Positioning Libraries in the Digital Preservation Landscape

S. K. Reilly
LIBER- the European Association of Research Libraries

Abstract
This paper draws on LIBER’s experience in several European best practice network projects related to the digitisation of cultural heritage (Europeana Libraries, Europeana Newspapers) and to digital preservation (APARSEN). Collaboration with a diverse set of practitioners, including both public and private stakeholders, means that libraries can stake their place in the common vision for digital preservation and ensure that the issues surrounding the preservation of digital cultural heritage are represented in this vision. In the long-term, this collaboration should ensure that digitized cultural content from libraries is made available in the most cost effective and sustainable way to ensure continued access into the future.

Abstract
Susan is the LIBER Project Officer and manages LIBER participation in several EU projects in the areas of Open Access, data exchange, digital preservation and digitisation. She holds an MSc in Information Management from the University of Sheffield, and has several years experience working across a range of libraries, including management of the Library service at the Irish Management Institute in Dublin. Her interests range from open data, digital heritage, collaboration and innovation, and research infrastructures.

1. Introduction
The Stichting LIBER Foundation (Stichting LIBER)\(^1\) is the principal association of the major research libraries of Europe. Its current membership includes 425 national and research libraries from more than forty countries, mainly but not only, in Europe. Its overall aim is to support a functional network across national boundaries in order to ensure the preservation of European cultural heritage, to improve access to collections in European research libraries, and to provide more efficient information services in Europe. In the 2009-2012 LIBER strategy\(^2\) (1) Digitisation and Resource Discovery, and (2) Heritage Collections and Preservation were identified as 2 of the 5 key performance areas for the Association.

Under digitisation and resource discovery LIBER set out to actively contribute to the digitisation of research library collections through the sharing of knowledge and best practice. In the area of Heritage

\(^1\) See [http://www.libereurope.eu](http://www.libereurope.eu)
Collections and Preservation the focus was on the development of standards, addressing access and preservation issues through and in digitisation, and the curation and preservation of born digital and digitised material.

For both key performance areas collaboration and participation in European projects were identified as vehicles for achieving their goals.

Over the past 3 years LIBER has worked towards developing its network and actively seeking out and building on opportunities to engage in collaborative projects to help develop best practice in digitisation and increase the visibility of library collections from across Europe.

2. Libraries & the Digital Agenda in Europe

As increased access to knowledge in the Digital Economy is now being viewed as key to economic development, libraries are under pressure to increase access to the knowledge within their collections. Some national libraries\(^3\) are even redefining their functions to more explicitly state their role in supporting education and research.

Institutions such as the European Commission have great influence over the development of the Information Landscape. Initiatives such as the Digital Agenda\(^4\) have brought European research libraries into the centre of political discussions on the empowerment of the European citizen. This has presented LIBER with the opportunity to position research libraries at the centre of the Digital Agenda, in particular via their engagement in the dialogue surrounding development of European research infrastructures.

The Commission has also set some clear targets for the digitisation of Europe’s cultural heritage.\(^5\) By 2015 all public domain masterpieces must be made available via Europeana and, by 2025, all of Europe’s cultural heritage. As the requirement for the legal deposit of born digital becomes a reality, the size of such national library holdings will increase dramatically. Although progress has been unsatisfactory thus far,\(^6\) the Orphan Works Directive, in conjunction with Arrow,\(^7\) should open up space for the mass digitisation of and increased accessibility of cultural heritage material.

The Digital Libraries Initiative\(^8\) sets out to make all of Europe’s cultural resources and scientific records – books, journals, films, maps, photographs, music, etc. – accessible to all, as well as preserve them for future generations. In its Recommendation on the Digitisation and Online Accessibility and Preservation of Cultural Heritage Material\(^9\) acknowledges that:

---


\(^7\) See [http://www.arrow-net.eu/](http://www.arrow-net.eu/)


1. There is a need to increase the volume and types (e.g. audiovisual) content available in the Europeana\(^{10}\) Portal;
2. That no clear and comprehensive policies are in place on the preservation of digital content in several EU Member States and the absence of which poses a threat to the survival of digitised material and may also result in the loss of material produced in digital format;
3. That digitisation is costly and steps should be taken to reduce cost through collaboration for economies of scale and public-private partnership.

The recommendations to address these issues include:

1. The use of domain specific cross border aggregators to bring about economies of scale;
2. Ensure the use of common digitisation standards defined by Europeana in collaboration with the cultural institutions in order to achieve interoperability of the digitised material at European level, as well as the systematic use of permanent identifiers;
3. Increased targets for digitisation activity;
4. Make the necessary arrangements for the deposit of material created in digital format;
5. In order to guarantee its long-term preservation.

The implication of these recommendations for Europe’s libraries are that (a) there is a need for a library domain aggregator, (b) that they must engage in the development of digitisation and metadata standards, and (c) find new ways of reducing the cost of and funding digitisation, and, if they are mandated to preserve born digital material, (d) ensure that they have the correct skills and infrastructure in place to do so.

A survey carried out by the NUMERIC project\(^{11}\) found that only 1% of library collections have been digitized but that a further 30% of the collection was going to be digitized. This represents a huge investment in digitisation, but to what end?

The New Renaissance Report,\(^{12}\) which examines the digitisation of European cultural heritage, sees this investment as beneficial in terms of:

- Wider access to and democratisation of culture and knowledge;
- The educational system - both schools and universities;
- The development of new technologies and services for digitization;
- For digital preservation;
- Interacting in innovative ways with the cultural material.

Applying this logic to the library domain indicates that for libraries to ensure that full benefits of their digitization activities are realized they must ensure that they maximize the visibility of their collections,

\(^{10}\) See [www.europeana.eu](http://www.europeana.eu)


not just to the general public but to those in the education system and must make it possible for individuals to interact with the content in new and innovative ways.

3. Collaboration

The first milestone in LIBER’s work towards seeking opportunities to engage in relevant collaborative projects was its involvement in Europeana Travel. Europeana Travel13 was a 2-year project, which started in May 2009, funded by the European Commission. The aim of the project was to digitise content based around the theme of travel and make it available in Europe’s portal for digitised cultural heritage, Europeana. Although LIBER was not an official partner in the project, Europeana Travel was a significant project for the Association because it brought together LIBER members from both national and university libraries as partners to develop best practice in digitisation. It also raised issues surrounding sustainability as the national and the other research libraries within the project used different aggregators to deliver their content to Europeana. The Consortium of European National Libraries (CENL) had already established aggregation service, The European Library,14 but this service was only open to members of CENL. Europeana Travel set up a LIBER aggregation service for research libraries, which was to be extended to other libraries after the life of the project. As the project began to come to a close it became apparent that the best solution for sustaining an aggregation service for both groups of libraries would be to open up The European Library service to research libraries.

The Europeana Travel project also conducted a survey of the state of digital preservation practice in LIBER libraries that were digitising their collections. The survey15 found that there was an under investment in skills for digital preservation and often a lack of written policies on the preservation of digitised material. It recommended (1) the sharing of best practice and (2) investment in digital preservation infrastructures.

4. Sustainability

Europeana Travel established the basis for collaboration between library networks. The Europeana Libraries16 project brings together several European library networks, incorporating both national and research libraries, in order to create a partnership which will provide value to both researchers and research libraries alike. LIBER, CENL, the Consortium of European Research Libraries, and the Europeana Foundation are all partners in the project.

It addresses the issue of sustainability by opening up the national library aggregation service, The European Library, to research libraries. It uses this service to aggregate a critical mass of valuable content from European research libraries. By the end of the project in December 2012 over 5.1 million objects, including 1,200 film and video clips, 850,000 images and 4.3 million texts (books, journal articles, theses, letters) will have been ingested from 19 research libraries. Much of this content is full text and of particular value to researchers. To maximise on the potential of this content, the project also set out to develop full text search capabilities and a search portal that provides tools specific to research.

---

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

Once the project ceases at the end of 2012, the aggregation service will be opened up to all research libraries in Europe. Much work has been done to engage the libraries within the CERL, CENL and LIBER network to define the value propositions of the aggregation service.

The value propositions for libraries include:

1. Widened access to library resources via a dedicated portal (run by libraries) for researchers of library content and catalogues.
2. A one-stop-shop for data processing, comprising a content ingestion workflow that is seamless, automated and extremely fast, together with data enrichment services.
3. Networking and knowledge sharing.

As the service is opened up to more libraries, the sharing of best practice amongst the network and continued development of standards will ensure that content and metadata remains at the highest standard and will support libraries in working more efficiently in the areas of digitisation and metadata.

5. Best Practice & Innovation

Sustainability is not ensured by merely opening up the aggregation service to more libraries. For The European Library there is a pressure to be increasingly innovation and progressive in developing the service in line with technological developments and changes in the behaviour of researchers. For libraries there is pressure to constantly increase the visibility, impact and accessibility of their collections. Collaboration and exploitation of the library networks enables libraries to contribute to the development of, and have access to, new digitisation technologies and best practice. This is exactly what Europeana Newspapers facilitates. The Europeana Newspapers aims at the aggregation and refinement of newspapers content, which will be made available through The European Library and Europeana. In addition it addresses challenges particularly associated with digitized newspapers. Through the project, 13

---

research and national libraries will make digitised newspaper content (from the early 19th century) more available and accessible that it has even been before.

By bringing a mass of newspapers content from all over Europe together in this way Europeana Newspapers achieves the following:

- Contributes to creating a critical mass of content on the First World War and providing an invaluable resource to researches and citizens alike by making newspapers published during this period available via Europeana;
- Adds new value and meaning to newspapers holdings by showing individual newspaper collections in a broader context;
- Creates new search and display capabilities though the development and application of new technologies.

There are other benefits to the work of this project as the procedures with which newspaper content will be upgraded include Optical Character Recognition (OCR), Optical Layout Recognition (OLR)/article tracking, Named Entity Recognition (NER), and page class recognition. OCR will allow users to search the full text of the newspapers. OLR will facilitate searching by article title or section and, the application of named entity recognition means that users can find content related to places, people or things. In other words, the digitised content can be searched more easily, deeply, efficiently, and is such a way as to unveil new relationships and contexts. For each of these technical tasks best practice recommendations will be identified and published. This will be of huge benefit to the broader network of libraries with the CERL, CENL and LIBER networks. It will help reduce the cost of newspaper digitisation projects and increase the accessibility of digital newspaper collections now and into the future.

6. Challenges

As was first recognised in Europeana Travel, digitisation necessitates investment in digital preservation. As libraries step up their digitisation activity the challenge of digital preservation increases. A common thread between all of LIBERs digitisation projects is that they highlight the need for libraries to engage with, and work in a collaborative way to address the challenges related to the preservation of digital cultural heritage.

So how can libraries collaborate to address some of the challenges related to the preservation of digital cultural heritage?

One way is to engage with the digital preservation community as a whole. Through the Alliance for Permanent Access to the Record of Science in Europe Network (APARSEN) project LIBER libraries have the opportunity to engage in a network of excellence for digital preservation and to contribute to a common vision. Collaboration with a diverse set of practitioners, including both public and private stakeholders, means that libraries can stake their place in the common vision for digital preservation and ensure that the issues surrounding the preservation of digital cultural heritage are represented in this vision.

A more unified vision also means a stronger business case for investment in digital preservation. Taking the findings from the Blue Ribbon Task Force, which recommended the definition of roles and responsibilities among stakeholders to ensure an ongoing and efficient flow of resources to preservation throughout the digital lifecycle, APARSEN partners are working together to assess how prepared we are in terms of economically sustainable preservation and how the following conditions for economically sustainable digital preservation can be created:

- Recognition of the benefits of digital preservation on the part of key decision-makers;
- Incentives for the decision-makers to act in the public interest;
- A process for selecting digital materials for long-term preservation;
- Mechanisms to secure an ongoing, efficient allocation of resources to digital preservation activities;
- Appropriate governance of digital preservation activities.

Participation in a network of excellence also facilitates the exchange of experience and expertise. In the case of APARSEN, this means that gaps in training are identified and filled and expertise is exchanged via various mechanisms such as multi-stakeholder workshops, Webinars and roadmaps. One of the roadmaps to be produced will be aimed at national and research libraries. Based on a survey of LIBER members, it will document the current situation, identify best practices and reveal the common digital preservation challenges that need to be addressed by libraries going forward.

The intended result of the APARSEN project is a stronger and wider body of expertise, but also a strengthening of, and a deeper understanding across, the network of stakeholders.

7. Conclusion

The push for the digitisation of cultural heritage is both a challenge and an opportunity for libraries in Europe. Digitisation is costly and libraries must find ways of ensuring that it is carried out in the most cost effective and efficient manner. One of the best ways to justify the expenditure of public money on such activity is to make the digitised content as accessible and visible as possible by applying best practice in metadata and pushing content out through an aggregator. The application of new techniques and technologies can add value to the content by making it more searchable, providing context and ultimately making it a useful resource for research and innovation.

Digital preservation is also a costly activity and it is up to libraries as custodians of digital cultural heritage to make the case for it and to engage in best practice.

Collaboration and the use of networks present the opportunity to for libraries to work more effectively and efficiently. For the LIBER libraries, collaboration has facilitated the development of best practice, enabled European libraries to be leaders in the use of digitisation technologies such as OCR, and given libraries the collective voice to engage in the wider dialogue surrounding digital preservation.

Sustainability is at the heart of digitisation, both in terms of accessibility and preservation. The next step for LIBER is to engage in the development of a sustainable European infrastructure for digitised cultural heritage (Europeana Cloud). The creation of a critical mass of high quality digitised content means that content providers and aggregators, across the European information landscape, urgently need a

---

cheaper, more sustainable technical infrastructure that is capable of storing both metadata and content. Researchers require a digital space where they can undertake innovative exploration and analysis of Europe’s digitised content. Europeana Cloud will create a cloud-based infrastructure capable of delivering cost-efficient content and metadata storage for stakeholders across Europe. It brings libraries and other content providers together with other stakeholders, such as technical solutions provider and researchers to provide a trusted, efficient, cheaper and more effective way of making Europe’s cultural heritage accessible.

Collaboration and engagement in the development of best practice is the key to discovering sustainable solutions to the challenges of digitisation and digital preservation. It encourages innovation and increases competitiveness. It can also lead to further collaboration and help to forge connections with new communities.