
Roadmap for Improving Access to Digitised Newspapers

Revision: 1.0

Authors: Marieke Willems (LIBER), Friedel Grant (LIBER)

Contributions: Melanie Imming, Susan Reilly, LIBER Digital Collections Working Group

Revision History

Revision	Date	Author	Organisation	Description
0.1	9/12/2014	Marieke Willems	LIBER	Collection of Information, layout of structure
0.2	6/01/2015	Friedel Grant	LIBER	Writing of first draft
0.3	7/01/2015	Melanie Imming, Susan Reilly	LIBER	Review, comments
0.4	12/01/2015	Digital Collections Working Group	LIBER	Review, Comments
0.5	19/01/2015	Friedel Grant	LIBER	Integration of comments, further revision
1.0	29/01/2015	Clemens Neudecker, Sandra Kobel	SBB	Internal review and final version

Table of Contents

1. Executive Summary	3
2. Core Recommendations	5
2.1 Aim for open, free and comprehensive access to digitised newspapers.	5
2.3 Establish Clarity in the Areas of Copyright and the Right to be Forgotten.	8
2.4 Work Towards Metadata Standardisation and Improved Refinement.....	9
2.5 Continue An Open Dialogue Between All Stakeholders.....	10
3. Concluding remarks	12

3. 1. Executive Summary

The Europeana Newspapers project began in early 2012 with a mission to draw together millions of digitised, historic newspapers from across Europe, giving researchers and the general public a single corpus through which they could study a diverse range of topics in a pan-European context.

Underpinning this objective was the recognition of the daily press as an outstanding information resource and as a strong source of influence on Europe's culture, sense of identity and the development of democracy.

As the project nears the end of its lifetime (it is due to finish on 31 March 2015), much progress has been made. For users of historic newspapers, the metadata of millions of newspaper pages have been added to Europeana¹ and an online content browser². For 10 million of these pages, the text displayed in the newspapers has also been made fully searchable. The project has also released a series of open-source software tools which any institution can use to evaluate and improve the quality of content refinement within its own digitised newspaper collections, thereby providing users with a better experience and search capabilities³.

These efforts do not stand alone but are rather part of a bigger movement to improve digital access to newspapers. Many other projects, institutions and industry groups such as publishing organisations are also striving to make newspapers available to a broader public.

Despite this, constraints such as a lack of funding, legal and technical challenges mean that many historical newspaper collections are not yet accessible in a digital format, and collections which are online are not always easily or freely accessible. They may be hidden behind a paywall or limited to local access points. Furthermore, the online presentation of some collections and the tools available to refine search results and analyse data held within the newspapers are often unsatisfactory for advanced users such as professional researchers.

With these challenges in mind, Europeana Newspapers is publishing this Roadmap as a guide to improving access to digitised newspapers. It aims to inform policy and decision makers within funding institutions and cultural heritage organisations of the main issues currently being faced, as well as to provide advice so that they are better equipped to improve future access to digitised historic newspapers.

The recommendations have been drawn from the experiences of the project and points brought up during a workshop held in London in September 2014. The event – *Newspapers in Europe and the Digital Agenda for Europe* – brought together policymakers, researchers, libraries, copyright experts, publishers and members of the Europeana Newspapers project to discuss current realities and challenges with regards to European newspaper digitisation⁴, as well as solutions for common

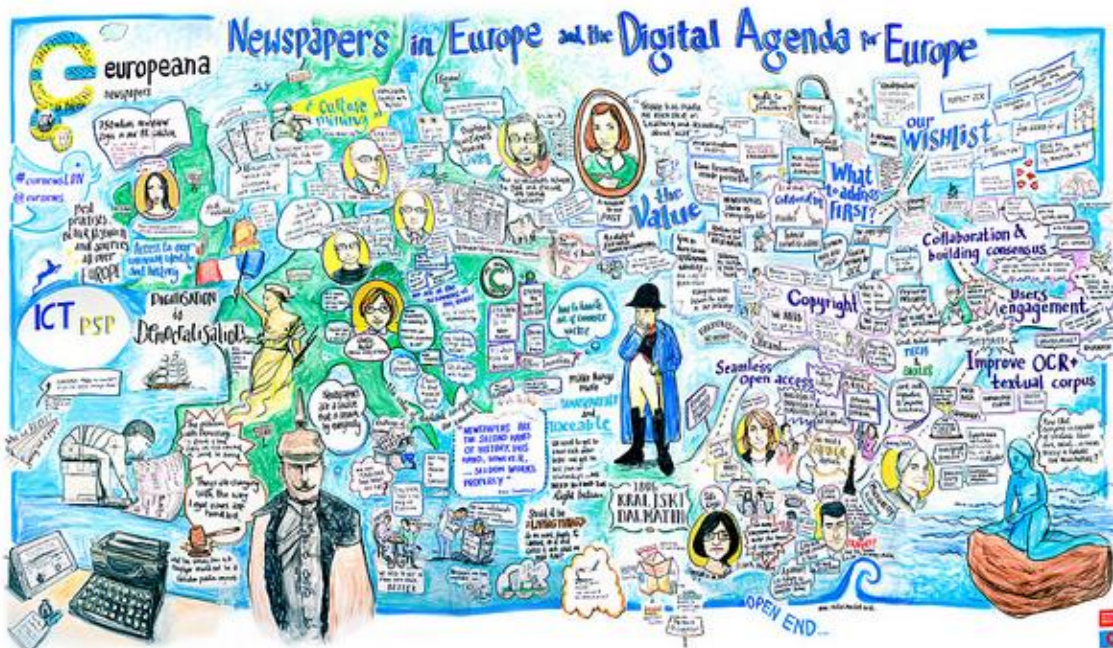
¹ <http://www.europeana.eu>

² <http://www.theeuropeanlibrary.org/tel4/newspapers>

³ <http://www.europeana-newspapers.eu/public-materials/tools>

⁴ Only 10% of all newspaper content is actually digitised. This is a very high barrier for access to historic newspapers. <http://www.enumerate.eu/fileadmin/ENUMERATE/documents/ENUMERATE-Digitisation-Survey-2014.pdf>

issues. For further details on the workshop itself, a separate report is available⁵. A high-resolution illustration can also be downloaded which graphically shows the main points of the debate⁶.



The issues raised in the London workshop are depicted in this diagram, which can be downloaded from the Europeana Newspapers website.

Although the aims in this Roadmap may not be immediately achievable, they are goals worth striving for. By digitising historic newspapers, the work of academic researchers, genealogists, local historians, teaching and learning communities and engaged citizens around the world will be supported.

The content made available will be a valuable resource for the study of European history, society, culture, languages, publishing, literature, art, design and much more. In addition, it has the ability to encourage new forms of research and re-use, particularly by the creative industries. As a result, new jobs and products can be created.

These goals, and the steps needed to reach them, should be taken into account by the stakeholders listed in Section 2.5 of this document.

⁵ http://www.europeana-newspapers.eu/wp-content/uploads/2014/10/D6.2.3_Report_on_Final_Workshop_London.pdf

⁶ <http://www.europeana-newspapers.eu/digital-newspapers-illustration>

4. 2. Core Recommendations

This Roadmap makes five key recommendations for improving access to digitised, historic newspapers.

2.1 Aim for open, free and comprehensive access to digitised newspapers.

Digitised historic newspapers are a relatively complex type of digital content. A corpus such as the one collected by Europeana Newspapers, for example, contains content which originates from 23 different countries, is printed in 20 different languages (with some newspapers containing more than one language) and uses multiple fonts, including Latin, Gothic, Cyrillic and Ottoman.

This creates challenges in terms of the digitisation and display of the newspapers. For example, performing OCR (Optical Character Recognition) on intricate, antique fonts may be impossible or result in many errors. This in turn makes it difficult for users to reliably search the full text of such newspapers.



The text used in newspapers such as this one, from the National Library of Latvia, can be difficult for computers to read, making it complicated to produce a reliable full-text version.

At the same time, because newspapers are such a rich source of information, containing news from all levels of society, reported on a daily or weekly basis, they are in great demand by researchers who have very high expectations in terms of access to digitised newspapers. These researchers are generally not interested in simply browsing digital newspapers but rather have a specific search strategy in mind and want the tools and refined search options to enable them to carry out this work.

During the London workshop, it was agreed that organisations should aim for the following in order to provide seamless access to digitised historic newspapers:

1. **Provide free access whenever possible.** In an ideal world, historic newspaper collections should be freely available, rather than locked behind paywalls. Participants acknowledged, however, that this recommendation, perhaps more than any other, is heavily reliant on the availability of funding from other sources if it is to be achieved (e.g. government funded projects to make content available).
2. **Encourage community engagement.** Those using the newspapers should be given the opportunity to improve the resource (e.g. by submitting corrections to OCRed text, tagging articles with keywords). This is a way of both developing a loyal audience and increasing the quality of the corpus. Such functions would be appreciated by many user groups and would improve the level of text accuracy in the corpus. As the highly successful digital library Trove noted:
*'The primary motivator for embarking upon collaborative text correction was to improve data quality and this has been a success. Another outcome is that the Library is now beginning to understand that engaging users in services, empowering them to make a difference, and building social networking communities is almost if not equally as important to the users as having high quality data. Giving control to users and entrusting the community to have such a crucial role in the development of a service helps build a dedicated, responsible and engaged user base - a major asset for a library wanting to remain relevant and visible in the digital age.'*⁷
3. **Make full-text content available.** The text within newspapers should be processed with OCR, OLR (Optical Layout Recognition) and NER (Named Entity Recognition) technology so that researchers can effectively search for terms, themes, specific people and places, and individual articles contained within the papers.
4. **Provide tools to allow analysis and extraction of data.** In addition to high-quality content, tools should be provided which allow researchers to perform an in-depth search and analysis of the material. For example, tools could be provided to translate content from other languages, to search specifically for images, to content mine the newspapers and to extract data.
5. **Give additional context.** Where possible, links should be provided to related concepts and other types of related content (e.g. the name of a famous person could be linked to the Wikipedia entry for that person).
6. **Label content clearly with licenses.** Content should be labelled clearly with a license indicating what it can and cannot be used for. This will enable users to exploit the content to its full potential (see 2.3 *Establish Clarity in the Areas of Copyright and the Right to be Forgotten* for more information).
7. **Make content citable and downloadable.** Newspaper content should be citable at an article level and individual pages or articles should be downloadable, so that researchers can create their own personal archive for individual projects.

⁷ <http://conference.ifla.org/past-wlic/2009/99-gatenby-en.pdf>

2.2 Ensure sustained funding of, and investment in, newspaper digitisation.

There are many reasons to invest in the digitisation of historic newspapers. In the broadest possible context, digitised historic newspapers can be seen as one part of a wider movement to open up data in the modern age.

This exposure of information to the public at large, offers great potential for research, collaboration and innovation, leading to benefits for the economy and society as a whole. The more newspapers are digitised, the greater this potential will be.

On a more micro level, the funding of newspaper digitisation comes at a time when demand for access to such material from researchers is growing. The availability of newspapers in a digital format gives researchers the opportunity to examine topics and questions from angles which were not previously possible.

For institutions, the successful opening up of a newspaper collection online can raise the public profile of the institution and help further convince funders and supporters of the institution's value. It can also help institutions to work more effectively by streamlining workflows and to better serve their users.

Despite this, participants at the London workshop noted that there was currently not enough funding to perform such work. Despite major contributions towards digitisation from EU Member States and the Commission⁸ the vast majority of Europe's cultural heritage, including newspapers, remains in analogue format and content-holding institutions continue to face budget cuts.



The digitisation of newspapers can be expensive and there is currently not enough funding to meet user demand for this type of resource.

⁸ http://www.slideshare.net/Europeana_Newspapers/eurnewsldnkrzysztofnychczynski

The hunt for funding has led some libraries, such as the British Library, to sign public-private partnerships. Patrick Fleming, Head of Business Change at the library, noted that their British Newspaper Archive⁹ was only made possible thanks to an agreement with DC Thomson. This agreement provided the library with critical funds and allowed them to give users wider access to content than the British Library could have done on its own¹⁰.

Concerns about funding were shared not only by cultural heritage institutions but also by publishing groups, which tend to hold large archives of digital newspapers, particularly from more recent years. These archives are of great interest to researchers but are often only accessible when specific conditions of access are met.

Satu Kangas – Director of Legal Affairs and Media Policy at FinnMedia, a member of the Finnish Copyright Council and of the Copyright Working Group of the European Newspaper Publishers Association (ENPA) – noted that publishers were already struggling to deal with rising costs at the same time as revenues were falling from printed newspapers¹¹. Demand was high for digital newspapers to be made available, Kangas said, but publishers also had a great need to monetise this content and it was not clear how this could be done.

2.3 Establish Clarity in the Areas of Copyright and the Right to be Forgotten.

There is a clear need to establish clarity on copyright as it applies to the digitisation of newspapers. The application and interpretation of copyright law varies significantly between countries across Europe and is impacting on the availability of historic newspaper collections.

When libraries and others holding digital newspaper collections are uncertain about the copyright status of those works, they may avoid digitising and making those collections available to a broader public. Users, meanwhile, often come across collections of material for which the copyright status is not clearly marked. This inhibits reuse, as the people accessing the resource cannot evaluate if it is possible, for example, to download and use the material for teaching purposes, to publish it in a book or incorporate it in a new, artistic creation.

In order to address this challenge, workshop participants proposed a broad collaboration between the cultural sector, libraries, publishers, authors' associations and European organisations such as LIBER, with the aim of building both a common European understanding of copyright and agreements and a simple licensing process.

The foundations of such collaborations have already been laid via Europeana Newspapers, and suggestions for continuing this work are covered under Section 5 of this document.

⁹ <http://www.britishnewspaperarchive.co.uk>

¹⁰ http://www.slideshare.net/Europeana_Newspapers/eurnewsldnpatrickfleming

¹¹ http://www.slideshare.net/Europeana_Newspapers/eurnewsldnsatukangas



Some of the issues related to Copyright are depicted in this diagram – part of a larger illustration created at the Newspapers in Europe and the Digital Agenda for Europe workshop.

In addition to copyright, the “right to be forgotten” emerged as another legal issue impacting on those who provide access to digitised newspapers. How should we address the desire of some individuals to live their lives, without being stigmatised by a past action which was documented in a newspaper?. Should the digital providers of historic newspapers be prepared to remove content if requested in such circumstances, and how could this be done without effectively re-writing history?

The answers to these questions are not yet entirely clear, but it is nevertheless felt that libraries, the industry (Wikimedia), public funders, law makers, publishers, authors and collecting societies must work together to build a set of transparent ethical and legal guidelines which can be used to evaluate and address this and similar situations.

2.4 Work Towards Metadata Standardisation and Improved Refinement.

A digital newspaper archive which truly satisfies the needs of researchers cannot be achieved without a significant investment in the metadata that accompanies each paper and the technology that makes it possible to display and search the material in a meaningful way.

The Europeana Newspapers project has already contributed towards the improvement of digital newspaper metadata and refinement. It has, for example, published a report on how to align newspaper metadata to the Europeana Data Model¹² and held numerous workshops on aggregation, refinement and quality assessment of newspaper content¹³.

¹² http://www.europeana-newspapers.eu/wp-content/uploads/2012/04/D4.4_ENP_EDMforNewspapers.pdf

¹³ <http://www.europeana-newspapers.eu/public-materials/deliverables/>

With the project now at an end, however, libraries with digital newspaper collections, aggregators, organisations such as the IMPACT centre of competence¹⁴ and front-line content providers must continue this work to improve the available tools and achieve better standards of precision.

This can be done through collaboration (worldwide, between sectors or via crowdsourcing), by sharing success stories, by establishing knowledge centres where others can see the research done to date and by creating software and training sets which share best practices with others.

2.5 Continue An Open Dialogue Between All Stakeholders.

During its 3-year lifetime, the Europeana Newspapers project collaborated directly with 35 networking partners¹⁵ and brought together countless other institutions via the three Workshops, 10 Information Days and continual media activity undertaken as part of the project¹⁶.

This was important both for the project as a whole and individual institutions, as it gave them an opportunity to discuss topics related to digitised historic newspapers with others trying to achieve the same goals. It also enabled the sharing of individual success stories, case studies and Best Practice. Through this process, relationships were forged both within sectors (e.g. library to library) and across sectors (e.g. cultural heritage institutions and publishers).



Events, such as this Europeana Newspapers workshop in Belgrade, should continue so that stakeholders with an interest in newspaper digitisation can continue to share Best Practices with each other.

¹⁴ <http://www.digitisation.eu/>

¹⁵ <http://www.europeana-newspapers.eu/consortium/project-partners/#networking>

¹⁶ Reports providing more detail on these many initiatives are available on the project website <http://www.europeana-newspapers.eu/public-materials/deliverables/>

In order to continue improving access to digitised newspapers, this collaboration must continue. This could be done, for example, via an email discussion list or face-to-face meetings. The format of the collaboration is less important than the fact that it continues in some way, ensuring that all parties interested in improving access to digitised newspapers have a venue for discussion and further fruitful collaborations between stakeholders.

During the London workshop, the premise of an ongoing cooperation was broken down further into the following actions:

Organisations such as EBLIDA, IFLA and LIBER:

- Develop relationships with stakeholders and encourage further dialogue on the topic of improving access to newspapers;
- Share best practices among members with newspaper content, especially with those who still haven't digitised their content;
- Provide online visibility for collections where possible;
- Help to establish clarity on copyright;
- Help to draw up principles regarding licensing agreements and the right to be forgotten;
- Share success stories, for example: the British Library Qatar Foundation¹⁷, Comellus project¹⁸, VanGoYourself (a Europeana "selfie" re-enactments of digital heritage¹⁹), or the examples of Poland and Finland, who have been very successful in finding regional structural funds for digitisation.

Infrastructure providers such as Europeana:

- Act as facilitators;
- Provide online visibility;
- Act as a platform to develop relationships with stakeholders;
- Identify themes and flowcharts to facilitate use of examples.

Individual libraries that have digitised newspaper collections:

- Enter into dialogue with researchers to learn about their needs;
- Enable re-use;
- Establish connections beyond the library itself (e.g. research libraries could reach out to public libraries and their users).

¹⁷ <http://www.bl.uk/qatar/>

¹⁸

http://www.ifla.org/files/assets/newspapers/Singapore_2013_papers/day_2_04_2013_ifla_satellite_kaukonen_m_hosio_m_preservation_and_access_of_digitally_deposited_newspapers.pdf

¹⁹ <http://www.pro.europeana.eu/web/europeana-creative/blog/-/blogs/hacking-cultural-tourism-with-%E2%80%9Cselfie%E2%80%9D-reenactments-of-digital-heritage:-an-interview-with-the-project-leaders-behind-vangoyourself>

Publishers:

- Enter into dialogue with libraries, policy makers and researchers on copyright.

Researchers:

- Discuss their needs with libraries and publishers;
- Look into the possibilities of a shared safe space for research use only.

3. Concluding remarks

Making more historic newspapers available to the widest possible audience is important on many levels: for individual research projects, for institutions who want to draw attention to their valuable collections and for society's quest to learn from and build on our shared history.

For all of these reasons, it is hoped that this Roadmap will help to inform people about the issues surrounding the digitisation and publication of historic newspapers, of their importance as an information source and as a catalyst for further discussion between all stakeholders with an interest in making digitised, historic newspapers available to the public.

We encourage all stakeholders to discuss the goals and messages contained within this Roadmap and to share the Roadmap with others so that the discussions and sharing of Best Practices started by the Europeana Newspapers project can continue into the future.