



# Europeana DSI 2– Access to Digital Resources of European Heritage

**DELIVERABLE**

## D2.4: EUROPEANA NEWSPAPERS COLLECTIONS PLAN

<b>Revision</b>	1.0
<b>Date of submission</b>	28 February 2017
<b>Author(s)</b>	Clemens Neudecker (SBB)
<b>Dissemination Level</b>	Public



Co-financed by the European Union  
Connecting Europe Facility

## REVISION HISTORY AND STATEMENT OF ORIGINALITY

### Revision History

Revision No.	Date	Author	Organisation	Description
0.1	17-10-2016	Clemens Neudecker	Berlin State Library	Initial draft
0.2	26-10-2016	Clemens Neudecker	Berlin State Library	Revised draft
0.3	10-01-2017	Clemens Neudecker	Berlin State Library	New structure
0.4	23-01-2017	Clemens Neudecker	Berlin State Library	Populated new structure
0.5	13-02-2017	Clemens Neudecker	Berlin State Library	Consolidated draft
0.6	14-02-2017	Nienke van Schaverbeke, Adrian Murphy, David Haskiya	Europeana Foundation	Comments and suggestions for revision
0.7	21-02-2017	David Haskiya	Europeana Foundation	Section on product design and development added
0.8	22-02-2017	Clemens Neudecker	Berlin State Library	Comments and suggestions addressed
0.9	23-02-2017	Douglas McCarthy	Europeana Foundation	Final review
1.0	27-02-2017	Clemens Neudecker	Berlin State Library	Finalised document

### Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

The sole responsibility of this publication lies with the author. The European Union is not responsible for any use that may be made of the information contained therein.

'Europeana DSI is co-financed by the European Union's Connecting Europe Facility'

## Table of Contents

<b>1. Purpose of this document</b>	4
1.1 Relationship to other documents	4
<b>2. Elevator Pitch</b>	4
<b>3. Introduction to Europeana Newspapers</b>	4
<b>4. Business Model</b>	5
4.1 Audiences	6
4.2 Value proposition	7
4.3 Market landscape	7
4.3.1 Market trends	9
4.3.2 SWOT analysis	9
4.4 Channels and networks	10
4.4.1 Advisory board	12
4.4.2 Events and workshops	13
4.4.3 Meetings and conferences	13
4.5 Key activities	14
4.5.1 Themes / Exhibitions	14
4.5.2 Browse entry points	14
4.5.3 Blog	14
4.5.4 Social media	15
4.5.5 User engagement activities	15
4.5.6 Activities timeline	15
4.6 Resources	16
4.7 Impact	16
<b>5. Product design and development</b>	19
5.1 Scope of product development	19
5.2 Evaluation	20
5.3 Product development methodology and policies	21
5.4 Roadmap	21

## 1. Purpose of this document

This plan covers the ambitions and high level milestones for the Europeana Newspapers thematic collection over the duration of DSI-2 and beyond. It will be evaluated on an ongoing basis against implementation, results and actual landscape to ensure effectiveness.

### 1.1 Relationship to other documents

This document follows strategy, direction and broad activities described in the Europeana Collections Plan<sup>1</sup> (Deliverable 2.1).

## 2. Elevator Pitch

*“Europeana must now be regarded as the first port of call for anybody who wishes to explore the continent’s newspapers”* - British Library Labs prize-winner Dr. Bob Nicholson, in his review of Europeana Newspapers for “Reviews in History”.<sup>2</sup> As Dr. Nicholson notes in his review, Europeana Newspapers has improved access to Europe’s digital historical newspapers dramatically.

Europeana Newspapers is currently the largest single collection of digitised historical newspapers from Europe, and the only online platform to provide a single entry point to digital newspaper full-text content covering four centuries, and more than a dozen languages, from major national and research libraries in Europe.

Furthermore, Europeana Newspapers raises the bar for open library data, with a majority of the content (images and OCRred text) being made available without any restrictions for anyone to use and reuse due to public domain licensing, and all metadata being released under CC0.

Thanks to its volume, variety and availability, the Europeana Newspapers collection is of tremendous value to a broad community of users, and can help enable new ways in research and education as well as support the further standardisation and adoption of best-practices for the large-scale digitisation and online presentation of historical newspapers in Europeana.

## 3. Introduction to Europeana Newspapers

Europeana Newspapers is foreseen to launch early in 2018 as a thematic collection on Europeana and will be managed by the Berlin State Library<sup>3</sup>. Providing access to around 20 million pages - 12 million pages fully searchable - of digital historical newspaper content from 23 libraries in Europe, it forms one of the largest collections of Europe’s historical newspapers currently available online.

---

<sup>1</sup> [http://pro.europeana.eu/files/Europeana\\_Professional/Projects/Project\\_list/Europeana\\_DSI-2/Deliverables/d2.1-europeana-collections-plan.pdf](http://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_DSI-2/Deliverables/d2.1-europeana-collections-plan.pdf)

<sup>2</sup> <http://www.history.ac.uk/reviews/review/1894>

<sup>3</sup> <http://staatsbibliothek-berlin.de/>

The Europeana Newspapers thematic collection will build on the results of the highly successful 2012-2015 ICT-PSP project Europeana Newspapers (Ref. 297380)<sup>4</sup>. The Europeana Newspapers project was a joint effort of 18 partners with the aim to make a significant amount of Europe's historical newspapers fully searchable by means of refining scanned newspaper pages from libraries throughout Europe with OCR and, to some extent, also article separation. At the close of the project, the historical newspapers portal prototype developed by The European Library (TEL)<sup>5</sup> provided access to more than 20 million pages of historical newspapers from as early on as 1618 and up to the 21st century - equivalent to more than 1,000 newspaper titles with around 3,6 million issues. As indicated below (4.7), the prototype has seen a very promising degree of use, with around 1.3 million visitors to the site in 2016 and an exceptionally high user session duration of over 16 minutes, indicating that users really dive into the full-text content of the newspapers.

Access to the collection is possible via a variety of entry points such as

- full-text search (string search), incl. autosuggestions
- search by newspaper title
- search by date of issue using a calendar
- search by country of publication on a map

Additionally, filters and facets like "language", "content provider" or "decade" allow for a further fine-tuning of the result set. A carousel with browsable title pages of newspapers published "on this day" completes the landing page of the portal. Usability testing has been done for the TEL portal, and the full report<sup>6</sup> is available online.

Three datasets have been released that compile subsets of the Europeana Newspapers collection targeted at specific audiences

- ENP Ground Truth dataset<sup>7</sup>
- ENP Named Entity corpora for French, Dutch, German<sup>8</sup>
- Plain text downloads of OCRred text<sup>9</sup>

Finally, work is undertaken to further develop a Europeana Newspapers API that will allow programmatic access to metadata, images and OCRred text.

It is the aim of Europeana to replicate the full feature set of the TEL portal prototype described above for the Europeana Newspapers thematic collection. The release of features and functionalities will happen in phases, starting with a public alpha release with a limited feature set in January 2018 and expanding from there based on user feedback and needs.

## 4. Business Model

Below we have attached the business model canvas of Europeana Newspapers. The thematic collection is central for its activities and scopes, as it is the main channel through which it can reach its target audience and engage its partners.

---

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

<sup>5</sup> <http://www.theeuropeanlibrary.org/tel4/newspapers>

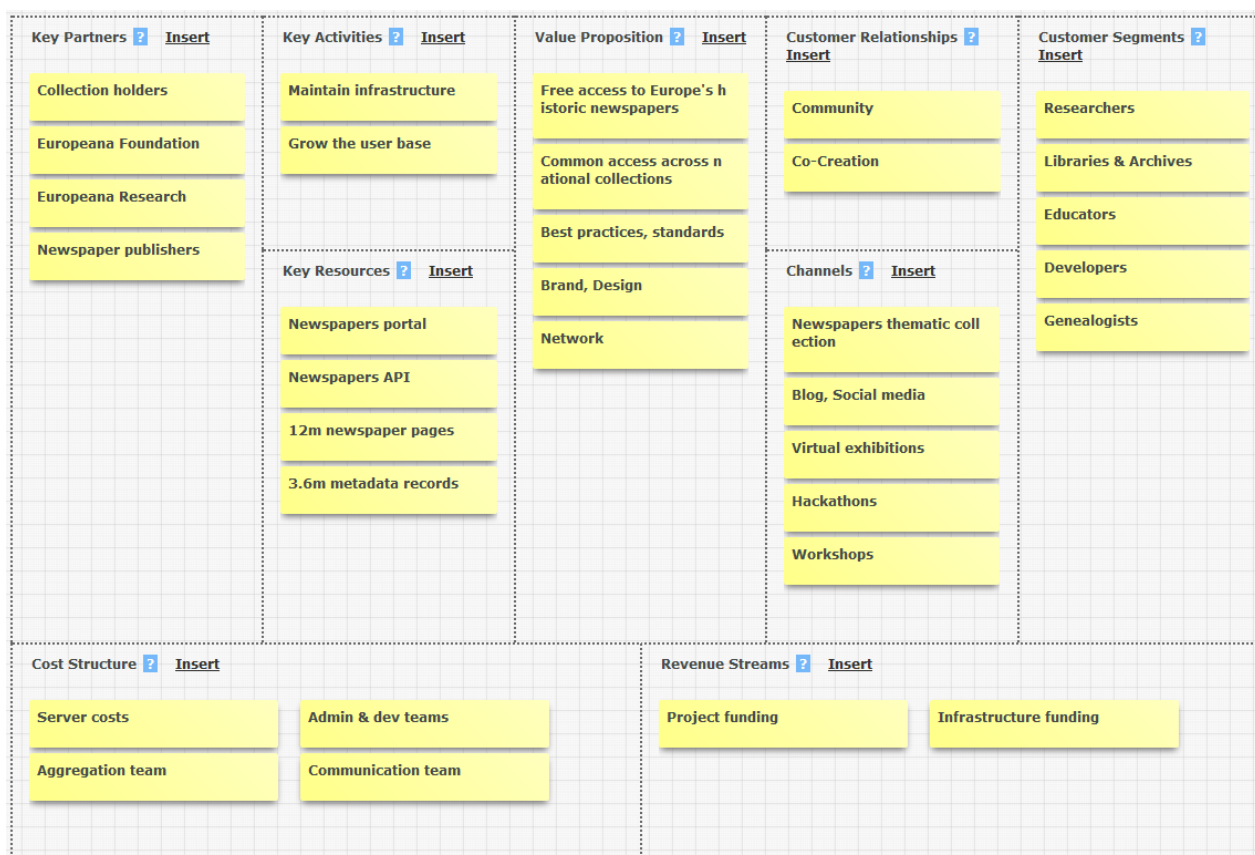
<sup>6</sup> <http://www.europeana-newspapers.eu/wp-content/uploads/2014/05/The-European-Library-Newspaper-Archive-Usability-testing-Report-April-2014.pdf>

<sup>7</sup> <http://primaresearch.org/datasets/ENP>

<sup>8</sup> <https://github.com/EuropeanaNewspapers/ner-corpora>; due for publication in ELRA

<http://catalog.elra.info/>

<sup>9</sup> <http://research.europeana.eu/itemtype/newspapers>



*Business model canvas for Europeana Newspapers;*  
<https://canvanizer.com/canvas/rkch03Nv5PazM>

## 4.1 Audiences

As seen from a series of interviews in 2014 and 2015<sup>10</sup>, this large and diverse historical newspaper collection has already attracted a broad user base. This includes scholars in the humanities and social sciences, researchers in computer science and language processing as well as teachers in schools and universities, but also important stakeholders such as libraries or other cultural heritage organisations holding historical newspaper collections. The core audiences for Europeana Newspapers can therefore be roughly prioritised into the following groups.

- **Research**

Researchers from the humanities and social sciences are particularly interested in historical newspapers, as they allow them to study public opinion and culture in the past and provide a perspective on the daily life of ordinary people that is often not accounted for by history textbooks.

Also researchers from the field of language studies, computational linguists and philologists show great interest in historical newspapers. A corpus of full-text covering four centuries, like Europeana Newspapers, enables them to look into the use and development of language over time.

Historical newspapers are also one of the main sources for genealogists or people wanting to perform research into their family history.

<sup>10</sup> <http://www.europeana-newspapers.eu/category/interviews-with-researchers/>

Finally, the pattern recognition community is researching novel algorithms for OCR and document analysis. Many of the unsolved challenges in this domain lie in the area of historical documents or the complex layout typically found in newspaper pages. The Europeana Newspapers data and ground-truth<sup>11</sup> open up many possibilities for the development and testing of innovative approaches and technologies for the recognition and analysis of historical documents based on a real-world representative dataset.

- **Education**

Due to its availability under open licensing and the accessibility of newspaper content, the Europeana Newspapers thematic collection is highly suited for reuse in education, either on the school or university level.

- **Libraries & Archives**

Libraries and other institutions holding large collections of historical newspapers that they are aiming to digitise are naturally a key audience for Europeana Newspapers. These organisations are mainly interested in knowledge sharing about best practices, standards and technologies in the community as well as how to get their data aggregated to and shown in the Europeana Newspapers thematic collection.

## 4.2 Value proposition

Europeana Newspapers offers value to a wide community of users thanks to a number of core strengths.

First and foremost, Europeana Newspapers offers the currently largest free online collection of historical newspapers from Europe. A majority of the content has been made available under public domain licensing, providing a vast resource that can be easily used and reused for various purposes.

Second, Europeana Newspapers is the first historical newspaper portal that gives access to a cross-language newspaper collection from numerous European cultural heritage institutions via a single user interface. With its representative collection from European member states, it will play an instrumental role in furthering the deeper understanding of European history through the scholarly analysis of Europe's historical newspapers. The cross-language and cross-border perspectives provided through the collection, make it much easier to conduct comparative research or teaching about the prevalent views on a particular subject in different European countries.

Finally, due to common standards and best-practices and the network established by Europeana Newspapers, it will support capacity building and the further harmonisation and interoperability of digital historical newspaper collections on the European (and international) level.

## 4.3 Market landscape

For a long time, Europe was trailing behind with regard to the online availability of digital historical newspapers. Due to fragmentation into national or regional collections, and the lack of common standards and best practices for newspaper digitisation, resources outside of Europe have attracted most users, including from Europe.

---

<sup>11</sup> <http://primaresearch.org/datasets/ENP>

One of the largest and longest running newspaper digitisation programmes is the Trove<sup>12</sup> service of the National Library of Australia. Launched as a project in August 2008, the site has grown steadily in content and now provides access to some 200 million newspaper articles. Trove has also made headlines with its spectacular crowdsourcing achievements<sup>13</sup>, attracting a large user group who continuously engage in the online correction of the OCRed historical newspaper text and having already contributed more than 40 million corrected text lines.

Another major newspaper digitisation initiative is the National Newspaper Digitisation Program<sup>14</sup> (NDNP) at the Library of Congress (LC). Initiated in 2003 as a partnership between the National Endowment for the Humanities (NEH) and the LC, in October 2015 the US national newspaper portal Chronicling America celebrated the publication of its 10-millionth historical newspaper page online<sup>15</sup>.

Within Europe, one of the longest running and most successful newspaper digitisation programmes is Austrian Newspapers Online<sup>16</sup> (ANNO) by the Austrian National Library. Since its kick-off in 2003, ANNO has continuously expanded its collection and now offers more than 17 million pages online. However, mainly due to the challenges in the OCRing of Gothic letters, ANNO does not currently include OCRed text for all titles in the collection - around 4.6 million pages were among the first to be OCRed in the frame of the Europeana Newspapers ICT-PSP project. All rights-free newspaper pages in ANNO have been integrated into the Europeana Newspapers collection, with metadata being aggregated too for those titles that still have access restrictions.

Another important newspaper digitisation programme in Europe is the Databank Digitale Dagbladen (DDD) launched by the National Library of the Netherlands in 2007. Around 11 million pages have been digitised since, always including OCR and article separation, and are presented in the Dutch national digital library Delpher.<sup>17</sup> All rights-free newspaper pages in Delpher have been integrated into the Europeana Newspapers collection, with metadata being aggregated too for those titles that still have access restrictions.

With around 18 million pages online, the British Newspaper Archive<sup>18</sup> also ranks amongst the largest collections of digital historical newspapers in Europe. However, since the digitisation was done mostly via Public-Private Partnerships, accessing the content is currently only possible with a paid monthly subscription.<sup>19</sup>

While also major commercial databases of historical newspapers have been assembled by e.g. Ancestry.com<sup>20</sup> and Newspapers.com<sup>21</sup>, access to the content is locked behind paywalls. Furthermore, there is a strong focus on newspapers from the US, with only marginal amounts of newspapers from other regions being included. In contrast, Europeana Newspapers offers free

---

<sup>12</sup> <http://trove.nla.gov.au/newspaper/>

<sup>13</sup> Holley, Rose. "Many hands make light work: Public collaborative OCR text correction in Australian historic newspapers." (2009). [http://www.nla.gov.au/ndp/project\\_details/documents/ANDP\\_ManyHands.pdf](http://www.nla.gov.au/ndp/project_details/documents/ANDP_ManyHands.pdf)

<sup>14</sup> <https://www.loc.gov/ndnp/>

<sup>15</sup> <http://blogs.loc.gov/thesignal/2015/10/extra-extra-chronicling-america-posts-its-10-millionth-historic-newspaper-page/>

<sup>16</sup> <http://anno.onb.ac.at/>

<sup>17</sup> <http://www.delpher.nl/nl/kranten>

<sup>18</sup> <http://www.britishnewspaperarchive.co.uk/>

<sup>19</sup> <https://www.britishnewspaperarchive.co.uk/account/subscribe>

<sup>20</sup> <https://www.ancestry.com/>

<sup>21</sup> <https://www.newspapers.com/>



and unrestricted access to all its assets, with the possibility to download and reuse a vast amount of content from the thematic collection, including images and OCRed text. More than that, Europeana Newspapers has already released several open datasets and continuously seeks out further ways to make the data easily accessible and available for reuse.

Compared to all of the above, Europeana Newspapers offers a range of added benefits. First and foremost, what makes the Europeana Newspapers collection unique is the European scale of newspaper aggregation. Secondly, thanks to the open licensing of the data, it can be freely used and reused by stakeholders in various contexts. Finally, Europeana Newspapers has established best practices and standards for scalable newspaper digitisation and presentation in Europe.

### 4.3.1 Market trends

In recent years, the digitisation of historical newspapers in Europe has seen an increase in both, scale and also quality. In addition to scanning, OCR and even article separation have now become standard in most newspaper digitisation projects. National libraries in particular have expanded their activities in this area with especially the Nordic countries<sup>22</sup> running massive projects, like e.g. the Danish National Library who are digitising 32 million pages<sup>23</sup>.

In Germany, the German Research Foundation (DFG) is currently preparing a master plan for the digitisation of the nation's historical newspapers<sup>24</sup>, and there are strong ties to the results and recommendations of Europeana Newspapers. The adoption of best-practices established by Europeana Newspapers will lead to further libraries joining the network and digitising newspaper content for aggregation into the German Digital Library (DDB) and thus, the Europeana platform.

At the same time, the use of digital historical newspaper collections for research has seen a move away from the study of national collections into the comparative analysis of large-scale collections across borders and languages. Examples are the Asymmetrical Encounters project<sup>25</sup> or the Viral Texts project<sup>26</sup>. Europeana Newspapers has supplied both projects with data and seeks further collaborations with projects and initiatives that have similar aims.

### 4.3.2 SWOT analysis

SWOT analysis covers Europeana Newspapers activities and the thematic collection which is central to these.

#### Strengths

- Large volume of collection with approx. 20 million pages, 12 million pages OCR equivalent to roughly 1,000 titles or 3.6 million issues
- Broad variety of topics and themes covered in newspapers attracts diverse users
- Free access and availability of a majority of the data under public domain license
- Unique assembly of cross-national newspaper content, covering 17 EU countries
- Cross-lingual collection in more than 20 different languages, searchable via a single portal
- Presentation of searchable OCRed text and images on a single user interface
- Strong European and international network

---

<sup>22</sup> <http://www.liber2015.org.uk/event/digitisation-of-newspapers-in-european-libraries/>

<sup>23</sup> <http://en.statsbiblioteket.dk/national-library-division/newspaper-digitisation/>

<sup>24</sup> [https://opus4.kobv.de/opus4-bib-info/files/2066/Workshop\\_Zeitungsdigitalisierung\\_DBT\\_Juni+2015.pdf](https://opus4.kobv.de/opus4-bib-info/files/2066/Workshop_Zeitungsdigitalisierung_DBT_Juni+2015.pdf)

<sup>25</sup> <http://asymenc.wp.hum.uu.nl/>

<sup>26</sup> <https://www.northeastern.edu/nulab/viral-texts/#content>

- Strong uptake by researchers across various domains

### Weaknesses

- OCR and OLR quality varies and is overall at about 80% accuracy
- Multilingual content and historical spelling can cause difficulties for users when searching
- Mixed licenses - while most content is PD, some still carries access restrictions
- Mixed levels of refinement - not all pages have OCR, only about 20% have article separation

### Opportunities

- Enrichments and visualizations via research projects and apps using the data
- Crowdsourcing for correction of OCR errors
- Tagging and transcription of full-text
- Extraction and publication of images
- Reuse in educational or entertainment apps

### Threats

- Copyright restrictions remain for some newspapers from the 20th century
- Unresolved errors in collection aggregated by TEL
- Stop of content aggregation from libraries until new aggregation model is found
- Issues and errors in TEL portal until re-launch as Europeana thematic collection cause frustration with users

## 4.4 Channels and networks

Various social media and dissemination channels are in place and actively curated:

- **Twitter** account <https://twitter.com/eurnews> (1,230 followers)  
Twitter is used mainly for announcements and to promote Europeana Newspapers content, new developments, themes or to retweet other sources that are relevant to the community. Tweets will appear at random frequency, but at least weekly and often more frequently.
- **Facebook** account <https://www.facebook.com/EuropeanaNewspapers> (1,367 likes)  
Facebook is used for promoting Europeana Newspapers content, new developments, themes or blog posts and to highlight trends and events of relevance to the community. Facebook posts will appear at random frequency, but at least monthly and often more frequently.
- **LinkedIn** group <http://www.linkedin.com/groups?gid=4425919>  
LinkedIn is used to announce major developments, events, workshops and conferences targeted at cultural heritage and newspaper professionals.
- **GitHub** account <https://github.com/EuropeanaNewspapers>  
GitHub is used for the development, management and distribution of open source software tools resulting from Europeana Newspapers.
- **Slideshare** account [http://www.slideshare.net/Europeana\\_Newspapers](http://www.slideshare.net/Europeana_Newspapers)  
Slideshare is used to publish presentations from events and workshops organised by Europeana Newspapers.
- **Flickr** account <https://www.flickr.com/photos/enewspapers/>  
Flickr is used to publish images from Europeana Newspapers events as well as selected newspaper pages and graphical content extracted from newspapers.

The Europeana Newspapers website and blog<sup>27</sup> is another important channel, frequently attracting feedback and comments from users. More information on the use and communication strategy of the blog can be found under [4.5.3](#).

While work in 2017 will progress the migration of the data and functionalities provided in the TEL portal prototype, communication activities of Europeana Newspapers will put particular focus on the various datasets that have been released, and how they can be/are being used in various research contexts.

Europeana Newspapers can build on an existing network established during the ICT-PSP project and comprising a diverse and healthy mix of stakeholders such as libraries, researchers and end-users (including scholars, educators, designers and computer scientist) of digital newspapers. The network currently comprises the 18 project partners, 11 associated partners and another 40 networking partners<sup>28</sup>. With the help and further growth of this network, Europeana Newspapers is in a comfortable position to continue and expand on the activities of the Europeana Newspapers project and make Europeana the best and most comprehensive platform for anybody who wishes to explore Europe's historical newspapers.

The Europeana Network Association with its members is another important community of experts and stakeholders that will be very useful to the further development of the Europeana Newspapers collection. Last but not least, the collaboration with other thematic collections on Europeana opens up many opportunities to explore synergies in the area of marketing and cross-promoting the collections.

Furthermore, several members of the Europeana Newspapers network actively participate in the IFLA News Media section<sup>29</sup> as well as in other library associations like LIBER, where information about the Europeana Newspapers thematic collection can be disseminated widely amongst cultural heritage professionals, thereby creating additional awareness and engagement in the library community.

Also already during the ICT-PSP project, a dialogue has been established between Europeana Newspapers, Europeana, the European Commission and ENPA<sup>30</sup>, the European Newspapers Publishing Association. Representatives of the EC and ENPA also participated in the main closing event of the Europeana Newspapers project where they contributed to the Roadmap for Improving Access to Newspapers<sup>31</sup>. The roadmap is intended 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 historical newspapers. It can therefore serve as a useful guideline for the long-term development of Europeana Newspapers.

---

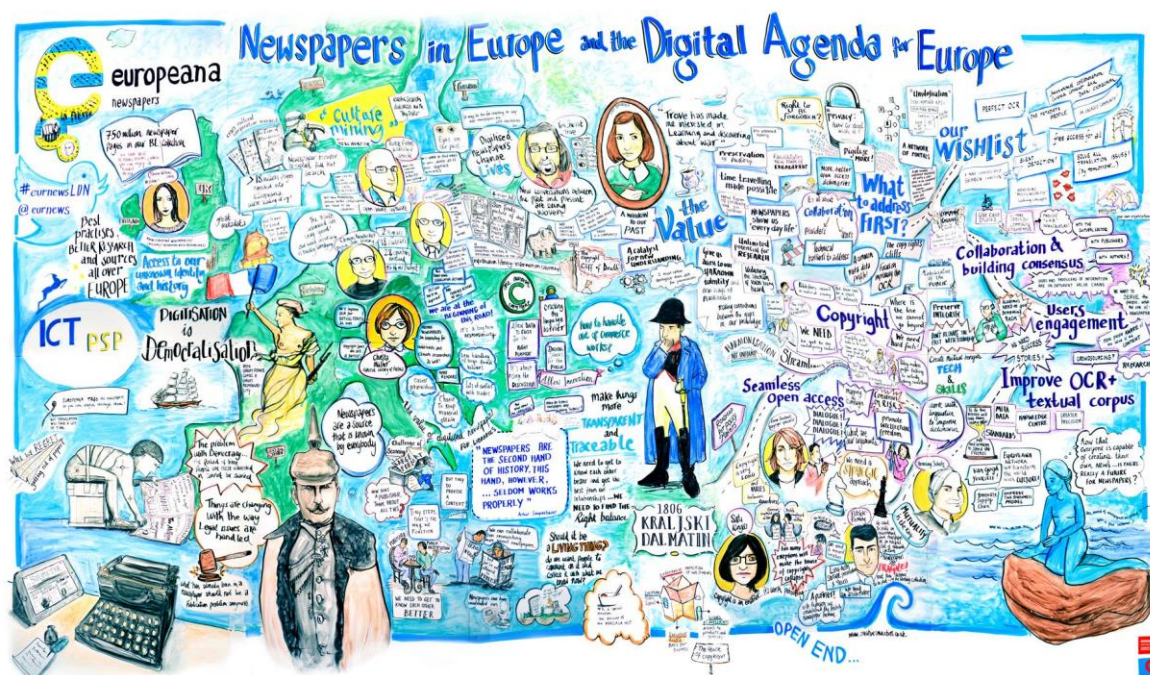
<sup>27</sup> <http://www.europeana-newspapers.eu/blog/>

<sup>28</sup> <http://www.europeana-newspapers.eu/network/>

<sup>29</sup> <http://www.ifla.org/news-media>

<sup>30</sup> <https://www.enpa.eu/>

<sup>31</sup> [http://www.europeana-newspapers.eu/wp-content/uploads/2015/05/Roadmap\\_for\\_Improving\\_Access\\_to\\_Newspapers\\_final.pdf](http://www.europeana-newspapers.eu/wp-content/uploads/2015/05/Roadmap_for_Improving_Access_to_Newspapers_final.pdf)



*Illustration of the Roadmap for Improving Access to Newspapers and the Digital Agenda for Europe*

Europeana Newspapers has been involved from the start in the Transatlantic Digitised Newspaper Symposium<sup>32</sup> (held at UCL in May 2016 and again in April 2017) and is engaged in an ongoing collaboration with one of its initiators, Ryan Cordell, principal investigator of the Viral Texts project<sup>33</sup> who can now expand the scope of the project to include European newspapers.

Last but not least, in 2016 collaborations have also been established with the DPLA, the Digital Public Library of America<sup>34</sup> regarding the workflow and best-practices for newspaper aggregation, refinement and presentation, as well as with the IIF (International Image Interoperability Framework) Interest Group for newspapers<sup>35</sup> and ICON (International Coalition for Newspapers), based at the Chicago Center for Research Libraries<sup>36</sup>.

To summarise - Europeana Newspapers is very well integrated with the major international newspaper networks and communities and continuously engaged in various activities to promote the further adoption of standards and best practices for newspaper digitisation and aggregation to Europeana.

#### 4.4.1 Advisory board

Europeana Newspapers will establish an Advisory Board to support and guide the further development of the Europeana Newspapers thematic collection based on user needs,

<sup>32</sup> <https://www.ucl.ac.uk/dh/events/archive/transatlantic>

<sup>33</sup> <http://viraltxts.org/>

<sup>34</sup> <https://dp.la/>

<sup>35</sup> <https://groups.google.com/forum/#!topic/iif-discuss/lwye6ZHkIBM>

<sup>36</sup> <http://www.crl.edu/events/framing-common-agenda-newspaper-digitization-and-preservation-icon-summit>



expectations of the scientific community and demands for reuse by e.g. creative industries, designers, artists etc.

The Advisory Board will comprise of representatives from a diverse range of backgrounds, including librarians, Digital Humanities scholars, researchers in OCR and historical document analysis, educators and designers, to make sure all relevant stakeholders are well represented and their perspectives covered and taken into account in the management of the Europeana Newspaper thematic collection.

#### 4.4.2 Events and workshops

A major event is foreseen for June 2017 to increase the use of and engagement with the Europeana Newspapers thematic collection. The Berlin Transcribathon Campus 2017 will bring together people who are passionate about transcription, enrichment and linking of various historical sources. The focus will be on sources from the First World War collected by Europeana 1914-1918 and Europeana Newspapers.

The three day event will be hosted by the Berlin State Library and organised together with Facts & Files and Europeana. The main part of the event will be the Transcribathon competition, where cross-generational and interdisciplinary teams compete with each other on the transcription and enrichment of documents from Europeana Newspapers and Europeana 1914-1918 such as letters, diaries, newspapers and photos. Participants will be drawn from the core audiences described in [4.1](#) and engaged through targeted advertising and promotion of the event via their respective channels and networks.

The teams have 40 hours of time to work and present their results. An international jury will award prizes in the following categories

- Transcription (quantity and quality)
- Enrichment (focus newspapers and other annotations)
- Enrichment (focus geolocations)
- Wikimedia-Editathon
- Talks and presentations from key user and stakeholder groups

The teams will be working within a campus environment with all technical equipment (computers, screens, digitisation equipment) and food provided to create a great and intense working atmosphere. Experts from different fields will be present and available to assist participants. In addition to the competition a number of talks, a podium discussion, screenings of WWI films and social events will take place.

#### 4.4.3 Meetings and conferences

In 2016, Europeana Newspapers participated in a number of conferences such as the IFLA News Media Conference 2016<sup>37</sup> in Hamburg, Germany, the 12th IAPR International Workshop on Document Analysis Systems 2016<sup>38</sup> in Santorini, Greece, and the 10th Language Resources and Evaluation Conference 2016<sup>39</sup> in Portorož, Slovenia. Presentations and publications resulting

---

<sup>37</sup> <http://blogs.sub.uni-hamburg.de/ifla-newsmedia/>

<sup>38</sup> <http://www.primaresearch.org/das2016/>

<sup>39</sup> <http://lrec2016.lrec-conf.org/en/>

from these can be found on the Europeana Newspapers website.<sup>40</sup>

Further presentations and publications of Europeana Newspapers are foreseen for the following 2017 scientific conferences

- DATeCH and Digitisation Days 2017, 1-2 June 2017, Göttingen, Germany<sup>41</sup>
- IFLA News Media Conference, 16-18 August 2017, Dresden, Germany<sup>42</sup>
- ICDAR 2017, 10-15 November 2017, Kyoto, Japan<sup>43</sup>

## 4.5 Key activities

### 4.5.1 Themes / Exhibitions

Every six months, a new exhibition and overarching portal theme will be presented in consecutive exhibition chapters to tell the story through a narrative in a concise, structured, and user-friendly way.

Currently planned exhibitions contain 4-6 chapters, each featuring several smaller sections devoted to particular topics or collections, and presenting up to 12 images.

The first theme will be devoted to World War I, in cooperation with Europeana 1914-1918 and in preparation of the Berlin Transcribathon Campus in June 2017.

Thanks to the broad variety of content covered by historical newspapers, they are perfectly suited to explore additional themes in collaboration with the other Europeana thematic collections, such as e.g.

- When photographs were first introduced in newspapers and the first photographers that were written about
- Newspaper articles about fashion and fashion events and trends
- Articles and reviews about artworks, artists and art fairs
- Articles and reviews about musical works and shows, composers and musicians

Other themes that will be evaluated are theater/acting, transportation/mobility and sports.

### 4.5.2 Browse entry points

Curated thematic (chronological / event / person / topic-centred) searches will guide users directly towards interesting, valuable collections or themes available from the landing page of Europeana Newspapers.

### 4.5.3 Blog

The Europeana Newspapers blog<sup>44</sup> is used to engage users with specific highlights in the collection or inform them of news and developments. It regularly publishes blog posts that

---

<sup>40</sup> <http://www.europeana-newspapers.eu/public-materials/>

<sup>41</sup> <http://ddays.digitisation.eu/datech-2017/>

<sup>42</sup> <http://www.ifla.org/news-media/conferences>

<sup>43</sup> <http://u-pat.org/ICDAR2017/>

<sup>44</sup> <http://www.europeana-newspapers.eu/blog/>

introduce either

- A subset of the collection, e.g. by a particular library<sup>45</sup>,
- A spotlight on major historical events and how they were covered in the news at that time<sup>46</sup>,
- An “on this day in history” feature<sup>47</sup>,
- A use case from researchers working with historical newspapers<sup>48</sup>

Over the course of March - November 2017, it is planned to publish at least one blog post per category per month.

#### 4.5.4 Social media

Various social media channels (see [4.4](#)) are used to publish announcements, links to new content on the Europeana Newspapers blog or website, or to engage in direct communication with users.

#### 4.5.5 User engagement activities

As specified under [4.4.2](#), a major user engagement event is planned for June 2017 in Berlin, Germany with the Berlin Transcribathon Campus.

Furthermore, data from Europeana Newspapers is reused in other user-centric activities and communities such as the Coding Da Vinci hackathon<sup>49</sup> in Germany. The possibility to join or assist in user engagement activities organised by other Europeana thematic collections will be considered.

#### 4.5.6 Activities timeline

Week	March 2017	April 2017	May 2017	June 2017	July 2017	August 2017
<b>1</b>	<b>New theme introduced</b>  Blog  Facebook and Twitter	Blog  Facebook and Twitter	Blog  Facebook and Twitter	<b>DATECH and Digitisation days</b>  Blog  Facebook and Twitter	Blog  Facebook and Twitter	Blog  Facebook and Twitter
<b>2</b>	Blog  Facebook and Twitter	Blog  Facebook and Twitter	Blog  Facebook and Twitter	Blog  Facebook and Twitter	Blog  Facebook and Twitter	Blog  Facebook and Twitter

<sup>45</sup> <http://www.europeana-newspapers.eu/category/featured-partner/>

<sup>46</sup> <http://www.europeana-newspapers.eu/category/celebrating-historic-newspapers/>

<sup>47</sup> <http://www.europeana-newspapers.eu/tag/this-day-in-history/>

<sup>48</sup> <http://www.europeana-newspapers.eu/category/interviews-with-researchers/>

<sup>49</sup> <https://codingdavinci.de/>

<b>3</b>	Blog  Facebook and Twitter	<b>Reading Europe Campaign</b>  Blog  Facebook and Twitter	Blog  Facebook and Twitter	<b>Berlin Transcribathon Campus</b>  Blog  Facebook and Twitter	Blog  Facebook and Twitter	<b>IFLA News Media Conference</b>  Blog  Facebook and Twitter
<b>4</b>	Blog  Facebook and Twitter	<b>Transatlantic newspaper summit</b>  Blog  Facebook and Twitter	Blog  Facebook and Twitter	Blog  Facebook and Twitter	Blog  Facebook and Twitter	Blog  Facebook and Twitter

## 4.6 Resources

Due to the transition from project to sustainable infrastructure and the closing of the TEL portal by CENL<sup>50</sup> (Conference of European National Libraries) in 2016, resources are currently used mainly for the migration of Europeana Newspapers from TEL to Europeana. This will implement features and functionalities provided by the TEL prototype portal on Europeana Collections.

20% of 1 FTE (full-time-equivalent) position is available from the Berlin State Library to curate, promote and further develop the thematic collection. Europeana staff is working on the migration and necessary technical developments and adaptations required for a successful migration and re-launch of Europeana Newspapers as a Europeana thematic collection (further details under [5](#)).

Google Docs is used for document sharing and collaboration. Basecamp and Assembla are also used for more specific tasks, such as communication and planning, issue management or the scheduling of technical developments. GitHub is used for the distribution of some open source software and related data from Europeana Newspapers.

## 4.7 Impact

Already at the closing of the Europeana Newspapers ICT-PSP project in March 2015, impact across a wide range of communities was observed. Looking forward, a key focus during DSI-2 will lie in the continued fostering of these in order to utilise the full impact of the re-launch of the Europeana Newspapers thematic collection on the Europeana Collections platform in 2018.

### Use statistics

Use statistics for the TEL prototype portal were tracked and analysed via Google Analytics, showing a very promising degree of user engagement:

- In 2016, there were 1,296,012 page views and 71,107 sessions recorded on the TEL newspapers site

<sup>50</sup> See the official press release at <http://www.cenl.org/wp-content/uploads/20161213-TEL-closure-press-release-final.pdf>



- The average session has lasted 16 minutes and 12 seconds. This is exceptionally high; the average for the TEL site as a whole (i.e. all its content, not just newspapers) is 4 mins and 41 seconds.
- Wikipedia is amongst the most important sources of users. While there are only around 20-25 links from Wikipedia to the TEL newspaper portal, they provide for nearly a quarter of incoming referrals.

## Research

The Europeana Newspapers collection has been used widely in research. Several datasets were released specifically targeting the needs of particular research communities, such as

- A scientific competition for Historical Newspaper Layout Analysis<sup>51</sup>, held at the ICDAR2013 conference, based on the Europeana Newspapers ground truth dataset<sup>52</sup>.
- Evaluation of Named Entity Recognition for historical sources, based on the Europeana Newspapers named entity corpora<sup>53</sup>.
- In collaboration with Europeana Research, a full dump of the public domain OCRed text from the Europeana Newspapers collection has been released<sup>54</sup>.
- A workshop<sup>55</sup> was held in 2015 to collect ideas on how to improve access to the data and find tools with potential to improve the research workflow.
- Sven Charleer, Visual Learning Analytics / UX researcher at KU Leuven developed a multi-device interactive prototype<sup>56</sup> that can serve as a first example for what creative designers can come up with based on the Europeana Newspapers collection.
- The European Research Academy in Bolzano, Italy, ran a project<sup>57</sup> for the scholarly analysis of historical newspapers from South-Tyrol based on the OCRed texts of newspapers from the Tessmann library produced by Europeana Newspapers.

## Education

Some examples for reuse of the Europeana Newspapers collection in education are provided by

- Amélia Del Rosario Sanz Cabrerizo<sup>58</sup>, a researcher and professor at Universidad Complutense de Madrid, who uses historical newspapers in her class on comparative literature and cyberculture<sup>59</sup>.
- Eva Maria Naimer, who provided insights on new teaching methods enabled through digital historical newspapers used in her second grade class alongside an Europeana Newspapers Information Day<sup>60</sup>; a book describes more thoroughly the use of historical newspapers for teaching history in schools<sup>61</sup>.
- Steven Stegers from the European Association of History Educators (EUROCLIO) highlights the rich possibilities for education enabled by combining digitised historical newspapers with other cultural heritage materials in his Europeana blog post<sup>62</sup>.

---

<sup>51</sup> <http://www.primaresearch.org/HNLA2013/>

<sup>52</sup> <http://www.primaresearch.org/datasets/ENP>

<sup>53</sup> <https://github.com/EuropeanaNewspapers/ner-corpora>

<sup>54</sup> <http://research.europeana.eu/itemtype/newspapers>

<sup>55</sup> <http://www.europeana-newspapers.eu/wanted-historic-newspaper-researchers/>

<sup>56</sup> <http://research.europeana.eu/blogpost/developing-an-interactive-visualisation-of-europe-s-historic-newspapers>

<sup>57</sup> <http://www.eurac.edu/de/research/projects/Pages/projectdetails.aspx?pid=11263>

<sup>58</sup> <http://www.europeana-newspapers.eu/qa-with-newspaper-researchers-amelia-sanz-cabrerizo/>

<sup>59</sup> <http://eprints.ucm.es/32725/>; <http://eprints.ucm.es/32726/>

<sup>60</sup> [http://www.slideshare.net/Europeana\\_Newspapers/enponbinfodaynaimer](http://www.slideshare.net/Europeana_Newspapers/enponbinfodaynaimer)

<sup>61</sup> [http://www.v-r.de/de/zeitungen\\_von\\_gestern\\_fuer\\_das\\_lernen\\_von\\_morgen/t-0/1039315/](http://www.v-r.de/de/zeitungen_von_gestern_fuer_das_lernen_von_morgen/t-0/1039315/)

<sup>62</sup> <http://pro.europeana.eu/blogpost/how-news-once-travelled>

### Libraries and archives

Libraries and archives holding digitised historical newspaper collections have explored novel ways of giving access and creating engagement with their newspaper collections thanks to Europeana Newspapers, as exemplified by

- The State and University Library Hamburg who launched a mobile app and a series of blog posts reusing World War One images and historical newspapers<sup>63</sup>.
- The National Library of France has launched a new portal<sup>64</sup> for their historical newspaper collection offering a wide range of themes and services.
- The launch of the Tasman digital<sup>65</sup> portal of the Dr. Friedrich Tasman library, where about half of the OCRed content provided was produced via Europeana Newspapers.

### SMEs

The historical newspaper data made available via Europeana Newspapers has already spawned some commercial interest and led to the following exploitation activities by SMEs

- The Bulgarian company Onto text AD who specialize on technologies for semantic knowledge extraction and management have used data from Europeana Newspapers for building a semantic demonstrator<sup>66</sup> in the frame of the Europeana Food & Drink project.
- In collaboration with Europeana Newspapers partner State and University Library Hamburg, the German digitisation company Content Conversion Specialists (CCS) has created the highly successful mobile app “Weltbrand” which presents images and newspapers from Hamburg during WWI<sup>67</sup> - the app was also prominently featured in printed news<sup>68</sup>.
- Following up the digitisation of historical newspapers from Berlin by the Berlin State Library under the Europeana Newspapers project, the Berlin State Library contracted the company Zeuschel to develop a specific newspaper module for the open source Goobi<sup>69</sup> software that is widely used throughout Germany and abroad to manage cultural heritage digitisation workflows.

### Other European projects and infrastructures

- In 2015, Europeana Newspapers has started collaborating with the European Data Infrastructure EUDAT<sup>70</sup>, with the aim of exposing and improving the Europeana Newspapers full-text corpus for more data driven usage (i.e. large-scale data analysis of the whole corpus) via the use of EUDAT infrastructure and services such as *B2SAFE* and *B2FIND*. A first client side implementation was done in 2016<sup>71</sup>. Following an initial pilot phase<sup>72</sup>, Europeana Newspapers was selected as one of the flagship pilot projects of EUDAT in 2017.
- There is ongoing collaboration between Europeana Newspapers and CLARIN<sup>73</sup> regarding the ingestion of the newspaper full-text in the infrastructure of CLARIN.

---

<sup>63</sup> <https://blog.sub.uni-hamburg.de/?p=14237>

<sup>64</sup> <http://www.retronews.fr/>

<sup>65</sup> <http://digital.tessmann.it/>

<sup>66</sup> <http://foodanddrinkeurope.eu/wp-content/uploads/2016/01/D3.19-Semantic-Demonstrator-Specification.pdf>

<sup>67</sup> <https://blog.sub.uni-hamburg.de/?p=14237>

<sup>68</sup> <http://www.abendblatt.de/kultur-live/article129947804/Weltbrand-1914-wie-der-Krieg-begann.html>

<sup>69</sup> <https://github.com/kitodo/kitodo-production/issues/41>

<sup>70</sup> <https://www.eudat.eu/>

<sup>71</sup> <https://github.com/nfreire/EUDAT-B2SHARE>

<sup>72</sup> <https://www.eudat.eu/enriching-europeana-newspapers-data-pilot>

<sup>73</sup> <https://www.clarin.eu/>

## 5. Product design and development

### 5.1 Scope of product development

The development of the Europeana Newspapers Collection services is essentially a migration and redevelopment project. What was developed in the context of the Europeana Newspapers project and implemented on the The European Library platform needs to be<sup>74</sup> re-implemented on the Europeana platform. In some cases it also needs to be further developed<sup>75</sup>. The data aggregated in the Europeana Newspapers project likewise needs to be migrated to a Europeana database.

The objectives of the migration and redevelopment are to create:

- A discovery site for end-users to search, browse and view newspapers. This will take the form of a thematic collection on Europeana Collections (cp. thematic collections for Art, Music and Fashion) and be based on a Europeana Newspapers API.
- A Europeana Newspapers API and newspapers data dumps for developers to access the newspapers (metadata, full text, possibly images) programmatically and in bulk. The API documentation and data dumps will be documented and featured on Europeana Labs and/or Europeana Research.
- Tool support for Europeana data partners and Europeana data officers to update existing newspapers datasets and provide new newspapers datasets for publication in the Europeana Newspapers API and the Europeana Newspapers Collections.

When redeveloping the software the goal is to, where needed and resources permitting, improve on what was developed in Europeana Newspapers. Planned improvements include:

- Re-develop the Newspapers API to be fully IIIF<sup>76</sup> compliant
- In as much as possible make changes to the UI and interaction based on the usability studies performed in the Europeana Newspapers project
- Create a tool that can update existing and publish new newspapers datasets that does not require developer intervention

To fulfil the objectives will require development beyond the time period of DSI-2 (see roadmap below).

---

<sup>74</sup> As The European Library ceased to operate on the 31st of December 2016.

<sup>75</sup> One example being the Newspapers API developed in the project and which was never really completed.

<sup>76</sup> <http://iiif.io/>





Mockup for the Europeana Newspapers thematic collection landing page

## 5.2 Evaluation

The objectives and goals of the development must be reflected in clearly defined metrics, performance indicators that will allow us to evaluate whether the objectives and goals have truly been met.

These performance indicators will be:

- User satisfaction of the Europeana Newspapers Collections and the Newspapers API. This will be measured primarily via user surveys. The goal is a Net Promoter Score<sup>77</sup> of 20+ and a 60+% proportion of users evaluating the services as 4 or 5 on a scale of 1-5.
  - Based on an analysis of the users and usage of the The European Library Newspapers discovery our initial hypothesis is that users performing formal academic (or private) research will be the primary target audience we will cater the Europeana Newspapers Collections too. This hypothesis will be tested early in the development process.
- Uptime/Stability of the Europeana Newspapers Collections and the Newspapers API. This

<sup>77</sup> [https://en.wikipedia.org/wiki/Net\\_Promoter](https://en.wikipedia.org/wiki/Net_Promoter)

will be measured by Europeana's already existing service monitoring system<sup>78</sup>. The goal is to have an uptime of 99.5% or higher.

- Software maintainability of the Europeana Newspapers Collections and the Newspapers API. This will be measured by SonarQube<sup>79</sup>. The goal is a maintainability rating<sup>80</sup> of a minimum of B.

These performance indicators are aligned with how other Europeana products and services are evaluated and may also be reflected in the Europeana Business Plans (from 2018 onwards at the earliest).

## 5.3 Product development methodology and policies

The Europeana Newspapers services will be developed using a combination of Scrum<sup>81</sup> and User Centred Design<sup>82</sup>. This is the standard way product development is performed at Europeana. The Scrum development team will have a core of a product owner (responsible for prioritisation and vision), two near full-time developers with other developers, designers and data specialists joining the team for shorter periods of time and for specific purposes.

As a matter of policy Europeana will only use software components that are open source and compatible with the European Union Public License (EUPL v1.1)<sup>83</sup>. All source code produced will equally be so licensed. The source code will be managed in GitHub<sup>84</sup> and be publicly available and open to other developers to use or contribute to.

## 5.4 Roadmap

Development is planned to begin in July 2017. A tentative release for a full production Newspapers Services system is set to mid 2018.

Below is a roadmap for the development. Note that the roadmap is subject to change in regards to exact times of release. The roadmap for the year of 2018 is very coarse. It can and will become more detailed as we progress on the part of the roadmap covering 2017.

July 2017	August	September	October	November	December
Setup of development environments		Newspapers API dev	Newspapers API dev	Configuration of new free text search server	Public alpha of Newspapers API
Technical design and specification		Create clickable wireframes of the UI	Create high fidelity clickable wireframes of the	Newspaper API dev	Public alpha of Newspapers Collections

<sup>78</sup> <http://status.europeana.eu/>

<sup>79</sup> <https://www.sonarqube.org/>

<sup>80</sup> <https://docs.sonarqube.org/display/SONAR/Metric+Definitions#MetricDefinitions-Maintainability>

<sup>81</sup> [https://en.wikipedia.org/wiki/Scrum\\_\(software\\_development\)](https://en.wikipedia.org/wiki/Scrum_(software_development))

<sup>82</sup> [https://en.wikipedia.org/wiki/User-centered\\_design](https://en.wikipedia.org/wiki/User-centered_design)

<sup>83</sup> [https://joinup.ec.europa.eu/community/eupl/og\\_page/european-union-public-licence-eupl-v1.1](https://joinup.ec.europa.eu/community/eupl/og_page/european-union-public-licence-eupl-v1.1)

<sup>84</sup> <https://github.com/europeana>

## D2.4: EUROPEANA NEWSPAPERS COLLECTIONS PLAN

<p>Review of existing user research<sup>85</sup></p> <p>Review all Europeana Newspaper datasets</p> <p>Backlog of user stories ready to start development</p>	<p>Reingest Europeana Newspapers datasets if necessary</p>	<p>UI</p> <p>Newspapers viewer and front-end dev</p> <p>Ruby client dev</p>	<p>Newspapers viewer and front-end dev</p> <p>Ruby client dev</p>	
---	--	---	---	--

January 2018	February	March	April	May	June
<p>Alpha testing and evaluation</p> <p>Modifications based on the testing results</p> <p>Addition of all features necessary for beta release of Newspapers API and Collections</p> <p>Technical design of a Newspapers ingestion service</p> <p>Release of beta versions of the Newspapers API and Collections</p>		<p>Beta testing and evaluation incl. load and performance testing</p> <p>Modifications based on the testing results</p> <p>Addition of all features necessary for first production release of the Newspapers API and Collections</p> <p>Stabilisation (technical debt reduction) of the Newspapers API and Collections</p> <p>Performance optimisation of the Newspapers API and Collections</p> <p>Development of a Newspapers ingestion service</p>		<p>Development and release of Newspapers API 1.0</p> <p>Development and beta release of a Newspapers ingestion service</p>	

<sup>85</sup> For example [http://www.europeana-newspapers.eu/wp-content/uploads/2015/05/D4.6-7\\_Interface\\_Development\\_and\\_Usability\\_Testing\\_1.0.pdf](http://www.europeana-newspapers.eu/wp-content/uploads/2015/05/D4.6-7_Interface_Development_and_Usability_Testing_1.0.pdf)