

Improving the discovery of European historic newspapers

Rossitza Atanassova

Digital Scholarship, British Library, London, United Kingdom.
E-mail address: rossitza.atanassova@bl.uk



Copyright © 2014 by Rossitza Atanassova. This work is made available under the terms of the Creative Commons Attribution 3.0 Unported License: <http://creativecommons.org/licenses/by/3.0/>

Abstract:

The Europeana Newspapers Project is about improving access to digitised historic newspapers through the use of refinement techniques and the launch of a browsing tool. Twenty three libraries across Europe are contributing content and by the end of the project in January 2015 users will be able to search ten million pages of historic newspapers and millions of metadata records. Named entities and structural elements of the newspaper layout will have also been identified in two million of the pages.

The focus of this paper is the user response to the Europeana Newspapers [browser](#) developed by The European Library, which is currently available as prototype. In April the tool will be tested by several user groups and the results will help assess its functionality and how well it meets users' needs and expectations. The feedback from the user testing will inform the release of an improved beta version of the browser later this year.

Taking into account the results from the user testing, the paper will reflect on the user experience of discovering content via the historic newspapers browser on The European Library website. In particular, it will consider the user response to the creation of a specifically European interface for historic newspapers, assess how successful the tool is for searching across European newspaper content and the implications it may have for the users' investigative practices.

To conclude, the paper will emphasise the value of the resource created by Europeana Newspapers and its intended users. Researchers are one of the main user groups for digitised European historic newspapers and some of their research interests have been highlighted in a number of interviews published on the Project [blog](#). Furthermore, the dataset of European newspapers creates opportunities for new discoveries for digital humanities scholars.

Keywords: digitised newspapers, interface, discovery, usability testing, research

Improving the discovery of European Historic Newspapers

Introduction to Europeana Newspapers

Europeana Newspapers is an ambitious EC-funded Best Practice Network, which is improving access to digitised historic newspapers through the refinement and aggregation of content from 22 European countries.ⁱ The project is creating 10 million pages of full text, through Optical Character Recognition, of existing digitised images from 12 full partner libraries. Two million of these pages are also being processed with Optical Layout Recognition software to create article-level records, and with Named Entities Recognition software to tag persons, places and organisations names.

Altogether the project will make up to 18 million historic newspaper pages from across Europe searchable. In addition, it aims to aggregate the metadata of a further 19 million newspaper pages held by 11 associate partners.

All of this content - the full text, images and metadata – will be made accessible via a search interface developed and hosted by The European Library. The interface currently allows anyone to search content from nine countries and provides access to full text for over two million newspaper pages and to metadata for over one million newspaper issues. The remaining content will be added to the site at the beginning of 2015 and issue-level metadata will also be available via Europeana.ⁱⁱ

As a Best Practice Network Europeana, Newspapers also builds and shares quality-assessment tools and metadata standards for digitised newspapers, and engages actively with researchers and the wider user community. The project website, maintained by LIBER, and social media channels are used to disseminate project activities and deliverables, including a recently published video, thematic articles, content highlights and interviews with researchers and people involved with other digitised newspaper archives.

The Europeana Newspapers interface

As part of this project, The European Library is delivering and hosting a cross-searchable newspapers interface for European historic newspapers.ⁱⁱⁱ This browser has been available as a prototype since January 2014 and will undergo several iterations before the final interface is released in November 2014.

The current prototype version of the project interface incorporates the basic search and browse functionalities associated with digitised newspapers sites. Users can search by keyword, over the full text or the newspaper titles. They can also refine their search by content provider, language and a set of years of publication, or browse by newspaper title, date and contributing country. The search results display issue-level metadata, and for 11 project partners, show the full-text and newspaper page image, with links to view the original in the source library if appropriate. Since the interface provides access to multi-national content, the aggregation and browser design have taken into account any restrictions imposed by contributing libraries due to national copyright laws and the libraries' business models for digitised newspapers. Therefore, the contributing partners were presented with four options to help them decide how their content should appear on the newspapers interface:

- Option 1 – metadata, full-text and full, zoomable images

- Option 2 – metadata, full-text and static images – either full size or snippets
- Option 3 – metadata and full-text only
- Option 4 – metadata only

The majority of libraries have accepted Option 1 and allow display of full-size zoomable images that are either ingested by The European Library or delivered directly from the library's own image server. With a more consistent user experience and better search functionality in mind, The European Library has been able to convince some partners to offer a static full size image rather than a snippet view, with some also considering zoomable images.

In order to inform the further development of the prototype interface, The European Library arranged for two rounds of usability testing this year. The first took place in April 2014 and the second is scheduled for September 2014. The main objectives of the tests are to understand the needs and expectations of users of digitised historic newspapers, to evaluate their experience of a pan-European newspapers interface and to recommend any changes that would improve the site functionality and design.

Feedback from the first usability test

Twelve participants from five countries (UK, Latvia, Austria, Italy and Finland) represented in the project took part in 60-minute long remote online test sessions that were conducted by UserVision, a company of independent usability experts based in Edinburgh, Scotland. The user group for the test were people with a professional or personal research interest in historic newspaper collections. Each participant joined an online meeting with a moderator from UserVision and was asked to complete six pre-determined tasks using the newspaper interface on The European Library portal. The scenarios were typical and included an exploration of the landing page, the performance of a basic search for a place name, a refinement of the search results by country of publication, a search by date, by title and by region. The consultants observed the participants' performance and offered assistance with the tasks where necessary. Before, during and after the sessions participants were asked questions to determine their expectations, identify and explore areas of concern and formulate recommendations as to how issues they encountered can be addressed.

The user feedback from the usability test was broadly positive. Participants reacted well to the layout and functionalities of the site. Their initial expectations of the Europeana Newspapers interface were high due to the quality of The European Library website and the scope of the content made available by the project. Because of this and few problems encountered during the test, the overall rating for participants' expectations of the site was "slightly above expectation".^{iv}

“Strong positive reaction to the availability of the archive”

The findings of the usability test report that an “aggregated view of content from many sources” is unquestionably of value to users with research interest in historic newspapers. They clearly welcome the opportunity to cross-search full text and issue-level metadata from 22 European countries. In the post-questionnaire interviews, participants stressed further the high value for them of “the breadth and depth” of newspaper content provided on the browser.

Furthermore, user feedback confirmed that the basic search functions on the site worked well and that the browse and facet search options were mostly effective. The presentation of the images was also appreciated. The rotation and zoom controls offered by the IIPImage Viewer were familiar to users and easy to operate, and the display of the OCRed text next to the image also met users' expectations.

The feedback from the tests helped identify and recommend actions about a few functionality and design issues, many of which have since been addressed by The European Library.

Browse options are now clearer and better integrated

The user feedback highlighted the value of the browse and advanced options on the landing page, and made recommendations for making these functionalities more prominent, intuitive and better integrated on the site. Accordingly, a new 'discover' tab, as well as 'explore by' links from the default search landing page, have been introduced to take users to the browse options. The 'browse by title' option has been also modified to present users with an alphabetical index for all available newspaper titles and thus help them browse and select relevant titles more easily and quickly. The previously overlooked 'further search options' feature is now labelled and signalled more clearly to invite users to 'filter by library, date and title'.

The hugely popular 'browse by country' option in the form of a geographical map of the site's content has also been modified in response to user feedback. The current default view of the map of Europe is better in line with the site's content and more effective to use due to the increased size and the choice of a better colour palette to indicate the number of issues available for each country. There is a further recommendation to provide pop-over boxes for all countries on the map, even if they do not contribute content to the site.

There are two features in particular that seemed to need more urgent attention in order to increase their value for users. The first one is the presentation of the 'browse by issue date' option which appeared unclear and cumbersome to use. The option has been modified to include a text input box for entering the year with auto-suggestions and to highlight the order of selection of months and dates for which content is available. The second round of usability testing will need to verify how successful these changes have been.

The 'this day in history' feature was originally presented in the form of a carousel which highlighted only a small selection of newspapers. The redesigned interface incorporates more clearly this interactive option which allows users to scroll through all relevant issues. This is an important element of the interface design, since it enhances the site visually with newspaper images that refresh regularly.

Requirements for enhanced search results functionality

Participants in the usability testing confirmed that search results displayed well and the information provided in the link, the short description and image thumbnail helped them assess easily the relevance of the result.

At the time of the usability test, however, the order of search results was not configurable and prevented participants from understanding the order of results and exercising control over it. The updated interface displays results by relevance and also allows users to sort by ascending

and descending date and configure the number of results displayed on a webpage. Another implemented change is the improved management of filters through the addition of an 'x' icon next to a selected filter.

The full-screen mode option relating to the newspaper image and full text pane presentation is an important enhancement and allows users to see complete lines of text without the need of scrolling. This addresses the navigation difficulties experienced during the usability testing, due to the restricted horizontal width of the content pane window.

Further changes relating to search results are recommended to be implemented to the 'browser' interface before its final release in November. These include the addition of navigation controls to allow back and forth movement between search results, and a 'back to search results' button control.

The usability test also strongly recommended the addition of a search input box on the results page "populated with the original search term(s) entered", to allow users to modify their search terms, or perform a new search, without having to go back to the landing page and thus resetting the already selected filters.

The lack of an option to download or save locally images and associated metadata was seen as an obvious shortcoming of the interface and also needs to be implemented. The development of this option, and a print option, will also need to meet the contributing libraries expectations.

User research practices and expectations of the Europeana Newspapers interface

Judging from the participants' performance of the tasks set in the usability testing and the feedback from the interviews, it seems their preferred method of using the historic newspapers archive is through controlled search options rather than through browsing. This is explained by the difficulty of browsing such a vast newspaper archive and by the participants' well established search strategies.

This particular group of users would like to see more advanced search options or facets to help them filter through and manipulate the search results. Such functionalities would be consistent with the advanced search options (Boolean search, article type facets, multiple layers of filters) implemented on other interfaces to digitised historic newspapers, such as the British Newspaper Archive, Chronicling America and Trove.^v Further options that participants mentioned include searching by 'subject area' and 'historical period' which points to a very specific rather than a general interest in the site.

In addition, the participants in the test expected to be able to create a user account offering a research space to save search histories. They would also like to be notified of newly published content on the site and to have the option to submit feedback. As already mentioned their research needs require the ability to download a local copy of an image, text or metadata for the purpose of building their own personal archives.

A unique value of the Europeana Newspapers interface is that it offers cross-searchability of content published in over 22 languages, but this could also present a barrier for users. They will need to know different languages to be able to make the most of the site. Even if one can confidently perform a basic search across all content, a deeper knowledge of languages is

necessary to interpret the results. In the first round of usability testing participants did not comment on this as a problem, as they mainly searched content in their own language and from their own country. However, it has been suggested by researchers and librarians that the interface should embed tools to assist users with translation. This is a challenge for open-source tools, such as Google Translate, which will not cope with translation of the type of content found on the Europeana Newspapers interface.

Researchers are a primary target user group for Europeana Newspapers

As confirmed by the first round of usability testing, there is a great demand for making digitised historic newspapers available and it is no surprise that the academic research community has received the Europeana Newspapers project with great enthusiasm. LIBER, who lead the promotional work on the project, and other partners, including the British Library, have been actively engaging this user group to promote the resource and to better understand what use researchers will make of the archive, and in particular the new research possibilities that this content opens up.

A series of Q&A interviews with newspaper researchers from different European countries is published on the project website and highlights some of the ideas researchers have for mining the content on the Europeana Newspapers interface. The researchers interviewed turn to newspapers to study a range of subjects and topics: 19th century popular culture and humour, history, literature, evolution of language, public discourse, reference cultures and professional careers. For them such an aggregation of millions of pages of European newspapers offers exciting new opportunities for “transnational comparative research” and computational analysis of the data. One of the interviewees, Professor Toine Pieters of Utrecht, sees the multilingualism of the archive not as a barrier but rather as a challenge that needs to be overcome and is already being addressed by a project which will explore reference cultures in Europe with the help of multi-lingual text mining techniques.^{vi}

The Europeana Newspapers Information Days, organised by project partners, have been another vehicle for engaging the researcher community. So far five Information Days have taken place in Turkey, Latvia, Poland, Germany and the UK and three more events are planned for later this year in Italy, France and Estonia. The UK Information Day held at the British Library in June 2014 brought together researchers in the fields of history, literature, print culture, media and social science, as well as digital humanities scholars.^{vii} This latter group is particularly excited about the possibilities of exploiting in news ways historical newspaper content and applying digital humanities research methods to the data aggregated by Europeana Newspapers.

Newspapers contain a plethora of illustrations, maps, photographs, and one idea would be to extract through algorithm the illustrations found in the newspaper pages in the archive and invite users to tag, organise thematically or link back to captions and descriptions found in the newspapers. The inspiration for this idea comes from the British Library Labs project which extracted one million images from 65,000 digitised 19th century books and released them on Flickr Commons for users to tag and re-use. This project enabled users to describe the images, create thematic albums, re-use creatively for commercial and educational purposes and use them for research in the areas of image recognition and automated classification of historical images.^{viii} The new ‘Victorian Meme Machine’ project conducted by Bob Nicholson of Edge Hill University and supported by BL Labs project, is creating a database of Victorian jokes extracted from digitised 19th century British newspapers and will

semi-automatically pair them with appropriate images from the Library's digital collections, to create new context and re-use for these Victorian jokes.^{ix}

Such innovative approaches to digitised historic newspapers could be applied to the corpus created by Europeana Newspapers and would help attract new professional and amateur audiences and engage the wider user community. The value of Europeana Newspapers is not limited to academic researchers, but also genealogists, local historians, the teaching and learning community and all citizens of Europe and outside Europe. The content aggregated by the project would be a valuable resource for the study of European history, society, culture, languages, publishing, literature, art, design and much more.

Next steps with the Europeana Newspapers interface

The prototype 'browser' for Europeana Newspapers will be developed further following the second round of usability testing in September. If time and resource permits, The European Library aspires to implement additional features, such as an option for users to correct the OCR'd text. This functionality has been successfully implemented by other newspaper archives, such as the British Newspaper Archive and Trove, and there are many good reasons for why it should be added on the Europeana Newspapers interface. The ability to edit and even tag articles would be appreciated by many user groups and will both improve the level of text accuracy in the corpus and increase user engagement with this digital archive. The European Library is also looking into the possibility of creating an API (Application Programming Interface) to provide access to large sets of data via machine harvesting and analysis.

The complexity of providing access to digitised historic newspapers is not unique to Europeana Newspapers, but the challenges are augmented by the size and range of the dataset involved. To create a good online experience, the project interface has to articulate well the representation and characteristics of the content, what is available and what is missing, and manage the expectations with regard to the quality of the images and full text. Some of these challenges are shared and addressed by other digitised historical newspaper interfaces^x, whilst other are more specific to the Europeana Newspapers project and affected by political, economic and legal policy issues.^{xi} The development of the project interface will also need to be sustainable after the project ends in January 2015. Many of the challenges and the opportunities to improving access to digitised historic newspapers will be discussed at the project's final public workshop, entitled 'Newspapers in Europe and the Digital Agenda in Europe', to be held at the British Library on 29-30 September 2014.^{xii}

ⁱ <http://www.europeana-newspapers.eu/>

ⁱⁱ Dunning, A. and Muhr, M. 2013. *Newspaper Aggregation and Indexing Plan*, http://www.europeana-newspapers.eu/wp-content/uploads/2012/04/D4-2_Aggegration_and_Indexing_Plan_V2.pdf

ⁱⁱⁱ <http://www.theeuropeanlibrary.org/tel4/newspapers>

^{iv} Blackwood, A. 2014. *The European Library: Newspaper Archive - Usability Testing*, <http://www.europeana-newspapers.eu/wp-content/uploads/2014/05/The-European-Library-Newspaper-Archive-Usability-testing-Report-April-2014.pdf>

^v <http://www.britishnewspaperarchive.co.uk/search/advanced>

^{vi} <http://www.europeana-newspapers.eu/qa-with-newspapers-researchers-toine-pieters/> and <http://asymenc.eu/>

^{vii} <http://www.europeana-newspapers.eu/enabling-access-to-digitised-historic-newspapers-june-9th-london/>

^{viii} For more information about the release of the images, see <http://britishlibrary.typepad.co.uk/digital-scholarship/2013/12/a-million-first-steps.html> For examples of a creative re-use of the images, see <http://blpublicdomain.wikispaces.com/>

^{ix} <https://www.youtube.com/watch?v=FN1ZSAz2vMg> and <http://labs.bl.uk/>

^x Diving into newspaper archives: Chronicling America. Interview with Deborah Thomas, <http://www.europeana-newspapers.eu/diving-into-newspaper-archives-chronicling-america/>

^{xi} Dunning, A and Neudecker, C. 2014 *Representation and Absence in Digital Resources: The Case of Europeana Newspapers* <http://dharchive.org/paper/DH2014/Paper-773.xml>

^{xii} <http://www.europeana-newspapers.eu/agenda-final-workshop/>