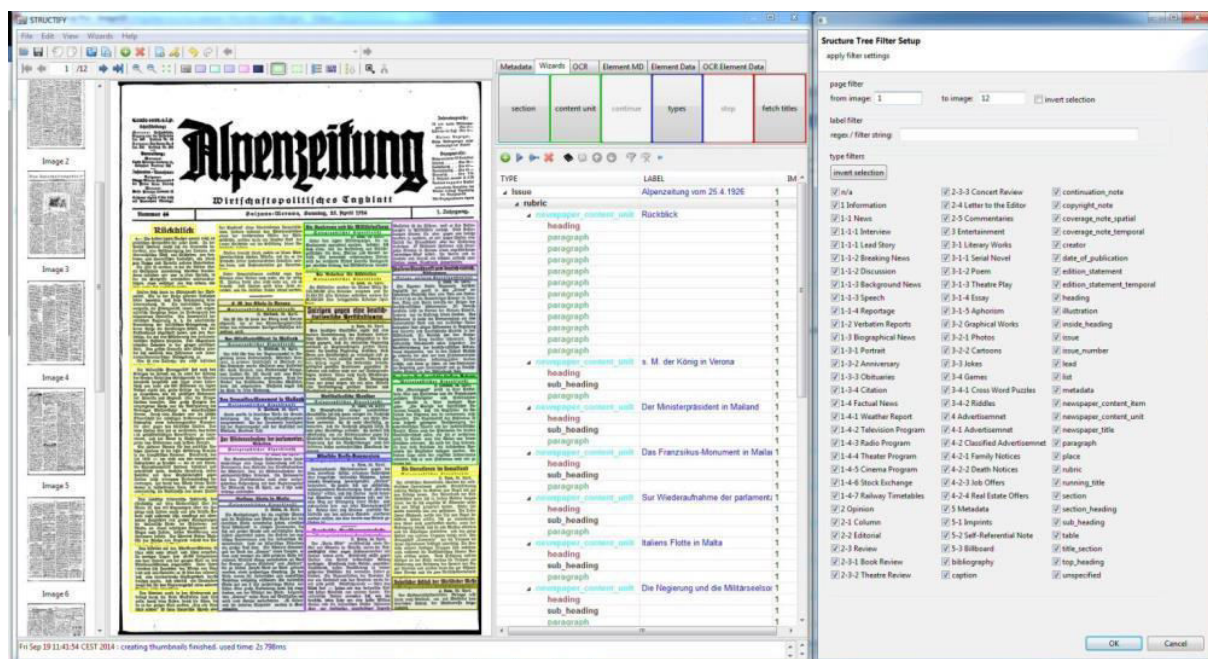


Fig. 11: Screenshot of the STRUCTIFY tool

8. WP6 - Dissemination and exploitation

The main purpose of WP6 was to engage all stakeholders of the newspaper community and make sure they are aware of the efforts of the project and that their interests are considered in the developments of the project.

For outreach and dissemination, the project website at www.europeana-newspapers.eu was the main channel. The website provides general information on the project, the consortium members and associated partners, news related to the project, activities within the project and events as well as contact details of the project coordinator. Additionally, public materials - deliverables and dissemination materials - can be downloaded.

Also social media like [Facebook](#), [Twitter](#), [LinkedIn](#), [Flickr](#), [Vimeo](#), [Youtube](#) and [Slideshare](#) were utilised to great success in supporting the dissemination of the project and its outcomes. Numerous [promotional materials](#) were created and can be downloaded from the project website.

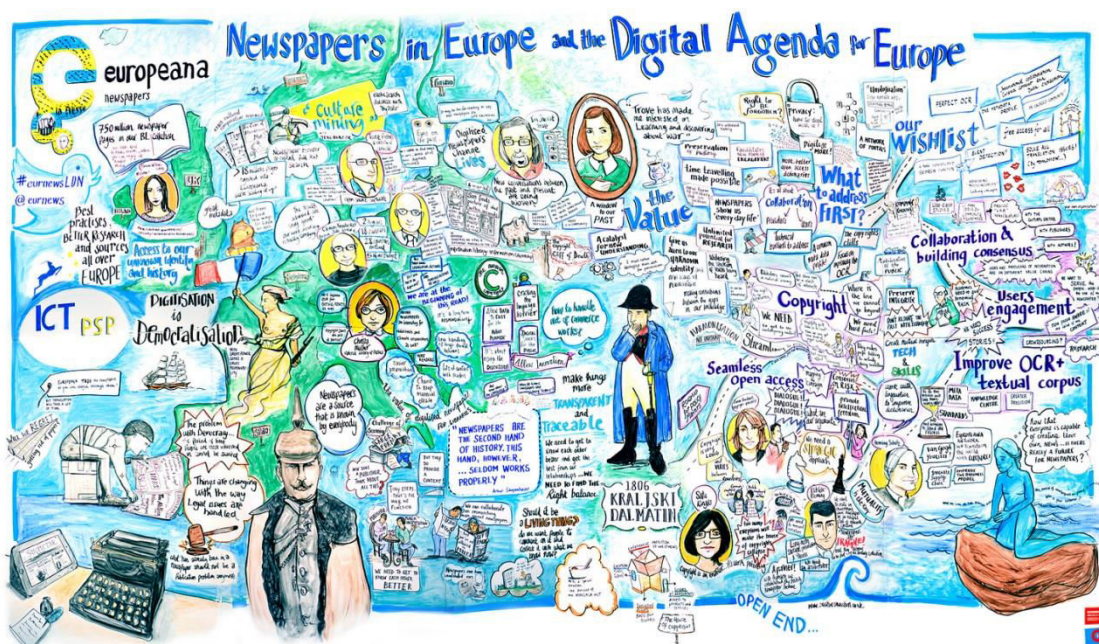
Several [short movies](#) and [animations](#) were produced as well as an 18 minute documentary-style film in cooperation with the [Europeana Collections 1914-1918 project](#).



Fig. 12: A digital commemoration of the First World War – documentary movie ([watch on Youtube](#))

The Europeana Newspapers project also organised three major workshops:

1. [Workshop on “Refinement and Quality Assessment”](#)
June 2013 in Belgrade, Serbia
2. [Workshop on “Aggregation and Presentation”](#)
September 2013 in Amsterdam, Netherlands
3. [Final Workshop on “Newspapers in Europe and the Digital Agenda for Europe”](#)
September 2014 in London, United Kingdom



In addition, a series of highly successful information days were organised by library participants in every country, and in their local language, to make sure the project and its activities are also well embedded in the national newspaper and digitisation communities:

- [Turkish Information Day](#)
- [Latvian Information Day](#)
- [German Information Day](#)
- [Polish Information Day](#)
- [English Information Day](#)
- [Italian Information Day](#)
- [Dutch Information Day](#)
- [Estonian Information Day](#)
- [Austrian Information Day](#)
- [French Information Day](#)

9. Concluding remarks

At the end of the project's duration all full and associated partners deem the project to be an overwhelming success. The number of refined newspaper pages as foreseen originally was exceeded and all project objectives were successfully reached.

Some additional work was done that was not planned. For example, the associated project partners provided metadata on the scale of almost 10 million pages to Europeana, thereby nearly matching the contribution of full project partners. Strong interest in further contributions to the Europeana Newspapers collection has also been expressed by others such as the National Libraries of Norway, Denmark and the Faroese Islands.

The overall amount of digitised newspapers to be accessed via Europeana has been increased dramatically. Europeana Newspapers has not only produced the largest contribution of records to Europeana by any single project, it has also contributed the largest ingest of data in the Europeana Data Model, thanks to highly efficient aggregation and normalisation workflows at The European Library. In terms of volume and functionality, the project output is on par with other major international newspapers digitisation programmes such as *Chronicling America*⁸ and *Trove*⁹ in Australia and has put Europe on the global map of leading providers of digitised newspapers.

In addition to that, the project made important improvements to the Europeana user experience possible, such as the presentation of images and full-text directly on the Europeana platform and the browsing of hierarchical objects.

For the first time OCR workflows for newspapers were implemented on such a large scale and across a dozen European countries and languages. Experiences gathered here will be extremely valuable for future large scale refinement projects and for further aggregation of these materials for Europeana. Important work on structural metadata was undertaken and generally received in a very positive way by expert audiences.

The series of national information days proved to be a most valuable form of spreading the project's results across Europe, particularly so, since project partners did an exceptionally good job in tying these information days to ongoing national or regional activities in the area of newspaper digitisation.

All partners are most pleased with the massive interest that the outcomes created already during the project lifetime and among different user groups and are convinced that the project's core mission, the provision of vast amounts of digital European historic newspapers to the public, is and will be of paramount importance to a wide number of stakeholders.

⁸ <http://chroniclingamerica.loc.gov/>

⁹ <http://trove.nla.gov.au/>