Digitisation of Library Material in Europe: Problems, Obstacles and Perspectives anno 2007

Erland Kolding Nielsen

Director General of the Royal Library, Copenhagen, Postboks 2149,
1016 København K, Denmark, ekn@kb.dk.

Abstract

This paper is a short introductory policy paper about the state-of-the-art of digitisation of library material in Europe, seen from the chief executive point of view of a big national and university library in the autumn of 2007. It focuses on current problems, obstacles, and some perspectives. What has been achieved, what are the problems and obstacles in terms of especially mass digitisation in the light of the so-called Google challenge and the response by the Commission of the European Union, and what are the consequences likely to be?

Introduction

Digitisation of library and archive material[1] has been part of library activities for about 15 years. Many national libraries and big university libraries, not to mention archives and other cultural institutions, have digitised – normally smaller – parts of their collections often without an overall plan or at any rate without coordination at the national level. National schemes for systematic digitisation are very rare, if existing anywhere at all. The efforts and results are substantial and manifold, but it is very difficult to gain an overview. The first documented overview of digitisation within national libraries was carried out by the National Library of Austria on behalf of the Conference of European National Librarians (CENL) in 2007, and will be presented at this conference. Some of the conclusions are astonishing and rather disturbing.

What then are the characteristics of the achievements of libraries during these first 15 years? The CENL investigation reveals that only 1% of the

holdings has been digitised so far, i.e., approximately 4.7 millon items, representing 17 million pages:

- The main emphasis of digitisation has been on newspapers, special collections and rare, fragile or heavily used material within this category, i.e., manuscripts, rare books, photographs, maps etc.
- The priority and reason for digitisation has been *access,* not preservation.
- Standards for digitisation formats have varied, and are only now in the process of being settled and agreed upon in terms of permanence and preservation.
- Mass digitisation has not been planned or carried out, with two exceptions.
- Books (and journals), comprising only 12%, i.e. 619,000 (of which 80% again are Russian dissertations) have – again with two notable exceptions – not yet been systematically digitised at any national level.
- Hardly any books from the 20th century have been digitised.
- European scholarly journals have not yet been retrodigitised in most European countries.

The investigation also shows what the situation will be like in 2012, if the present policies, priorities and financial conditions continue. The coming European Digital Library (EDL) will be a library without books!

The Case of the Danish Royal Library

As a typical example of what a large – if not one of the largest in Europe – national and university library like the Royal Library has done since it published the first digital texts on CD-ROM in 1989 until the present time, when everything can be accessed over the Internet on the website, is managed by a CMS, a content management system, and is stored in a so-called DOMS, a digital object management system, the following short overview can be considered representative for the present situation.

We have now digitised c. 175,000 digital objects, varying from fiction books to manuscripts to photographs, often as collections with new names for marketing purposes.[2]

- Books: Access to a selection of important Danish literary classics until 1937. Special fulltext database *The Digital Archive of Danish Literature* (fiction). 2,300 books, 310,000 pages since 1996.
- Government reports since 1848: c. 1,200, c. 200,000 pages, 2005–06.
- Manuscripts, archives, and rare books: Access to a small selection of important manuscripts and early prints (mostly Danish and European). 350 mss. and rare books, 71,000 pages since 1996.
- Music: Access to a selection of musical scores, incl. manuscripts, mostly Danish composers. 3,600 prints and mss., 106,000 pages since 1996.
- Photographs and maps: Access to a selection from the Danish National Photo Archive and the map collection. Subjects: Danish topography and portraits. 148,500 items digitised since 1996.
- Serials: Access to a small selection of Danish journals. A Danish counterpart to *JStore* is being published this autumn with full retro-digitisation of the first 10 main journals (ca. 35,000 articles) from the 19th century until today, called 'tidsskrift.dk' (journal.dk). 14 serials digitised, 316,000 pages, since 2003.

This overview is probably more or less representative of the current situation of many European national libraries anno 2007. It is apparently not enough, and it certainly does not address the problems of mass digitisation of books and journals, the core collections of national and university libraries.

What are the Present Obstacles?

What are the obstacles for speeding up this situation and providing more digital content in the years to come?

Technology

The technology of digitisation has developed rapidly over the last 5–8 years, and today I do not think that technology poses a problem, except perhaps

a financial one. Scanners for different types of material, different sizes and conditions have been developed, including those that are necessary for valuable and/or fragile material. In many institutions and countries a bigger problem seems to be an organizational one: how to organize the digitisation business – the production flow – most effectively and efficiently, especially on a broader scale, regionally or nationally.

Digitisation Formats and Preservation

The first 15 years showed a range of different formats, some of which were not suitable or sufficient in terms of preservation, and that means that the digital material cannot always survive in new digital surroundings without enormous cost of preservation. It is now clear that at least parts of what has already been digitised has to be digitised yet again in order to secure that output meets the requirements of current e-publishing and preservation standards.

At present the cost of the full digitising process is still very high, and we can hardly imagine the cost of repeating the process even though technological advances make this desirable. This indicates that applicable standards for digitising must support a compromise between the two extremes in financial terms: 1) digitising for access on *de facto* browsers/players and 2) digitising for substitution. It is important from the perspective of a European Digital Library that the libraries can agree on formats suitable for access as well as long-term preservation. Focusing on a rather limited number of open formats combined with strong collaboration within the library world should make it possible to define a dynamic set of best practices safeguarding the investment for as long as possible.

Finance

Most national and some university libraries have already redirected quite large financial resources to digitisation purposes, but as they have rarely got sufficient money for their overall activities it is impossible to finance really big programs, e.g., mass digitisation of books and journals. There is only one exception (France), and perhaps one or two under way, but not altogether clear yet.

Organization

In Denmark we have an ongoing debate on who should digitise how much for how many? This year, we developed a business case that shows that the cheapest way of organizing digitisation on a large scale is to concentrate the process and build up advanced digitisation competence within a few large institutions.

The situation seems to be similar in other countries. Too many – often small – institutions or institutions with relevant collections of too modest a volume want to digitise too little at too high a price without being able to justify distributed costs of investment and management.

Copyright

By far the biggest obstacle today to digitisation of material even after 1880 – apart from the financing – is the present legal situation of European copyright and the conditions and possibilities of negotiating and acquiring the right to digitise objects within the 70 years' limit of the death of the copyright holder.

The extension of the copyright limit from 50 to 70 years after the death of the copyright holder was simply a catastrophe and an enormous obstacle to developing a relevant, adequate and comprehensive EDL with 20th-century material of sufficient importance. The frequently emphasized balance between the interests of the copyright holders and the users, *in casu* the institutions trying to convert the physical material into the digital, has completely tilted to the advantage of the copyright holders. The legal demands of investigating and finding the heirs etc. are simply prohibitive for mass digitisation projects with contents from the 20th century.

The sooner the European Commission understands this and acts accordingly, the better the chances of developing a comprehensive and relevant EDL at a level of cost (both in terms of production and administration) within the range and possibilities of the institutions in question.

Erland Kolding Nielsen

The Google Challenge

The announcement in December 2004 by Google that they would start a massive digitisation programme of books – digitizing 'the world's knowledge' (15 million books from originally six major research libraries) based on entire university library holdings from the 19th and 20th centuries especially from the USA and UK[3] was considered an enormous challenge to almost all European countries, as it could be foreseen that only fragments – and even arbitrary parts – of the national imprint would be incorporated, and the consequence of that could be that Anglo-American books would in the future dominate at all levels of education, research, scholarship and public use. I did agree then – and still do - with most of the main points of criticism voiced by my former French colleague, Jean-Noel Jeanneney in his book *Google and the Myth of Universal Knowledge: A View from Europe*.[4]

The Response of the European Union

The response came quickly, but – in my opinion – inadequately, from the European Union on September 30, 2005, with the communication called *i2010: Digital Libraries*,[5] followed by an extensive hearing process within the library, archive, museum and cultural sectors of Europe,[6] and finally the *Recommendation on the digitisation and online accessibility of cultural material and digital preservation* by the Commission of August 24, 2006, to which all ministers of culture agreed in November 2006. This is the framework for digitisation policy actions of the European Commission in the years to come, including the EDL project, based on the already established TEL service, The European Library, a portal introduced by the Conference of European National Librarians some years ago.

I assume that you are all aware of the vision and content of the communication. There is, of course, in my opinion, nothing wrong with the vision (a European Digital Library with more than 12 million objects by 2012), but the way the Commission addresses the financial problem of mass digitisation and especially its expectations as to the possible results of public-private partnerships are unrealistic, as there is no viable market in most European countries for digital products of this kind, with the exception of the English

and Spanish speaking world – not even the French speaking world is large enough.

Conclusions

At the first presentation of the CENL-survey a month ago it was concluded: 'On institutional level systematic content digitisation is daily practice in many European National Libraries. On the national and EU level there is a need for co-ordinated funding of mass digitisation and building up a digital library infrastructure'.

Today we can foresee that the European Union will reach its goal in terms of digital content, defined as expected number of digital objects, in the EDL, even if the present situation in terms of priorities and level of activity should continue, *but* both the national libraries, the member states *and* the Union will have to address the problem that this constitutes a great risk. It can be predicted, too, that if priorities in financing and resource allocations are not changed, the EDL in 2010 or 2012 will still consist of mostly digital heritage objects (which is of course in itself not bad at all) and very few books and journals.

Why is that? Well, simply because if governments refuse to pay for mass digitisation of their national imprint of books and journals, this will either not be done or done only by private firms on terms that we normally are not willing to accept in Europe.

Accordingly, the political issue to be addressed at both European and national level in all 47 European countries is: who is going to pay for mass digitisation of books and journals, and what restrictions to public or general use shall, will or must we accept in the future, if digitisation is entirely carried out by the private sector, i.e. Google or Microsoft?

A European digital library without access to the most important contributions from a variety of scholarly angles – books and journals – is a chimaere even in the age of the Internet. But it is a prediction that may come true in due course, simply because the governments are too slow or do not see the threat, and because the Commission has really not understood the urgency of the problem, as it emerges today.

The problem of mass digitisation might be stated this way: it is either the state (the public sector) or Google! So what do we want? Free access or restricted access to what has been free so far in the physical world, but now on market terms in a marketplace without real competition?

I hope that this conference shall address this problem among many others.

Websites Referred To In The Text

CENL, Foundation Conference of European National Librarians, http://www.nlib.ee/cenl/index.php

EDL, European Digital Library, http://www.edlproject.eu/

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

Notes

---

[1] This paper deals only with *retrodigitisation* of library material in physical formats, *not* with digital born material whether this is acquired or harvested by libraries or archives.

[2] The digitised collections are described at the URL which gives access to the collections, central URL: http://www.kb.dk/da/nb/materialer/e-ressourcer/index.html

[3] Cf. Ronald Milne: 'The Google Mass Digitisation Project at Oxford', *LIBER Quarterly*, 16 (2006) 3–4.

[4] French ed. April 2006, Eng. s.y. Cf. also David Bearman**:** 'Jean-Noël Jeanneney's Critique of Google: Private Sector Book Digitisation and Digital Library Policy', *D-Lib Magazine*, 12 (December 2006) 12.

[5] http://ec.europa.eu/information_society/eeurope/i2010/index_en.htm. The website has a good overview of the policy actions and documents within the field.

[6] Responses from a.o. LIBER and CENL, cf. their websites.