



Building Bridges: from Europeana Libraries to Europeana Newspapers

Susan K. Reilly
LIBER
The Hague, Netherlands
E-mail: susan.reilly@kb.nl

Session: **119 – Users and portals: digital newspapers, usability, and genealogy – Newspapers Section with Genealogy and Local History Section**

Abstract:

Studies show that ease of access, and particularly the one-stop shop approach, are favoured by researchers as a clean and efficient way to accessing digital content. As well as ease of access, quality-assured content is of prime importance.

The Europeana Libraries¹ project is addressing both of these issues by selecting 5.1 million images, books, videos and theses and articles directly from 19 of Europe's leading research libraries. The source of the data means that confidence can be placed in the metadata and the quality of the imaging, while ease of access will be guaranteed through a single search of all objects on The European Library² and Europeana³ websites.

This partnership has laid the foundation for further collaboration and innovation. Already, the sustainable aggregation infrastructure and full-text search capabilities created through the Europeana Libraries project are set to be applied to a new body of content through the Europeana Newspapers⁴ project. This project will make 29 million pages of newspaper content across Europe available through The European Library and Europeana platforms.

¹ <http://www.europeana-libraries.eu>

² <http://www.theeuropeanlibrary.org>

³ <http://www.europeana.eu>

⁴ <http://www.europeana-newspapers.eu/>

Introduction:

Studies show that ease of access, and particularly the one-stop shop approach, are favoured by researchers as a clean and efficient way to accessing digital content. As well as ease of access, quality-assured content is of prime importance.

This paper outlines some of the work of the Europeana Libraries⁵ project, which will make national and research library full text content searchable through a single portal. It illustrates the motivations for, and benefits, of collaboration across organisations to achieve a common vision. It also outlines the motivations behind, and what will be achieved, in the Europeana Newspapers project, a project which benefits from the portal and full text search capabilities developed through Europeana Libraries.

The Foundation Stone:

The Europeana Libraries project is a two year project which began in January 2011. The idea for the project came on foot of the identification of a need to have a single aggregator for European research libraries, both national and university. Such libraries had worked together in the past to provide thematic content to Europeana, the Europeana cultural heritage portal, through the Europeana Travel⁶ project. Within this project content on the theme of travel and tourism was aggregated through two separate aggregators, one for national libraries and another aggregator established especially to aggregate content from the other libraries in the project. The project was highly successful both in terms of supplying high quality digital content to Europeana, and in terms of establishing collaboration between national and other research libraries, but it did raise a question mark over the sustainability of the use of two separate aggregators.

The Europeana Libraries project addresses this issue of sustainability by opening up the national library aggregation service, The European Library, to research libraries. It uses this service to aggregate a critical mass of valuable content from European research libraries. By the end of the project in December 2012 over 5.1 million objects, including 1,200 film and video clips, 850,000 images and 4.3 million texts (books, journal articles, theses, letters) will have been ingested from research libraries. Much of this content is full text and of particular value to researchers. To maximise on the potential of this content, the project also sets out to develop full text search capabilities and a search portal that provides tool specific to research.

The Partnership:

Bringing together content from research and national libraries also facilitated the coming together of key European library networks, namely CENL, CERL, and LIBER.

The Conference of European National Libraries (CENL) represents 48 national libraries and currently owns The European Library aggregator, which is the only

⁵ <http://www.europeana-libraries.eu>

⁶ <http://www.europeanatravel.eu/>

European library domain aggregator. Up until now it has only aggregated content from national libraries.

The Consortium of European Research Libraries' focus is improving access to and exploiting European printed heritage and has 33 full members from research and national libraries. CERL has a particular interest and expertise in indexing and metadata.

LIBER, the association for European Research Libraries has over 420 members (national, university and other research libraries) from Europe and its borders. 19 LIBER members provide the content for the project and, ultimately, the service developed within Europeana Libraries will be extended to all of LIBER's members.

Europeana Libraries is the first opportunity that these three organisations have used to work together over one very strong commonality, and that is that their member institutions all have content that is valuable research material and all want to make their content accessible and usable for the research community. The sustainability of the project outcomes will be ensured through the exercise of dissecting this commonality into actual and agreed value propositions and cementing the relationship between the networks.

Defining the Value of Partnership:

The projects value not only lies in the creation of a single aggregation service for libraries, although this is a significant aspect, it also lies in the potential it offers to bring research content from libraries to researchers world wide. Potentially, it extends the reach of the collections of both national and research libraries beyond the boundaries of their established research communities and regions. It exploits the collective reputation of libraries as trusted providers of quality information and good metadata. It is a well established fact that libraries are positively associated with books.⁷ Providing the full text content of digitized book collections alongside other digital content such as images, videos and audio files, not to mention scholarly content such as articles and theses, means that researchers can obtain richer search results. Through augmenting the visibility of such content in this way libraries can increase the impact of the significant investment they make in digitization.⁸

For researcher the value lies in being able to access and search a critical mass of cultural heritage and related research content in one place. Studies show that researchers are carrying out more complex research and also have less time for their research activities⁹, hence the one stop shop is an attractive proposition.

⁷ De Rosa, Cathy et al. 2005. Perceptions of Libraries and Information Resources: A Report to the OCLC Membership. Dublin, Ohio: OCLC. <<http://www.oclc.org/reports/2005perceptions.htm>>.

⁸ European Commission, Maurice Lévy, Elisabeth Niggemann, and Jacques de Decker. The New Renaissance. Brussels: European Commission, 2011.
http://ec.europa.eu/information_society/activities/digital_libraries/doc/reflection_group/final_report_%20cds.pdf

⁹ Bulger, Monica, et al., Reinventing Research? Information Practices in the Humanities. London: Research Information Network, 2011.

Designing the Portal:

Most of the content aggregated through the Europeana Libraries project will be available through the Europeana portal but, considering the nature of the content being aggregated, The European Library portal was redeveloped with the humanities researcher as the end user in mind. Such a proposition is particularly relevant to the humanities research community, for whom the definition of research data is complex: *“The humanities community needs a critical mass of digital resources and needs common tools, services, and repositories if they are to move beyond ‘boutique projects’ to a solid foundation of theory and method.”*¹⁰

Several rounds of workshops and end user testing, as well as significant desk research, has fed into the design of the new portal which now features:

- the ability to search full text content
- the opportunity to inspect the raw metadata record of individual objects with the aim of eventually enabling access to large datasets for research purposes.
- the increasing widely used CERIF subject headings. This makes research possible across a corpus of objects, which are linked by a common theme or timeframe.
- Pan-European collection development in cooperation with the national, research and university libraries of Europe. This is an extension of current virtual exhibitions¹¹, which display content from a range of sources across Europe.
- Timelines showing the occurrence of a particular search term through the centuries.
- APIs which will allow for the content of the database to be analysed and displayed in contexts outside of The European Library portal. This means that researchers can bring their content into their own research environments and explore new ways of exploiting the content.
- Direct export of records to popular reference management services such as Mendeley and Zotero.

Once launched, the portal will be constantly redeveloped in line with emerging research practices in the humanities and digital humanities. Further studies will be made into how researcher can exploit and use this unique content and a content strategy will be developed in line with this.

¹⁰ Christine L. Borgman. "The Digital Future is Now: A Call to Action for the Humanities" *Digital Humanities Quarterly* 4.1 (2010). Available at: <http://works.bepress.com/borgman/233>

¹¹ www.theeuropeanlibrary.org/exhibition

Building bridges:

As well as the aggregation service, the portal and full text search capabilities developed in the Europeana Libraries project are now to be utilised to expose a new type of content: newspapers. Recent developments in OCR made through the IMPACT¹² project are now to be applied to newspaper content from 12 national and research libraries from across Europe. 18 million pages of newspapers will be refined and made available through Europeana and The European Library portal. The Europeana Newspapers project sets out to address the very specific challenges that making the full text of old newspapers searchable presents. It will make use of refinement methods for OCR, OLR/article segmentation, and named entity recognition (NER), and page class recognition.

Much of what has been developed and learned through Europeana Libraries will now be applied to Europeana Newspapers:

1. Aggregation:

Four types of existing digital newspaper collections can be identified:

- a) Images with only structural metadata
- b) Images with structural metadata and full text for searching (OCR)
- c) Images with structural metadata, article recognition (OLR/ article tracking) and OCR
- d) Images with structural metadata, OLR, OCR and semantic enrichment.

All data available will be harvested by The European Library. Data will be transformed to EDM, the data model of Europeana, and distributed for Europeana.

2. Metadata standardization.

A variety of metadata formats are currently in use. To improve access to digital content, common standards must be adopted. All existing metadata formats will be identified and best-practice solutions will be provided to the community.

3. Better Display Capabilities.

Making newspapers easy to search and presenting them attractively online is currently a challenge. The Newspapers Online project will look at the work done by Europeana Libraries and build on this work to specify appropriate search and presentation requirements, which can be used by Europeana.

Now, the text of newspapers from the past, as far back as the 18th century, will be fully searchable online. What's more, users will be able to view these papers in context, alongside art images, photographs, relevant these, books and articles. By presenting newspaper content in this way new connections may be made and doors opened for new types of research and collaboration.

There are other benefits to the work of this project as the procedures with which newspaper content will be upgraded include OCR, OLR/article tracking, NER, and page class recognition. For each of these technical tasks best practice

¹² <http://www.impact-project.eu/>

recommendations will be identified and published. This will be of huge benefit to the broader network of libraries with the CERL, CENL and LIBER networks. It will help reduce the cost of newspaper digitisation projects and increase the accessibility of digital newspaper collections now and into the future.

Conclusion:

Europeana Libraries was a best practice network that addressed a very practical need for a single aggregator for European libraries. In doing so it also brought together key library networks. Through working towards a common vision, the networks have created a resource which could have huge potential value for the research community. It has also created the conditions for national and research libraries to work together more fluidly, building on a vision to connect content and the researcher. The content held in Europe's libraries is rich and diverse. This is particularly true for newspapers holdings. Bringing these holdings online at a time when refinement technology are being developed to expose the full text in a meaningful way create a huge opportunity for researchers to interact with and drawn new connections between Europe's rich cultural heritage material. It is also the very embodiment of how organisations, networks, and institutions working together can produce innovative results, improve efficiency, and deliver on accessibility. Such work will have far reaching effects, not just for libraries or even for the accessibility of European cultural heritage, but for every country in the world with a mass of printed cultural material.

References:

Bulger, Monica, et al., Reinventing Research? Information Practices in the Humanities. London: Research Information Network, 2011.

Christine L. Borgman. "The Digital Future is Now: A Call to Action for the Humanities" Digital Humanities Quarterly 4.1 (2010). Available at: <http://works.bepress.com/borgman/233>

De Rosa, Cathy et al. 2005. Perceptions of Libraries and Information Resources: A Report to the OCLC Membership. Dublin, Ohio: OCLC.
<<http://www.oclc.org/reports/2005perceptions.htm>>.

European Commission, Maurice Lévy, Elisabeth Niggemann, and Jacques de Decker. The New Renaissance. Brussels: European Commission, 2011.
http://ec.europa.eu/information_society/activities/digital_libraries/doc/reflection_group/final_report_%20cds.pdf

Europeana (2012) Retrieved from <http://www.europeana.eu> on 19th April 2012.

Europeana Libraries (2012), Retrieved from <http://www.europeana-libraries.eu>, on 9th March 2012.

Europeana Travel (2012), Retrieved from <http://www.europeanatravel.eu/> on 19th April 2012.

IMPACT, Retrieved from <http://www.impact-project.eu/> on 19th April 2012.

The European Library(2012), Retrieved from <http://www.theeuropeanlibrary.org> on 9th March 2012.