Bringing Research Libraries into Europeana: Establishing a Library-Domain Aggregator

Sally Chambers

The European Library, Koninklijke Bibliotheek/National Library of the Netherlands, PO Box 90407, 2509 LK The Hague, The Netherlands, sally.chambers@kb.nl

Wouter Schallier

LIBER, Koninklijke Bibliotheek/National Library of the Netherlands, PO Box 90407, 2509 LK The Hague, The Netherlands, wouter.schallier@kb.nl

Abstract

Europeana's future strategy requires a domain-level aggregator of library content, positioned alongside other domain aggregators for museums, archives and audio-visual collections. To enable this, LIBER (Association of European Research Libraries), CENL (Conference of European National Librarians) and CERL (Consortium of European Research Libraries) have joined forces to submit the 'Europeana Libraries' project proposal to the European Commission in the context of the CIP-ICT PSP-2010-4 'Digital Libraries' programme in early Summer 2010. This paper outlines the vision of the 'Europeana Libraries' project, which will enable The European Library to become the library-domain aggregator for Europeana.

Key Words: LIBER; The European Library; CERL; Europeana; digital collections; aggregation

Introduction

The mission of Europeana is to enable people to explore the digital resources of Europe's museums, libraries, archives and audio-visual collections. By

Summer 2010, Europeana will be launched as an operational service which aims to give access to ten million items of Europe's cultural and scientific heritage through Europeana.eu. Europe's national, research and public libraries have a significant role to play in the realisation of the Europeana vision.
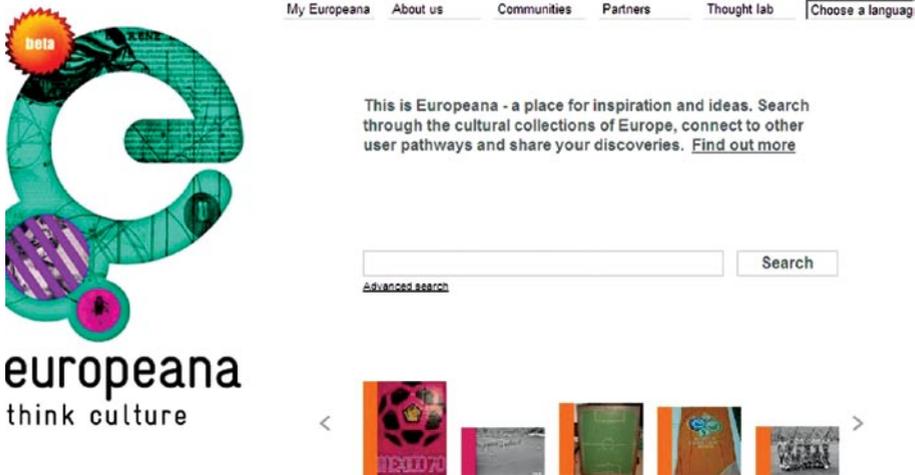
As outlined in Europeana's content strategy, there are thousands of cultural and scientific institutions in Europe with content and collections that are of interest for Europeana. However, it is not sustainable for Europeana to work with all of these institutions directly. A content aggregation model is therefore of crucial importance in enabling Europeana to reach its objectives.

Europeana's future strategy requires a domain-level aggregator of library content, positioned alongside other domain aggregators for museums, archives and audio-visual collections. To enable this, LIBER (Association of European Research Libraries), CENL (Conference of European National Librarians) and CERL (Consortium of European Research Libraries) have joined forces to submit the 'Europeana Libraries' project proposal to the European Commission in the context of the CIP-ICT PSP-2010-4 'Digital Libraries' programme in early Summer 2010. This paper outlines the vision of the 'Europeana Libraries' project, which will enable The European Library to become the library-domain aggregator for Europeana.

What is Europeana?

On 30 September 2005 the European Commission published the i2010: communication on digital libraries, where it announced its strategy to promote and support the creation of a European digital library. This European digital library, Europeana, provides access to the digital resources of libraries, museums, archives and audio-visual archives across Europe. Europeana.eu went live on 20 November 2008, launched by Viviane Reding, European Commissioner for Information Society and Media. In Summer 2010, the first operational version of Europeana will be launched, and will include over ten million digital objects from museums, libraries, archives and audio-visual archives from across Europe.

Fig. 1: Europeana.eu.

What is The European Library?

The European Library provides access to the digital and bibliographic collections of Europe's national libraries. The 49 European national libraries participating in The European Library are all members of the Conference of European National Librarians (CENL), a foundation aiming at increasing and reinforcing the role of national libraries in Europe. Members of CENL are the national librarians of the Council of Europe member states. This means that The European Library's partnership extends beyond the European Union, so it is pan-European in the widest sense. The operational service of The European Library is funded and supported by CENL. The European Library Office is based at the National Library of the Netherlands.

The European Library of today has evolved from a number of earlier projects. Its starting point was ten years ago, when the 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 (OPACs), 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 was part-funded by the European Commission's IST research programme and the partners included the Conference of European National Librarians (CENL), the national libraries of Finland, Germany, Italy (Florence), the Netherlands, Portugal, Slovenia, Switzerland and the United Kingdom, together with ICCU, the Central Institute for the Union Catalogue of Italian Libraries. Its achievement was to set down policies and to undertake the technical groundwork 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 both aggregates the digital collections of Europe's national libraries and brings together the catalogues of its partner libraries to create a union catalogue for Europe. Since Spring 2010, The European Library has taken on a new role as the aggregator of digital content from national libraries for Europeana, providing monthly deliveries of digital content from national libraries to Europeana.
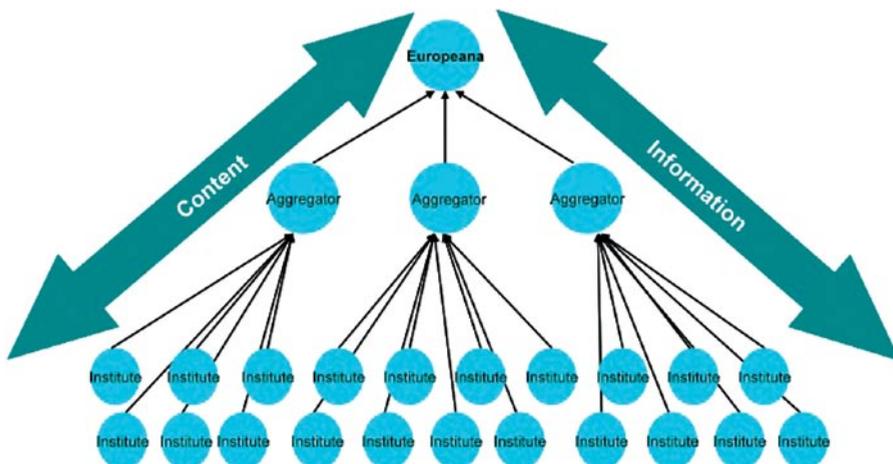
Fig. 2: The European Library.

Aggregation and Europeana

Europeana aims to give access to the digital content of many thousands of cultural institutions throughout Europe. It is therefore practically impossible for Europeana to work individually with each of these content providers. As a result, Europeana has developed an intermediate layer of content providers who aggregate content on behalf of a group of organisations. These aggregators could be, for example, national aggregators, collecting content on behalf of a number of organisations in one country; aggregators collecting content on a particular theme on behalf of Europeana, such as in the Europeana Travel project. Alternatively, aggregation could happen at the level of the type of organisation aggregating content, for example, an aggregator for libraries in Europe.

As outlined in the Europeana Aggregator's handbook, an aggregator, in the context of Europeana, is an organisation that collects metadata from a group of content providers and transmits them to Europeana. Aggregators gather material from individual organisations, standardise the file formats and metadata, and channel the latter into Europeana according to the Europeana guidelines and procedures. Aggregators also support the content providers with administration, operations and training. In comparison, a content provider is any organisation that provides digital content that is accessible via Europeana.

Fig. 3: Aggregators in the Europeana organisational model.

Aggregating Research Library Content for Europeana

Europeana's strategy requires a domain-level aggregator of library content, positioned alongside other domain aggregators for museums, archives and audio-visual collections. To enable this, LIBER (Association of European Research Libraries), CERL (Consortium of European Research Libraries) and CENL (Conference of European National Librarians) decided to join forces to develop a library-domain aggregator for Europe. As The European Library is already aggregating from national libraries for Europeana, it was agreed to extend The European Library's existing aggregation infrastructure to enable this.

Competitiveness and Innovation Framework Programme (CIP)

To realise this vision, LIBER, CERL and CENL agreed to submit a project proposal entitled 'Europeana Libraries', under the 'Digital Libraries theme' to the European Commission's Competitiveness and Innovation Framework Programme (CIP) under the specific Information and Communication Technologies Policy Support Programme (ICT PSP). This 'Digital Libraries' themeaims at improving the accessibility, use and preservation of Europe's rich and diverse cultural and scientific heritage in the online environment. The 2010 work programme intends to support activities that will continue to make more cultural and scientific content available online, whilst at the same time supporting the next development phase of the central Europeana service as an access portal to this content.

Within the 'Digital Libraries' theme, there are several objectives, the most relevant for the LIBER, CERL and CENL vision being *Objective 2.2 — Enhancing/ Aggregating content in Europeana.* As outlined in the work programme, this objective covers two main aims, the first being to enhance the Europeana content base by aggregating cultural heritage content and making it available through Europeana, and the second to enhance the searchability of cultural content already available through Europeana.

Europeana Libraries: Aggregating Digital Content from Europe's Libraries

The aim of the 'Europeana Libraries' project is to both significantly increase the amount of high-quality digital content from the libraries of Europe

in Europeana and to produce a sustainable model for it to continue as a major aggregator for Europeana into the future. This will not only enhance Europeana's current content base, but it will also improve The European Library's ability to incorporate such content cost-effectively by developing a domain-level aggregator for libraries, positioned alongside existing domain aggregators for museums, archives and audio-visual collections.

The key objectives of Europeana Libraries are to:

- increase the extent and range of quality content in Europeana by collaborating to establish a library aggregator for important complementary digital collections held by libraries across Europe, principally those of research libraries, thereby establishing channels for libraries to contribute their digitised content to Europeana;
- scale up existing infrastructure to manage the large-scale aggregation and ingestion of content into the libraries domain aggregator;
- improve interoperability by creating a single workflow for the ingestion of content from libraries into The European Library and Europeana;
- improve the quality of metadata to address the need for better search results in both The European Library portal and the Europeana service;
- offer the users of both The European Library and Europeana effective full-text search, making it possible not only to discover digital texts at the item level but also to search inside pages of books and articles;
- develop an exploitation plan to prepare the service for its extension, initially to other libraries within the membership of LIBER, CERL and CENL as well as beyond the consortium partners.

Aggregating High-Quality Content from Europe's Leading Research Libraries for Europeana

'Europeana Libraries' will aggregate high-quality content from nineteen leading research libraries from thirteen countries: Austria, Belgium, Estonia, Germany, Hungary, Ireland, Romania, Serbia, Spain, Sweden, Switzerland, UK, and Wales. The partners providing digital content to Europeana are all members of either LIBER or CERL. In total 5,168,453 pages, images, books and theses, AV clips and articles will be made available to Europeana as an outcome of the 'Europeana Libraries' project. Here is an analysis of the type and quantity of content to be provided:

Fig. 4: Analysis of digital content to be provided by Europeana Libraries.

Pages	Images	Books/Theses
3,319,045	848,078	598,130

Film/ Video Clips	Mixed content	Articles
1,200	34,000	368,000

Europeana Libraries will add content in a number of thematic areas which have been identified by analysing the content strengths of the Europeana Libraries partners:

- national history, culture and area studies
- travel and exploration
- European treasures
- European history, culture and area studies
- history of science and medicine
- creating a modern digital Europe
- history of art
- history of science and medicine
- literature and letters
- European cultural capitals

From within these themes, here are some highlights of the content that will be made available to Europeana via the project:

- Metadata for the monographs digitised by Google in the Bodleian Libraries, University of Oxford, the University of Ghent and the Bavarian State Library.

- Digital content from Europe's Special Collections including:
  - 80,000 images of unique material important for the history of science and medicine from Uppsala University Library.
  - 10,000 historically important portraits including those of important artists such as Van Dyck, Delacroix, the Galle dynasty, Wierikcx from Leuven University Library.
  - A special collection on the topic of Alexander the Great from the University Library in Belgrade.
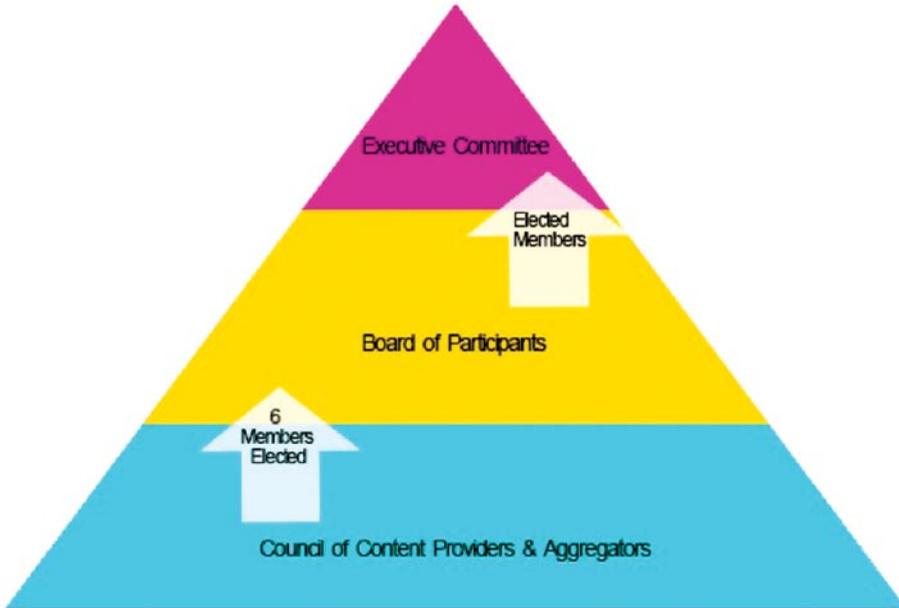
- Scholarly content including:
  - 50,000 records from universities across Europe by harvesting metadata for this material from LIBER's DART-Europe portal run by UCL (University College London). This will bring in university theses from countries such as the UK, the Nordic Countries, Spain (including the Catalan region), Belgium, Portugal, Ireland, Estonia, Switzerland and Hungary.

- Periodical literature is represented by a number of important contributions. Highlights include:
  - Lund University is offering 270,000 scholarly articles published in open access (from the Directory of Open Access Journals).

- Images are being offered by a number of partner libraries. Highlights include:
  - Tartu University is offering a collection which comprises 1,504 old autographs and paintings and graphics, 2,422 old photos, 118 old maps (starting from 1497), c. 5,044 digital letters of important persons all over Europe (Immanuel Kant's, I.J. von Essen's, Fr. L. Schardius's etc. correspondence);
  - The Wellcome Library is offering 108,000 images from its incomparable collections on medical history;
  - The Bodleian Libraries, University of Oxford, are offering 260,000 images. The outputs of more than a dozen collections of selected Special Collections materials drawn from libraries across Oxford, these items represent some of the hidden treasures of the University;
  - Ghent University, for example, is contributing practically the whole oeuvre of Piranesi, the eighteenth-century architect, in terms of drawings and etchings;
  - 287 drawings (1752–1914) from the Complutensian University in Madrid;
  - 50,000 engravings from the Complutensian University in Madrid.

- Moving film is also being offered by the partners. Highlights include:
  - 200 videos (1 hr each) from the University of Vienna, which is a record of current research activities, e.g. talks by ten Nobel Prize winners;

- The Wellcome Library is offering 900 clips from around 450 films. This archive of moving images has been selected from the historical medical film collections of the Wellcome Library. Topics cover all aspects of medicine, health and social care provision and practice in the UK;
- Legislative material: the Library of the Hungarian Parliament is offering approximately two million pages of historical parliamentary records, official gazettes, legal, historical and political literature (books and periodical);
- 1,500 items on the theme of European cultural capitals are also represented in the ephemera from Ghent University Library and by the Library of Lucian Blaga, with 1,000 images from the European event *Sibiu-European Capital of Culture 2007*.

Facilitating Research Library Content Aggregation

Within the Europeana Libraries project there are three key work packages that will work together to ensure the efficient aggregation and ingestion of the digital content from the research libraries into Europeana. *Work Package 3 — Aggregating research library content for Europeana*, will both project manage and undertake the actual aggregation of the digital content from all content partners and facilitate its ingestion into Europeana. Within this work package, an Aggregation Steering Group will be established, to ensure that all the aggregation-related work of the project is carried out cohesively and efficiently. This Aggregation Steering Group will also develop formal links with the Europeana Foundation Council of Content Providers and Aggregators, attending relevant meetings as necessary. The Europeana Foundation, Europeana's governing body, encourages content providers and aggregators to actively participate in the Council of Content Providers and Aggregators. As described within Europeana's Aggregator's Handbook, every type of organisation or project that provides content to Europeana can be a member of this council. The Council of Content Providers and Aggregators is intended to be a pan-European, cross-domain forum where the views of content providers can be reflected. It represents museums, libraries, archives and audio-visual content providers from the public or private sectors across a wide range of Member States. It is therefore the task of the Aggregation Steering Group to ensure that the European library-domain is represented at the Council of Content Providers and Aggregators and beyond.

Fig. 5: Council of Content Providers and Aggregators within the Europeana Foundation structure.

Extending The European Library Aggregation Infrastructure

A second key task as part of the aggregation activities for Europeana Libraries is to extend The European Library's existing aggregation infrastructure to enable the aggregation of digital content from libraries in Europe for Europeana. This will build on the tools that are currently used within The European Library technical infrastructure and also on those that have been made available to the Europeana network via the Europeana Labs software repository. A first task will be to create the functional requirements for the library-domain aggregation infrastructure, especially taking into account the requirements from the research library community. The software development process will also be undertaken using iterative software development practice. The usability and functionality of the tools will be validated across the library-domain community.

Enhancing the Searchability of Existing Library-Domain Content in Europeana

Enhancing the searchability of cultural content already available through Europeana is key to *Objective 2.2 — Enhancing/Aggregating content in Europeana* of the CIP IST PSP call. The Europeana Libraries project proposal has a particular work package focusing on this. The objective of this work package will be to undertake preparatory alignment of library-domain metadata that has already been provided to Europeana in the current Dublin Core-based Europeana metadata format Europeana Semantic Elements (ESE), to the Europeana Data Model (EDM).

The Europeana Data Model

The Europeana Data Model is the new proposal for structuring the data that Europeana will be ingesting, managing and publishing. It will help enable users to browse the content in Europeana in new ways and will facilitate Europeana's participation in the semantic web. EDM is intended to build on, and be backwardly compatible with, ESE. The philosophy behind EDM is that is should not build on any particular community standard but rather adopt an open, cross-domain semantic web-based framework that can accommodate the range and richness of particular community standards. In summer 2010, the validation of the proposed EDM by the libraries, archives, museums and audio-visual domain is reaching completion. Over summer 2010, the first phase of prototyping of EDM will take place. The first alpha version of EDM is expected to be released in autumn 2010. From January 2011, it is expected that EDM-based content will be ingested into Europeana. If successful, the research library content which will be provided to Europeana as part of the Europeana Libraries project will be provided in EDM format. The European Library has been both actively involved in the validation of EDM and will continue to follow progress during the prototyping stage. This involvement will form the basis of the continued library-domain involvement in the development and implementation of EDM.

Modelling the Library Landscape to Create a Sustainable Business and Organisational Model for the Aggregator

Developing an efficient, sustainable and cost-effective library-domain aggregator is essential to the success of the Europeana Libraries project. It is

therefore important that an appropriate aggregation infrastructure is developed that will be compatible with the library-domain landscape in Europe. A specific work package has been developed to model the library landscape from an aggregation viewpoint in order to recommend a sustainable, domain-wide framework for aggregating library content into Europeana. Building on previous aggregator surveys undertaken within the Europeana network, Europeana Libraries will undertake desk research to model the library-domain landscape in Europe. To support this research, interviews with key stakeholders from the national, research, university and public libraries in Europe will be conducted. This preparatory work will enable a range of scenarios for the aggregation of digital content from libraries in Europe into Europeana to be developed. The library community will then validate these scenarios. Once validated, the costs of implementing the scenarios will be identified. By the end of the project, a report recommending a sustainable aggregation infrastructure for Europe beyond the life time of the project will be produced for implementation.

Next Steps

The 'Europeana Libraries' consortium successfully submitted the project proposal to the European Commission by the 1st June 2010 deadline. The project proposals will be evaluated by external experts in June and July. A number of proposals will then be invited for negotiation during September and October. The successful projects will start towards the end of 2010/ early 2011. We hope that 'Europeana Libraries' will be one of the proposals selected.

Websites Referred to in the Text

CENL (Conference of European National Librarians), http://www.cenl.org

CERL (Consortium of European Research Libraries), http://www.cerl.org/web/

Competitiveness and Innovation Framework Programme (CIP) Information and Communications Technologies Policy Support Programme, http://ec.europa. eu/information_society/activities/econtentplus/index_en.htm

Europeana, http://www.europeana.eu; the 'provide content' section of the Europeana Version 1.0 website provides information about the Europeana

Content Strategy, the Europeana Aggregator's Handbook as well as the technical requirements for providing content to Europeana: http://version1.europeana.eu/web/europeana-project/provide-content

Europeana Council of Aggregators and Content Providers, http://version1.europeana.eu/web/europeana-foundation/content-council

Europeana Data Model prototyping pages, http://europeanalabs.eu/wiki/EDMPrototyping

Europeana Labs, the central resource for all developers working on the core Europeana system as well as for contributing developers, http://europeanalabs.eu/

i2010 Digital Library Strategy, http://europa.eu/i2010

LIBER (Association of European Research Libraries), http://www.libereurope.eu/

The European Library, http://www.theeuropeanlibrary.org