

Facilitating Access and Reuse of Research Materials: the Case of The European Library

Nuno Freire¹
The European Library

Abstract. *The European Library* provides access to research materials from the collections of Europe's national and research libraries, representing members from 46 countries. This paper presents the current status, on-going work, and future plans of the resource dissemination services provided by *The European Library*, covering resources such as national bibliographies, digital collections, full text collections, its access portal and API, open linked data publication, and integration in digital humanities infrastructures. In the coming years, *The European Library* will work to provide the means and tools for digital humanities researchers to easily use research materials from libraries in their research activities.

Keywords. digital libraries, full text, digital humanities, portals, research infrastructures.

Introduction

The European Library provides access to research materials existing in the collections of Europe's national and research libraries. The European Library's partnership with libraries extends beyond the European Union, covering all the Council of Europe member states. It is a pan-European service in the widest sense.

The European Library centres its activities in the provision of services based on exploiting the centralization of pan-European bibliographic data and digital content. Its most visible service is its portal, which is designed with researchers as the target user. The portal provides value to the libraries by showcasing their collections of research materials to a wide audience, and drawing new customers to their repositories. The homepage of the portal is shown in Figure 1. The centralization of resources from European libraries also allows the provision of other services that promote the reuse of these digital resources in many contexts.

This paper starts with a presentation of *The European Library*, and follows with a description of the main digital resources that underlie its service. These include national bibliographies, bibliographic databases, digital objects and full text materials.

¹ Corresponding Author: Senior researcher, The European Library, Europeana Foundation, Willem-Alexanderhof 5, 2509 LK The Hague, Netherlands; Email: nfreire@gmail.com

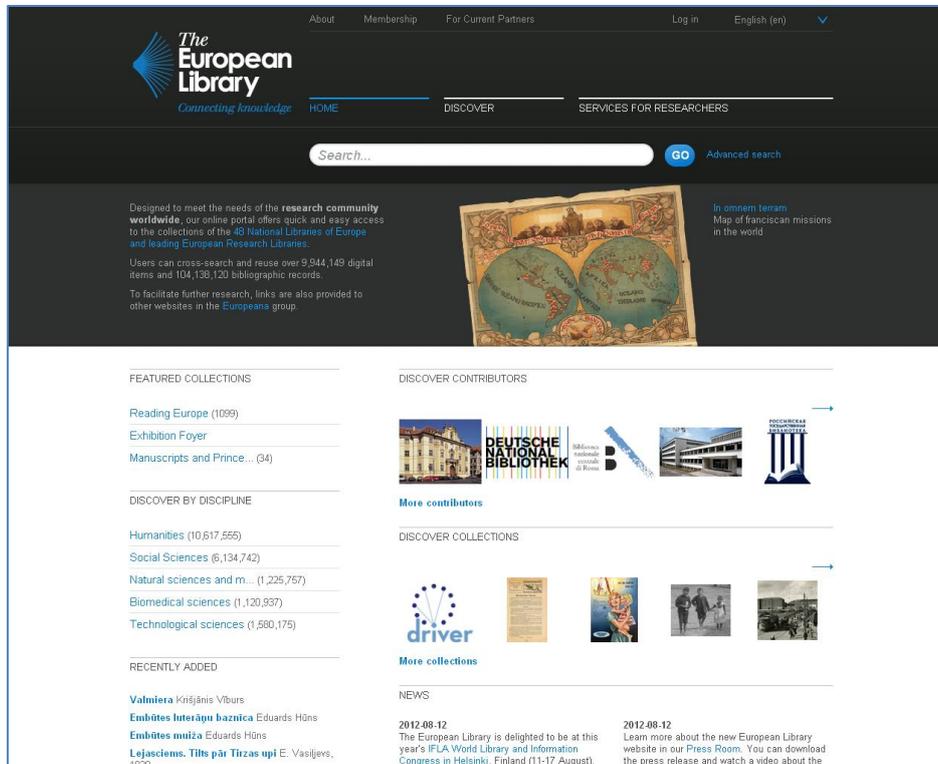


Figure 1. The European Library portal (<http://www.theeuropeanlibrary.org>).

1. The Origin and History of *The European Library*

The European Library of today has evolved from a number of earlier projects. Its starting point was in 1997, when Project GABRIEL (Gateway and Bridge to Europe's National Libraries) set out to establish a joint web portal of European national libraries.

At a basic level, the portal provided information about each library's collections and access to their online public access catalogues, together with their associated web services.

Building on GABRIEL's success, the TEL (The European Library) Project (2001–2004) went on to create a framework for access to key national and deposit collections within Europe. The project partners included the CENL, the national libraries of Finland, Germany, Italy (Florence), the Netherlands, Portugal, Slovenia, Switzerland and the United Kingdom, together with the Central Institute for the Union Catalogue of Italian Libraries (ICCU). Its achievement was to set down policies and to undertake the technical ground-work for a sustainable pan-European digital library, based on the distributed collections of the participating libraries. The overall outcome of the project was agreement to take it forward as an operational service under the auspices of CENL and hosted by the National Library of the Netherlands in The Hague, The Netherlands.

The European Library was launched as an operational service in March 2005. It aggregates the digital collections of Europe's national and research libraries and brings together the catalogues of its partner libraries to create a union catalogue for Europe.

Between 2005 and 2007, the TEL-ME-MOR project helped to incorporate 10 more national libraries from new European Union member states as full partners of The European Library. By the beginning of 2008, a further nine national libraries had joined the service. *The European Library* was further enlarged by the EDL project, during which national libraries continued to join The European Library. The project also focused on multilingualism, undertook the first steps towards a European Metadata Registry and created a roadmap for potential digitization efforts in the national libraries.

The European Library provided much of the organisational structure and expertise required to launch Europeana. When the European Parliament called for Europeana to be established, *The European Library* was asked to apply to run the project under the eContentplus programme. Work began on a prototype in 2007 and Europeana was launched in November 2008.

Europeana now operates independently but *The European Library* continues to work closely with Europeana. In project Europeana Libraries, which ran from 2010 until 2012, the library aggregation model for Europeana has been established, where *The European Library* had the mission to become the library-domain aggregator service for Europeana. This aggregation service provided by *The European Library* is available for libraries across Europe.

Currently, *The European Library* is led by three major library associations from Europe:

- The Conference of European National Librarians² (CENL) - The CENL foundation promotes the role of national libraries in Europe, in particular in respect of their responsibilities for maintaining the national cultural heritage and ensuring the accessibility of knowledge. Members of CENL are the national librarians of all Member States of the Council of Europe. The conference currently consists of 49 members from 46 European countries.
- Association of European Research Libraries³ (LIBER) – LIBER is the main network for research libraries in Europe, comprising more than 400 national, university and other libraries from more than 40 countries.
- Consortium of European Research Libraries⁴ (CERL) - CERL is a library consortium focused on the European printed heritage from the hand-press period (up to c. 1830). It aims to the sharing of resources and expertise between research libraries.

These three associations include in its members over 400 national and research libraries. For a library to be a member in *The European Library*, it is not required a membership in CENL, LIBER or CERL.

The European Library has launched a new web portal in 2012, aiming to open up new possibilities for academic research, and to build on The European Library's already unique position of giving researchers free, online access to research materials from Europe's national and research libraries.

² <http://web3.nlib.ee/cenl/>

³ <http://www.libereurope.eu>

⁴ <http://www.cerl.org>

2. Resources Aggregated by *The European Library*

This section outlines the main research resources available through *The European Library*. It also describes the unique characteristics of these resources, and summarises our recent and on-going activities.

2.1. Bibliographic Resources

The European Library holds several millions of bibliographic records from the national and leading research libraries of Europe. This bibliographic universe is organised into *collections* – groups of objects clustered around a topic or theme. The concept of collections is an extremely important one for *The European Library* because it represents a curatorial attempt at categorising content. Hundreds of collections are available and together they cover all the general areas of knowledge.

The National Bibliographies, one of the main bibliographic data sources in each country, are a key resource for *The European Library* users. Their purpose is to list every publication in a country, under the auspices of a national library or other government agency. Depending on the country, all publishers will need to send a copy of every published work to the national legal deposit, or in other countries, a national organisation will need to collect all publications. Given that the publisher domain is very heterogeneous and that thousands of publishers might exist in a country, national bibliographies are effectively the single point of reference with which to comprehensively identify all the publications in a country. In Europe, national bibliographies are typically created and maintained by national libraries. Whenever a book is published in a country, it is recorded in the corresponding national library catalogue from where the national bibliography is derived.

The European Library also offers access to collections of digital resources. These include a diverse wealth of content, such as e-theses, books, manuscripts and films, including many sizeable collections grouped around specific themes, topics or notable personalities. Examples of these are: digitised letters from prominent thinkers such as philosopher Immanuel Kant, pamphlets detailing the history of the Low Countries from 1490 to 1800, and images covering the papers, photographs and correspondence of prominent physicists such as Nobel prize winner Erwin Schrödinger. There are also digitised manuscripts created in European parts of the Ottoman Empire.

Currently, *The European Library* holds approximately 112 million bibliographic records in its centralized repository. This number is constantly increasing, as part of its data aggregation role in the Europeana context, and more national and research libraries' catalogues are included in the centralized repository. By the end of 2013, the total bibliographic universe of *The European Library* is expected to be approximately 200 million records.

2.2. Textual Resources

The European Library hosts a growing centralized index of textual resources. It currently contains over 24 million pages of full-text content, originating from 14 national libraries.

These textual resources were created mostly from Optical Character Recognition performed during digitization projects at the national libraries. The quality of the text

varies, depending on the quality of the original material, and the use of special fonts, some of which remain challenging for even modern readers.

That said, the size of the full-text collection, and the fact that it is relatively heterogeneous in terms of types of materials, languages and publication period, makes it a relevant resource for digital humanities research. An overview of the available resources is provided in Table 1.

Country of origin	Material type	Pages	Temporal coverage	Languages
Austria	Newspapers, governmental material	534.000	1862 – 1925	German
Czech Republic	Books, newspapers	2.579.511	1800 – 1989	Czech, German
Estonia	Newspapers, journals	713.933	1821 – 1940	Estonian
France	Books, periodicals	8.242.908	1650 – 1930	French (some others)
Hungary	Periodicals, newspapers, journals, books, monographs, pamphlets	237.914	1590 – 1992	Hungarian, Latin, English, German
Iceland	Newspapers, journals	5.727.149	1773 – 2002	Icelandic, Faroese, Greenlandic
Latvia	Newspapers, books	195.075	1900 – 1952	German, Latvian
Lithuania	Newspapers	125.477	1904 – 1940	Lithuanian
Norway	Books, journals	1.600.000	By authors dead for more than 70 years	Norwegian (others)
Poland	Newspapers, books	436.198	Before 1939	Polish, German, Czech, Ukrainian, Belarusian, Yiddish
Slovakia	Newspapers	185.000	Before 1918	Slovak, Hungarian, German
Slovenia	Newspapers, books, journals	328.502	1500 – 1945	Slovenian
Spain	Newspapers, books	3.033.525	17 th – 19 th Century	Spanish
Sweden	Newspapers, books, journals, printed ephemera	253.653	Until the 20 th century	Swedish

Table 1. Textual resources currently indexed at *The European Library*.

These textual resources will be expanded during 2013, thanks to the *Europeana Newspapers* project⁵. In this project, a group of 17 European institutions are working to provide more than 18 million newspaper pages for *The European Library* and Europeana. Each library participating in the project will distribute digitized newspapers and full-texts that are in the public domain and free of any legal restrictions.

3. Resource Dissemination Services

Although *The European Library* service was founded as a traditional library service, focused on making the library resources available through traditional online catalogue features, it is now aiming to facilitate the use of these resources in order to improve existing services, enable the development of new ones, or provide a good setting for research. This section presents the main services and activities undertaken by *The European Library*.

3.1. The European Library Portal

The European Library's portal main focus is to provide access to the collections of the European national and research libraries. It provides a web interface for searching and browsing through libraries, collections, bibliographic records, digital objects and full text contents. A new version of the portal was launched in May 2012, which included a complete functional redesign, focused on a shift towards functionalities for digital humanities researchers.

While previous versions of the portal were based on the distributed search paradigm, this new version is based on the centralization of the digital resources, which enables much functionality that nowadays web users expect in resource discovery portals.

Data mining technologies are being applied for automatic linkage of bibliographic data across countries and libraries (some examples are shown in Figure 2). On-going work is focusing on the linkage of the main entities used to describe bibliographic resources: persons, organizations, locations, historical periods, and subjects, in order to enable better functionality in the portal. This on-going work, once completed, will allow the exploration of an author's work across time and space. Users will be able to locate different editions of a work throughout Europe, browse all catalogues with a unified subject system, and browse across historical periods.

Access to the bibliographic resources under a unified subject system is also a core functionality of the portal. Subject information is frequently used in bibliographic data and very relevant for the portal functionality, given the multilingual nature of *The European Library*. However, unifying subject systems is great challenge because of the diversity of languages and knowledge organization systems in use across European libraries. Unified subject access is currently based in the results of the project Multilingual Access to Subjects⁶ (MACS). The MACS project has produced manual

⁵ <http://www.europeana-newspapers.eu/>

⁶ <http://www.d-nb.de/eng/wir/projekte/macs.htm>

alignments between three major systems: the Library of Congress Subject Heading⁷ (LCSH), the Répertoire d'autorité-matière encyclopédique et alphabétique unifié⁸ (RAMEAU) and Schlagwortnormdatei⁹ (SWD), covering the English, French and German languages. Ongoing work is targeting wider coverage of subject systems, by exploring the alignment of language independent subject classification systems (such as Dewey Decimal Classification¹⁰ and Universal Decimal Classification¹¹), and by researching semi-automatic subject alignment techniques.

Although subject information in our datasets is based on the above mentioned library subject systems, the unification of subject access is being implemented based on the R&D subject indexing scheme of CERIF.

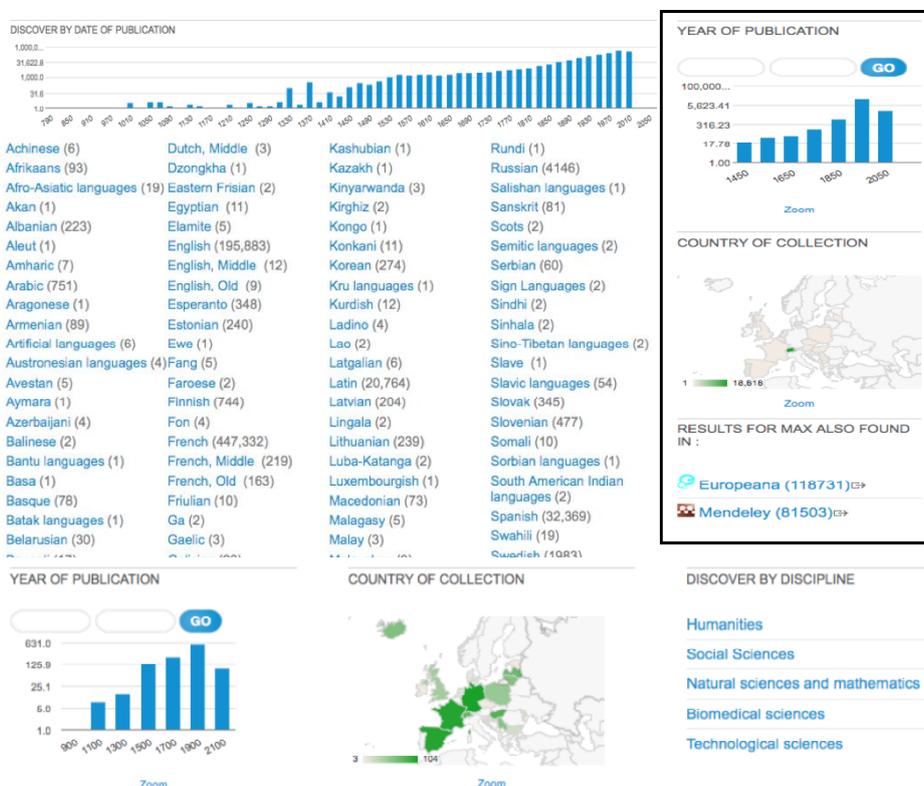


Figure 2. Mining bibliographic data at *The European Library's* portal.

⁷ <http://id.loc.gov/authorities/>

⁸ <http://rameau.bnf.fr/>

⁹ <http://www.d-nb.de/standardisierung/normdateien/swd.htm>

¹⁰ <http://dewey.info/>

¹¹ <http://www.udcc.org/about.htm>

3.2. *The European Library API*

The European Library OpenSearch¹² API allows the development of applications by third parties, which allow the search and display of *The European Library*'s collections, bibliographic data and previews. Along with standard keyword searching, the API also offers faceted searching. Currently, the OpenSearch API is available for non-commercial use to all members of *The European Library*.

We are also working towards making the resources hosted at *The European Library* available through third party applications in the research domain, such as Mendeley¹³ and the Summon discovery service¹⁴.

3.3. *Aggregation for Europeana*

Europeana¹⁵ provides access to digital content from libraries, museums, archives and audio-visual collections across Europe. Data about its digital contents follows the Europeana Data Model [1] (EDM) which is an open, cross-domain, semantic web-based framework for structuring the data about the digital content that Europeana ingests, manages, publishes for end-user access, and also publishes as open data. This data model allows Europeana to keep the richness of domain-level metadata and facilitates the participation of Europeana in the semantic web.

The European Library conducts activities within the library domain, involving EDM, in order to promote the representation of library specific information in EDM alongside authoritative and curated information from other domains. Of particular relevance is analysis of how bibliographic data (e.g. MARC21, UNIMARC, MODS, METS and Dublin Core-based application profiles) can be aligned to EDM, since in its role as aggregator for Europeana, *The European Library* supports all these bibliographic data formats for conversion and delivery to Europeana, according to EDM.

3.4. *Intellectual Property Rights Infrastructures*

The European Library's bibliographic data is enabling an innovative approach towards intellectual property rights clearance processes for supporting mass digitization projects. In the ARROW rights infrastructure¹⁶ (Accessible Registries of Rights Information and Orphan Works), a framework is being established to assist in determining the rights status of works.

ARROW provides automated ways to clear the rights of the books to be digitised. This rights clearance process is time consuming, since a library has to go through the following steps for each book:

- To identify the underlying work incorporated in the book to be digitised;
- To find out if the underlying work is in the public domain or in copyright, and whether it is an orphan work or out-of-print;

¹² <http://www.opensearch.org/>

¹³ <http://www.mendeley.com>

¹⁴ <http://www.proquest.com/>

¹⁵ <http://www.europeana.eu>

¹⁶ <http://www.arrow-net.eu/>

- To clearly describe the use that is requested for the book, such as digitisation for preservation, electronic document delivery, etc.;
- To identify the rights holder(s) or their agent, such as a collecting society;
- To seek the appropriate permission, if necessary.

The successful completion of this process depends on the availability of existing bibliographic and rights data, and national bibliographies are effectively the single point of reference with which to identify all the publications of an intellectual work. In this infrastructure, *The European Library* provides web services for the processing and use of national bibliographies during rights clearance processes. The service from *The European Library* is thereby a bibliographic record clustering service for national bibliographies. The service provides a set of related publications from the national bibliographies for a deeper analysis in a rights clearance process.

3.5. Linked Open Data

The European Library is currently working in its open linked data prototype. This prototype aims to make openly available the results of the data mining processing applied while aggregating library catalogues, digital libraries and open repositories. This process aims to establish links between entities defined in the Functional Requirements for Bibliographic Records [2] (FRBR) such as:

- Linking between the FRBR group 1 entities, particularly *work* and *manifestation*, across libraries.
- Linking between FRBR expressions within the same country (a result of the processing for the ARROW rights infrastructure)
- Linking of FRBR group 2 entities *person* and *corporate body*, between the aggregated bibliographic data and VIAF [3] (when they are not already linked at the source).
- Linking of the FRBR group 3 entity *concept*, across the aggregated bibliographic data sources, by exploiting the results of the MACS project.
- Linking of the FRBR group 3 entity *place* between the aggregated bibliographic data and Geonames [4] (a geographic ontology available as open data).

The European library's open data prototype aims to make this data enrichment available in a way that complements the ongoing activities of libraries with open data.

3.6. Digital Humanities Research Infrastructures

In order to provide the means and tools for digital humanities researchers to exploit the research materials held by libraries, a new research infrastructure, Europeana Research, will be created by extending the currently existing portal of The European Library.

Early stages of the project will analyse how academic users locate data and how they perceive the value of the content within Europeana. This analysis will be the basis of the content strategy of the Europeana Research platform, and will also provide the understanding of scholarly workflows which will be supported by Europeana Research. This analysis will be carried out jointly with the DARIAH network of arts and humanities researchers, and with the Council of European Social Science Data Archives (CESSDA). Academics working in these domains will be the most fertile

exploiters of libraries research materials, and therefore they have key roles to play in shaping the future of Europeana Research.

This analysis will allow the identification of tools that allow researchers to manipulate and exploit research materials in innovative ways. The project will therefore develop a suite of tools that allows scholars to interact with the content that they require from Europeana Research. The areas to be approached are:

- Accessing and analysing big data - permitting scholars to download, manipulate and analyse large data sets.
- Annotation - allowing researchers to annotate documents and to share these annotations
- Transcription - allowing users to transcribe and interpret documents
- Discovery and access - ensuring that services are tailored so that research material is discoverable by the scholarly community, possibly with integration in other research infrastructures in the field of digital humanities.

The creation of statistical datasets, resulting from data analysis and mining conducted by *The European Library*, is also being considered. Although no concrete plans are ready at this time, we expect these kinds of data sets to emerge as a result of our involvement with the digital humanities research community.

4. Conclusion

This paper presented *The European Library* and the main digital resources that underlie its services. *The European Library* provides services that add value to resources made available by libraries, by exploring the possibilities given by centralization of data and resources.

The European Library is entering a new phase as national and research libraries are coming together in a partnership to bring libraries' research collections to a worldwide audience. *The European Library* aims to support libraries to play a pivotal role in Europe's research agenda, supporting the growth strategy of the European Commission, which highlights higher education as key to fostering innovation, employment and long-term economic prospects. These policy developments present exciting opportunities for national and research libraries to work with Europeana to develop a research platform to support innovation and growth.

References

- [1] Europeana v1.0, The EDM Specifications V5.2.3. Online at <http://pro.europeana.eu/web/guest/edm-documentation>
- [2] IFLA Study Group on the Functional Requirements for Bibliographic Records. Functional requirements for bibliographic records: final report. München: K.G. Saur. UBCIM publications; new series, vol. 19. (1998) ISBN 3-598-11382-X. <www.ifla.org/VII/s13/frbr/frbr.pdf>
- [3] R. Bennett, C. Hengel-Dittrich, E. O'Neill, B.B. Tillett. VIAF (Virtual International Authority File): Linking Die Deutsche Bibliothek and Library of Congress Name Authority Files. World Library and Information Congress: 72nd IFLA General Conference and Council (2006)
- [4] B. Vatant, M. Wick, Geonames Ontology (2006). Online at <http://www.geonames.org/ontology/>