The Bookshelf: digitisation and access to copyright items in Norway

Vigdis Moe Skarstein
National Library of Norway, Oslo, Norway

Abstract
Purpose – The purpose of this paper is to provide an overview of the challenges faced in digitising copyright-protected books and other materials within the National Library of Norway.

Design/methodology/approach – Following a description of Bokhylla, or The Bookshelf, the paper outlines the strategies adopted within the National Library of Norway for digitising a range of materials, including books.

Findings – The National Library of Norway has negotiated a contract with Kopinor, an organisation that represents rights-holders from about 30 different organisations, and this will enable users to access some 50,000 copyright-protected books within the next three years.

Originality/value – Copyright presents challenges in any digitisation project. Norway has adopted a unique solution for enabling its citizens to gain digital access to materials deposited within the National Library.

Keywords National libraries, Digital libraries, Digital storage, Multimedia, Copyright law, Norway

Paper type Case study

1. The Bookshelf
1.1 General description
On the 27 May 2009 the National Library of Norway launched its newest digital service: Bokhylla.no, or The Bookshelf. This service makes available published copyright-protected material in full text on the internet, and within three years about 50,000 copyright-protected books will be available through the National Library of Norway’s web site. At the launch date, 14,000 books from the 1990s were included as well as some 3,500 published much earlier. The project aims to showcase Norwegian literature and, with much of the information on the internet published in English, it aims to provide a web presence for the cultural and literary heritage of one of Europe’s less major languages.

The web site is in Norwegian, and the opening page is shown in Figure 1 (www.nb.no/bokhylla).

The collection can be searched by author, title, words, text or quotations. Figure 2 shows the result of searching for a quote from Dr Stockman in Ibsen’s play, The Enemy of the People “Flertallet har aldri retten på sin side” (which can be translated as “the majority never has right on its side”).

This paper is based on a presentation made by the author at the International Federation of Library Associations and Institutions Conference in Milan in August 2009. Proceedings are at: www.ifla.org/files/hq/papers/ifla75/190-skarstein-en.pdf
Figure 1. Opening page of Bokhylla

Figure 2. Results of a search on The Bookshelf
Clicking on a retrieved item provides the basic metadata and when the URL for the book, *En Folkefiende*, is clicked, then the user is shown the specific page with the quote being highlighted as seen in Figure 3.

The background for this project is a contract between the National Library of Norway and Kopinor – an organisation which represents rights-holders from about 30 different interest organisations, such as authors, publishers, graphic artists and so on.

The National Librarian of Norway was given the authority to negotiate such a contract with Kopinor from the Minister of Cultural Affairs, who officially launched the service and said it will be a pilot to see if one can find permanent principal criteria to pay rights-holders for giving access to copyright-protected material.

1.2 The scope of the contract

The BokHylla project relates to printed books published in Norway during the years 1790-1799, 1890-1899 and 1990-1999, including literature translated into Norwegian. The contract regulates copyright-protected material and will involve up to 50,000 books with an estimate of 185 pages per book.

The right given to the National Library of Norway that enables it to make available the copyright-protected material includes the necessary reproduction for viewing copies on the digital storage, which the National Library of Norway makes pursuant to the regulations issued in the Norwegian Copyright Act (www.kopinor.org/opphavsrett/norwegian_copyright_act). The contract applies to Norwegian and foreign publications that have rights-holders represented by Kopinor and
competition law rules. The provision also stated “pursuant to the extended collective licence in the Norwegian Copyright Law, that published materials created by rights-holders that Kopinor does not represent are also covered by the contract”.

The material is to be made freely available on the National Library of Norway’s website for users with Norwegian IP addresses. The material is to be made available as individual pages in a digital format to be viewed on a computer screen. The users will be able to search in the text and search for covers, title pages and tables of contents, as well as to browse sequentially through individual pages. No arrangements are to be made for downloading or printing the material until the copyright-protected period has expired. Publishers will receive €0.06 per page distributed via Kopinor. The material will be made searchable via search engines and solutions to provide information on where the material can be borrowed or bought will be implemented.

1.3 The High North project

BokHylla is the second project in the National Library of Norway related to making available copyright-protected printed material in digital format on the internet. The first project, called The High North, is an agreement with about 6,000 different rights-holders to give access to 1,400 books in full text on the internet. Figure 4 shows the homepage of this project (www.nb.no/highnorth/).

A report written in 2008 after two years of operation provided a statistical analysis of the behaviour of users of this online service, of library lending figures and of sales statistics (National Library of Norway, 2008a). The findings from an online questionnaire
showed interesting results on how, and how much, the users read on the net. In particular, it shows how older content scores highly – as Anderson (2006) predicted in his *The Long Tail*.

It is the knowledge obtained from the High North project that made the Minister of Culture establish a group to suggest criteria for negotiations with rights-holders on how to pay for access to digital full text on the net in the National Library. On this basis the National Librarian negotiated the agreement that enables the library to give access to about 50,000 books, from different periods of Norwegian literature, in full text.

2. The Digital National Library
2.1 General overview
The vision of the National Library of Norway is to be a living memory bank, by being a “Multimedia Centre of Knowledge” with a focus not only on preservation, but also on mediation.

To succeed with this ambition, one of our main goals is to be a Digital National Library. It is important to have as much digital material as possible, not only historical material but also the modern part of the cultural heritage, and to give access to as much content as possible, to as many as possible, whenever required.

To be able to do this, we have two main important strategies:

1. To get as much material as possible deposited in digitised format.
2. To scan our total collection as fast as possible.

In 2006 the Norwegian National Library therefore started a systematic digitisation of its whole collection. According to Norwegian Copyright Law it can do this for preservation purposes.

As far as possible, we are now streamlining the process from when an object is selected for digitisation until it is placed digitally in our digital long-term repository, simultaneously offering web access for authorised users within the National Library’s premises. With the multimedia collection of the National Library this poses special challenges, as we need to establish separate production lines for the digitisation of different types of material. In addition, there are usually several variants within each type of material, which again demand different processing.

Implementing this is a challenge, in terms of technology, logistics, organisation, manpower and financing. Not least it is a challenge when it comes to copyright issues. At the beginning of 2009 we invested €2.3 million on machines and software, and reorganised our budget to give priority to digitisation. We also have received an increased budget to pay for copyright.

2.2 Modern Legal Deposit Act
Based on a modern Legal Deposit Act the National Library of Norway receives a copy of everything that is produced in Norway and is of interest to the public, be it books, newspapers, periodicals, photos, film, music or broadcast programmes.

All broadcasters with a licence in Norway may be asked to deliver copies of their programmes for preservation and we have extensive collaboration with the national broadcaster (NRK) for preservation and dissemination. At present, the programmes of
five radio channels are transferred every night to the National Library as digital files via the internet. The NRK can find and use its own material from the Digital Radio Archive and part of the material is also available at certain libraries. In 2008 we also started the digital legal deposit of the television programmes from two broadcasters.

It is a part of our strategy to negotiate with the publishers to get as much material as possible deposited in digital formats. The National Library’s agreement with NRK, both to have their programmes transferred digitally and to function as its archive, has been a model for others.

We have also recently signed an agreement with the most important newspaper and publishers in Norway to acquire digital deposited material and to give access to the material on certain conditions (National Library of Norway, 2008b). We also preserve those digital signals that remain digital all the way from production until they reach the user. Since 2003 we have harvested large parts of the Norwegian web domain.no. Some material is thus already delivered to us in a digital format. We have a relatively large digital collection when it comes to the audio-visual material, although we have less when it comes to printed material, but this is increasing daily.

With as much raw data as possible, we will establish a “data bank” that will both be a repository for future use, but also a source both for the National Library and other institutions to build various digital services to give access to the material today.

2.3 Systematic digitisation

Much of 2006 was spent in carrying out the process of inviting tenders for the purchase of digitisation software, digital storage and other ICT solutions required. In addition, a project manager and some system development staff were hired. Since 2007 we have systematically digitised material. Table I shows the status of our digital collection at the end of 2008. Depending on the funding, collaborators and of course technological developments, we estimate to digitise most of our collection within 20 years. Naturally, there is considerable uncertainty associated with an extensive and long-term effort such as this. We expect developments in technology, both in terms of digital storage, digitisation equipment and of the tools that will enable retrieval in a way that reduces

<table>
<thead>
<tr>
<th>Type</th>
<th>Digital now</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newspapers</td>
<td>208,000</td>
<td>5,000,000</td>
</tr>
<tr>
<td>Pictures</td>
<td>450,000</td>
<td>1,300,000</td>
</tr>
<tr>
<td>Books</td>
<td>110,000</td>
<td>450,000</td>
</tr>
<tr>
<td>Hours of moving images</td>
<td>600</td>
<td>250,000</td>
</tr>
<tr>
<td>Manuscripts</td>
<td>23,000</td>
<td>4,000,000</td>
</tr>
<tr>
<td>Maps</td>
<td>50</td>
<td>180,000</td>
</tr>
<tr>
<td>Audiobooks</td>
<td>40</td>
<td>2,500</td>
</tr>
<tr>
<td>Hours of music</td>
<td>5,000</td>
<td>75,000</td>
</tr>
<tr>
<td>Posters</td>
<td>7,200</td>
<td>80,000</td>
</tr>
<tr>
<td>Hours of radio</td>
<td>390,000</td>
<td>1,200,000</td>
</tr>
<tr>
<td>Periodicals</td>
<td>300</td>
<td>2,000,000</td>
</tr>
<tr>
<td>Ephemera</td>
<td>1,000</td>
<td>1,900,000</td>
</tr>
</tbody>
</table>

Note: In total this currently amounts to 3,000 Terabyte (Tb), or 3 Petabytes (Pb) of digital content.
the need for manual indexing. However, it is difficult to predict accurately the consequences of such a development in terms of both cost and increased efficiency.

From 2009 we have had a substantial increase in our funding for this project.

2.4 Digitisation on-demand
Systematic digitisation is the foundation of the digitising process. In addition, we will carry out on-demand digitisation which can be both initiated by our users and based on the strategic priorities of the National Library. On-demand digitisation will be given priority over the systematic work. We have agreements with several other libraries, museums and research institutions to digitise material of interest to users, be it material from the National Library or of material interest to us from the other institutions. In addition, we digitise material for use in national anniversaries where the National Library is responsible, for example the celebration of authors like Ibsen or Hamsun.

2.5 Prioritisation
There is also a challenge in deciding how to prioritise material for digitisation. Should we start with what we consider most important in a cultural heritage perspective, or should we cover one medium at a time? Alternatively, should we be most concerned about what is important from a preservation aspect – or should we primarily digitise and give access to what is demanded? Should we give priority only to the material where the rights are free, out of print, or orphan works, in order to minimise the difficulties in negotiating agreements with the rights holders? Our solution is a combination of mass digitisation and digitisation on demand. This is a challenge for our workflow, since it takes different work tools and different approaches for the digitisation of the various types of media.

3. The digital long-term repository
The digital long-term repository is an infrastructure for the long-term preservation of digital objects. Everything that is digitised as part of the National Library’s digitisation programme is to be preserved as digital objects in the National Library’s digital long-term repository.

The digital long-term repository separates the use of digital content from the technology that is employed for the storage. This allows for easy migration to new generations of storage technology without affecting the systems for retrieval of the digital content. This is very important for the 1,000-year perspective. All digital content is stored in three copies on two separate storage media in the digital long-term repository. At present one copy is stored on disc while two are on tape.

The National Library of Norway has the following strategies with respect to its digital long-term repository:

- *To be a trusted repository for digital material in Norwegian society.* In a digital environment where information on the internet changes frequently, an important issue is which will serve as the stable element when we go from the permanent to the transient. Who will secure the data integrity and organise the media so that they can be found in 1,000 years’ time, convert them when necessary and be trusted repositories?
An important strategy for the National Library is therefore to be such a trusted repository.

- *To be a digital archive for other institutions, e.g. publishers and newspapers.* Apart from the national broadcasters and newspapers that have already been mentioned, the National Library of Norway functions as a digital archive for various photograph and music collections. As part of this, we also have agreements to disseminate information, within the scope of copyright of course. Different institutions, within and outside the library world, want to discuss with us the possibility of us functioning as their digital archive. This is an important strategic challenge. Which criteria should be used to define the priorities and who should pay? We have chosen to give priority to newspapers that are much in demand, and, if appropriate, we also may split the cost with the newspaper publishers. It is also important for us to be able to give access to as much as possible, for research and documentation.

Of course, the national libraries around the world will also be trusted repositories for the cultural heritage of their respective countries. In Norway, we are developing systems and standards for preservation and are aware of the necessity for a sustainable quality, particularly for others who need trusted repositories, such as for research data, and also want to co-operate with us.

In future there will have to be co-operation between the public and private sector, this is an important strategy for us.

- *To find strategic partners for co-operation.* It is important to co-operate with other libraries, knowledge and cultural institutions to see to it that we do not duplicate the digitisation of material that may be found in more than one place. We regularly meet with the university libraries in Norway and the Norwegian national agency for ALM – archives, libraries and museums (Østby, 2003) to co-ordinate the development of digitisation. In Spring 2009 the Norwegian government presented a white paper to the Parliament on issues in the ALM sector, where the National Library, in co-operation with the National Archives, is given responsibility for the long-term storage of digital cultural heritage.

In addition to the strategic partners in broadcasting, newspapers, publishers and so on, we are also in contact with the Research Council of Norway to see how we can be partners in research programmes. We also have meetings with system providers and others to follow up on relevant developments in the technology. An example of this is our participation in the LongRec national research project, which has the primary objective to develop a “persistent, reliable and trustworthy long-term archive of digital information records with emphasis on availability and use of the information” (www.longrec.com). International strategic co-operation, by being involved with other institutions in the International Internet Preservation Consortium (IIPC (www.netpreserve.com)), is also important to us.

4. The Bookshelf and implementation challenges
There are several organisations offering cultural institutions the ability to digitise their book collections for free, or at a very low price, in return for asking for the right to store
the digital copies and to offer search and display of the books. Examples of this are Google Books and the Internet Archive. In our National Library book collection there are relatively few books that are no longer protected by copyright and to which free access accordingly can be offered. It is important to the National Library that books still under copyright will only be stored in our digital long-term repository. Access to such books will only be given in accordance with agreements made with rights-holders. Also, it has been a fundamental principle in the National Library that other service providers must be given equal opportunity to offer services based on our collections. We invite search engines to disseminate what we have digitised and invite other institutions to establish services as added value to our digitised material.

It has also been part of the picture that the National Library has been able to re-assign existing staff to tasks associated with the digitisation production line. This has meant that the total cost of in-house digitisation of books is lower than it would have been for outsourcing the digitisation. For other types of material, such as newspapers on microfilm, we have chosen to outsource the digitisation.

The challenges that we have faced in digitising our book collection include the following areas:

- **Technical quality.** The quality needed for preservation purposes must be much higher than what is normally needed for dissemination. At the same time the formats used for mediation have a shorter lifespan, partly because new advanced formats are constantly developed and partly because of improved versions of existing formats. We have therefore chosen to generate formats for dissemination based on the formats for preservation. The resolution requested by the user is generated at the moment when the image is sought by the user. Our preservation format is JPEG2000 and the dissemination format is PT. JPEG.

- **To give access to as much as possible of our cultural heritage, e.g. through search engines.** The issue now is how we should take advantage of having established a digital repository for preservation in order to give access to this material both to scholars, students and the public. This is a challenge in many respects, not only when it comes to the technical quality as discussed above, but also when it comes to copyright and dissemination aspects.

  In order to enable searching across large aggregations of data, the National Library has chosen to employ search engine technology rather than traditional database solutions. Both metadata and full text are indexed by the search engine, and searches are performed without regard to types of material. We have also implemented a so-called “drill-down” search in the metadata. Metadata for objects satisfying the search criteria are analysed in real time during searches, and alternate paths of navigation and different ways of narrowing the search results are built and displayed to the user. The search engine used is delivered by FAST.

- **Dissemination pedagogy.** National libraries have always given access to their material through the use of metadata. The digital access challenges the way we manage our collections and organise that knowledge.
Digital access to fully digitised material on the net also creates expectations about what kind of services we should provide. To what extent should we give added value to the material, such as in providing a blog, or giving access to the material in a way that is done in social networking sites such as YouTube and Facebook? We are in constant and running discussion with experts about this.

- **Internet pedagogy.** It is important to take part in discussions related to the development of the pedagogy of the internet. Librarians' knowledge about their users' needs is a good basis for developing a search methodology at the intersection of metadata and the methodology of search engines. Our aim is to give access to information, knowledge and experience on a given topic across media types. We create a common search engine for all media types through the integration of different catalogues.

- **Copyright and user behaviour.** The intellectual property rights are a challenge. Negotiating with the rights-holders is therefore a vital issue. What kind of models should there be for paying for use? In the Nordic countries there is a tradition of collective models. When digital use makes it easier to define and trace the use of each rights-holder's artistic works, will this be a pressure towards individual payment for use? How can we define who the rights-holders are? To what extent will there be licensing instead of permanent agreements for use? Is it possible to have agreements based on the principle for printed material also in the digital, or is the digital environment a separate starting-point for finding solutions? Is the fact that books are available through a Digital National Library a marketing point for commercial use, and can it be a factor in negotiations? Should the public fund the libraries' payment for rights – or will this new situation open up more pay-services in national libraries? This is of course already different among the European countries and national libraries in the whole world.

5. **Conclusions**

The agreement that the National Library of Norway has signed with Kopinor will provide insight into the issues of copyright and user behaviour. Through this project we hope to understand more about user patterns, frequency of use, interest in digital text on the internet and so on. Will the users read books online, will they download and if so, when and how? Are there differences between different kinds of material, such as fiction and reference works?

To be in the forefront of the development of digitisation services in libraries, it is important to analyse how digital services are used in other sectors than libraries, and to try and foresee the alternative scenarios for development. Who will own the digital future? Who adds value or where is value added in relation to knowledge? What will the end product be? Will the digital content of the national libraries be widespread in a large number of copies, or will everyone access it from centralised repositories? These are important challenges for all national libraries.

**References**


Further reading

Corresponding author
Vigdis Moe Skarstein can be contacted at: vigdis.skarstein@nb.no

To purchase reprints of this article please e-mail: reprints@emeraldinsight.com
Or visit our web site for further details: www.emeraldinsight.com/reprints