

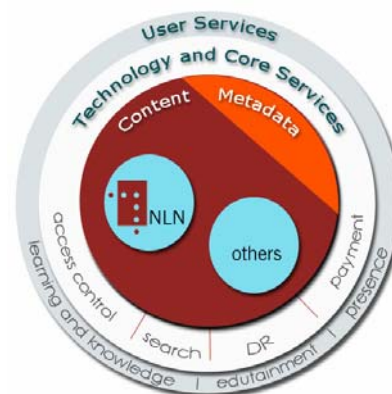
Digitising the Collection at the National Library

The National Library of Norway, May 2006

1 The Strategy

The National Library of Norway has a strategic manifesto concerning the Digital National Library. The manifesto concludes that

- The digital library shall be presented in ways users prefer and in a context which is tailored for their needs. The user shall be able to search both metadata and content using his preferred tools.
- The digital resources at the National Library shall be readily available for use in education and research, e.g. as learning objects suitable for use both in general and in specific contexts.
- Information and knowledge shall be combined into new contexts to give the users new experiences and knowledge. Services including digital resources in challenging and surprising ways as networked exhibitions or edutainment, shall be developed continuously.



To implement this strategy, the National Library decided to focus along three main areas in its digital library programme.

- New services must be developed, and the services must be attractive enough to compete with other alternatives on the Internet.
- The collection at the National Library must be digitised to improve availability for a broad audience and for professional use.
- The digital collection at the National Library shall be safe-guarded through further development and application of a digital repository at the National Library. The National Library shall also offer digital preservation services to other institutions in Norway.

As a consequence of the manifesto, the National Library has launched a strategic initiative where the objective is to digitise all of the collection at the National Library to make it available in the Digital National Library, and to preserve the collection in a digital format.

In addition to the digitisation initiative, the National Library of Norway shall establish systems for depositing digital born material to the library, even when the material is published as physical items or broadcasted. This will reduce the growth of material queuing up for digitisation. A few digital deposit systems are already in place, and a dialogue with a broad range of publishers to make necessary agreements, have been initiated.

2 Digitisation Initiative

For more than 10 years, a large variety of media types have been digitised at the Media preservation division at the National Library. The main focus of this digitisation has been photos, sound recordings and microfilm (newspapers). As a result of this, the digital collection currently holds more than 50 000 hours of radio broadcasts (in linear quality), more than 200 000 photos, and more than 200 000 newspapers.

To speed up the digitisation of the complete collection, a new digitisation initiative has been established in 2006. The existing digitisation of audiovisual material continues, and in addition € 1,25 million have been allocated in 2006 for the new initiative.

The new initiative will focus on digitisation of books and journals. As a starting point, all the Norwegian material in our safe deposit storage in Oslo will be digitised. The safe deposit storage contains the oldest, most fragile, and most valuable books, and counts a few thousand items. For this purpose we will use scanners where operators manually turn the pages between each scan. The most fragile material will be digitised using high quality scanners in collaboration with conservators in our conservation laboratory in Oslo. All of this material is too old to have any copyright constraints, and will thus be made available in our digital library without any restrictions.

The high quality scanners will also be used for digitisation of selected historical maps, manuscripts, and posters.

In addition, we will digitise material from our repository library in Mo i Rana. To get an effective digitisation, we will select material where we have several duplicates in our collection. One of the duplicates will be de-assembled and digitised in automated scanners. After digitisation, the de-assembled book will be thrown away. Our first calculations show that we can digitise more than six times more with automated scanning as compared to scanning with manual turning of pages.

The automated scanners will be placed in our microfilming section. Over the next year, conventional microfilming activity will be reduced (to be replaced by digitisation and digital deposit of newspapers from the publishers). As a part of this process, the staff at the microfilming section will be trained to operate the scanners. We also plan to recruit people internally for retrieval and transport of material from the collection to the digitisation, and also for the process of preparing the material for digitisation (including de-assembling and scanning covers).

Apart from this, more of our microfilmed newspapers will be digitised. This part of the digitisation will most likely be outsourced.

The budget for this new initiative will thus be used for digitisation equipment, IT staff for establishing automated post-processing of the digitised material, and outsourcing of some of the digitisation.

In addition to this, € 2 million have been allocated in 2006 for further development of our digital repository, and for increasing the digital capacity of the repository.

The books, the journals, and the newspapers may be copyright protected material. To make these items available in the digital library, we need agreements with the copyright

holder's organisations. The dialogue with these organisations will be carried out in parallel with the initiation of the digitisation work.

3 Handling the Digital Material

The most important objective for the digitisation is to make the collection at the National Library available in the National Library's digital library. It is therefore an important issue to make an item from the collection available in the digital library in a suitable format as soon as possible after digitisation.

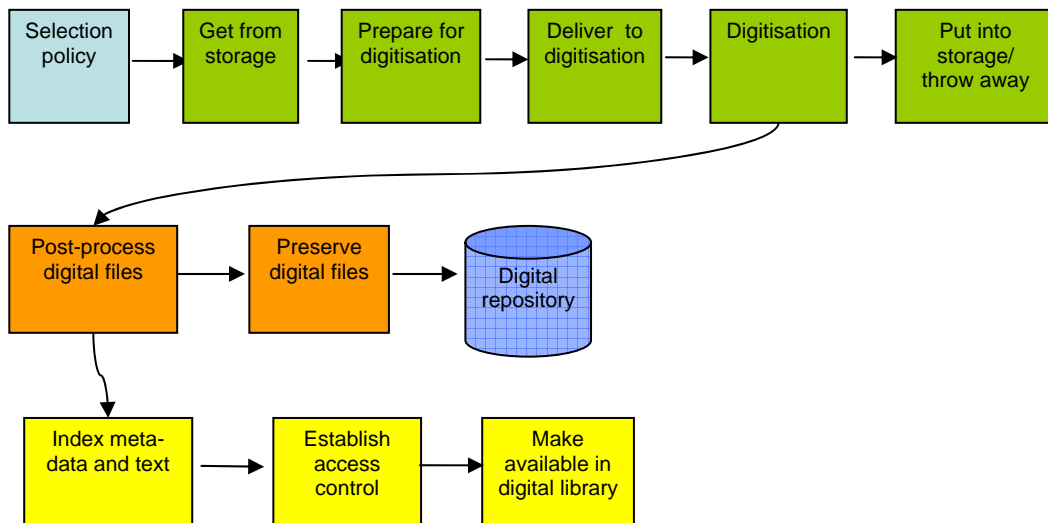
An automated post-processing chain is therefore being established. Some of the actions to be performed during the post-processing are:

- to bind together digital images constituting an item in its physical format
- to update metadatabases (remove the physical item and insert the digital id)
- to perform OCR on textual information
- to identify structure in the textual information and generate XML-tagged format
- to index metadata and textual information in our search engine, and make the index a part of our digital library search service
- to assess access restrictions and insert them into the authentication and authorisation system
- to generate access formats of the digital information
- to insert all digital data and metadata into the digital repository

We plan to use off-the-shelf software whenever possible. However, it will be necessary to do in-house development and integration to establish a streamlined operating system adapted to our digital library infrastructure. We will use a flexible architecture when engineering the system, making it easy to insert, remove and update processes in the post-processing.

This may be an interesting area for collaboration between national libraries around the world which are carrying out digitisation programmes - exchanging experience on off-the-shelf software, digitisation hardware, and even exchanging in-house developed processing modules.

A simplified model of how the processes connected to the digitisation may look like:



4 Timeframe

A first pilot for the digitisation of books is already in place. The purpose of this first pilot is to evaluate the scanning environment, and to produce digital material for the evaluation of OCR software and software for structure recognition. The next step is to have the new scanners in place during the summer, and to start small scale production early autumn 2006, to be increased in scale during the winter. Full scale production is hopefully in place early 2007.

It is still too early to estimate the total production objectives for 2006 and 2007, but our aim is ambitious.