

## Comellus-project - a step towards digital newspaper deposition

### Abstract

Historical newspapers are an important resource for researchers and also of interest for the common people. They contain a great deal of information on the history of politics, society and everyday life. The Comellus project aims to improve the production and accessibility of online newspaper material by piloting a fully digitalised newspaper handling process at the National Library of Finland.

The National Library has digitised and made available a large part of its newspaper collections in digital format through its Historical Newspaper Library ([digi.nationallibrary.fi](http://digi.nationallibrary.fi)). The digital newspapers are derived from printed newspapers by microfilming, scanning and post-processing the scans. This demands a great deal of time and human resources.

The piloted new process is based on the utilisation of print-PDF files, digital printing plates of a kind, used in today's newspaper printing process. They contain the exact images of the printed pages and can thus be used as digital surrogates for them. The aim of the project is to create a process where the newspaper publishers will automatically deliver the print-PDF files and associated metadata to the National Library.

By utilising the provided metadata along with the print-PDFs, the deposited newspapers can be made available in digital format almost instantly. In addition, the long term preservation of the digitally deposited newspapers is guaranteed by producing a microfilm using Computer Output Microfilm technology.

### Introduction

The central objective of preserving our published cultural heritage is to meet the needs of scientific research. Newspapers provide important information to both scholars and the general public. They are vital source material particularly for historical, cultural and social scientific researchers. Through newspapers, we obtain information about everyday history and political movements, from major events to microhistory. Communication studies is a key discipline that uses mass communication resources: each year, some 400 major research projects and more than 100 Master's theses are completed in the discipline in Finland.

The National Library preserves our published cultural heritage especially for research purposes. In its digitisation policy from 2010, the National Library stated that its materials should be provided to the general public as freely and comprehensively as possible within the scope of the Copyright Act. Currently, Finnish newspapers published prior to 1912 are freely available in digital format in the National Library's Historical Newspaper Library.



*The National Library's Historical Newspaper Library ([www.digi.nationallibrary.fi](http://www.digi.nationallibrary.fi)) provides access to Finnish newspapers published prior 1912 free of charge.*

In our increasingly digital world, consumer demands have increased in line with requests for online access to resources. Both scholars and members of the public can benefit from being able to access the digital format, as it enables them to carry out extensive searches in newspapers which may be difficult when done manually. The use of source material is directly related to ease of access for researchers.

Newspapers are usually printed on relatively cheap paper that is difficult to preserve. The National Library microfilms newspapers to preserve their content and then copies the films for customer use. The National Library's high-quality microfilms are expected to last some 500 years, guaranteeing that the content of these fragile newspapers is preserved far into the future.

### **Towards a digital newspaper process**

The Comellus project aims to create a new, efficient process in which publishers deposit newspapers in the National Library in digital format. Newspaper printing plates are created using digital (often PDF) page images, or print-ready PDFs. In practice, such files are digital replicas of the print newspaper. The new digital newspaper process will mean that the print-ready PDFs of publishers as well as related metadata from editorial systems are automatically deposited in the National Library.

It will be easier and more cost-effective to provide scholars and the public with access to digitally deposited newspapers via the National Library's Historical Newspaper Library, as the material need not be digitised or post-processed.

The National Library will exploit the metadata deposited together with the print-ready PDFs. Editorial systems may contain data about page structures, articles and authors or even geographic

information, some of which may be used directly in the National Library's own publishing system when providing access to deposited digital newspapers.

The National Library will ensure the long-term preservation of deposited material by converting print-ready PDFs to microfilm through the Computer Output Microfilm (COM) process. This stage will replace the microfilming of printed newspapers in the digital newspaper process.



*The microfilms produced from the digitally deposited newspapers by using COM-printing technology are expected to last at least 500 years.*

### **Digital processes for maximising efficiency and usability**

The project aims to develop a more efficient work process for handling newspapers at the National Library's Centre for Preservation and Digitisation. The digital process will increase efficiency, primarily by reducing the need to manually scan and post-process newspapers. In addition, some time and resources will be saved in the preparation of newspapers for microfilming

The National Library is also seeking to offer a greater variety of digital newspaper material. The utilisation of metadata from editorial systems in the National Library's publishing systems will allow the construction of more user-friendly interfaces with diverse search functions and speed up the actual process of making material available. With less manual work required, deposited digital material can ideally be made available to customers almost immediately. Copyright issues naturally impose some limitations.

### **Challenges involving the diverse data systems of newspaper publishers as well as legal issues**

The Comellus project must address the technical challenges posed by the various editorial systems and metadata production methods of newspaper publishers. As a rule, a separate system integration must be carried out for each editorial system to enable the automatic submission of metadata and

print-ready PDFs. The extent and quality of the metadata produced varies a great deal from newspaper to newspaper. The National Library will face a challenging task in incorporating heterogeneous metadata from various editorial sources while ensuring that the metadata can be used appropriately.

Other challenges include adjusting new digital operating models to other in-house processes. For example, the information systems developed in the project must support the parallel processing of printed and digital newspapers.

Major challenges are also related to copyright issues and the Act on Collecting and Preserving Cultural Material. As the Act currently requires publishers to deposit printed newspapers rather than print-ready PDFs, separate deposit agreements must probably be signed with newspaper publishers.

Due to copyright issues, the provision of digital access to deposited newspapers is also likely to be based on agreements. Negotiations on such agreements must usually be conducted separately with each newspaper publisher, which takes time and resources. What is more, individual journalists and photographers may also hold the copyright on some material, in which case agreements with publishers are not sufficient. Kopiosto (the Finnish copyright organisation for authors, publishers and performing artists) may play a significant role here.

### **Beginning with system design and process modelling**

The Comellus project commenced with the design and implementation of the information system needed to receive print-ready PDFs and ensure quality. Developing this system is a key component of the project, for the new digital newspaper process will be largely based on the system.

All system design and development will be conducted within the project to ensure that the final product is compatible with in-house processes and can be integrated seamlessly with existing in-house systems. Where possible, development work will consider not only newspapers, but also other types of publications, such as journals and monographs.

The system developed has already been tested for the deposit of print-ready PDFs, and the results were encouraging. In future, the aim is to expand the process so that the National Library also receives the descriptive metadata associated with the print-ready PDFs. Detailed planning and modelling of the digital newspaper process has begun and will continue as the project progresses.

### **Project funders and partners**

The Comellus project will strengthen Mikkeli's status as a hub of digitisation, archiving and electronic services in Finland. The project is part of the establishment of a research, development and training centre within the Digitalmikkeli network to promote knowledge transfer and synergy in related development projects.

The Comellus project is funded by the European Social Fund (ESF/"Leverage from the EU"), the South Savo Centre for Economic Development, Transport and the Environment, the City of Mikkeli and the National Library.

The National Library's partners in the Comellus project include the publisher of the Länsi-Savo newspaper, Etelä-Savon Viestintä Oy, and the publisher of the Etelä-Suomen Sanomat newspaper, Esan Kirjapaino Oy. Also involved is Anygraaf Oy, system supplier for the above newspaper

publishers. Anygraaf holds a major market share in Finland in the provision of editorial systems for newspapers and magazines. By cooperating with Anygraaf and the pilot newspapers, the project aims to develop an approach and metadata specification that serve the publishing sector as widely as possible.

## **Summary**

The main goal of the Comellus-project is to create a process for digital deposition of newspapers. The focus is on sub-processes related to deposition, reception and handling of digital newspaper material at the National Library of Finland. During the project, the information systems required by the new processes are implemented. The long term preservation of the digitally deposited newspapers will be guaranteed by COM-printing them on microfilm. Along with the digital page images, also some related descriptive metadata will be deposited. The metadata can be utilised for instance when the deposited newspapers are made available in digital format

## **Links**

National Library of Finland URL: <http://www.nationallibrary.fi>

Historical Newspaper Library URL: <http://digi.lib.helsinki.fi/sanomalehti>

Kopioisto URL: [http://www.kopioisto.fi/kopioisto/en\\_GB/](http://www.kopioisto.fi/kopioisto/en_GB/)

Digitalmikkeli URL: <http://www.digitalmikkeli.fi/inenglish>

European Social Fund (ESF) URL: <http://www.rakennerahastot.fi/rakennerahastot/en>

South Savo Centre for Economic Development, Transport and the Environment URL:  
<http://www.ely-keskus.fi/en>

City of Mikkeli URL: <http://www.mikkeli.fi/en/english>

Länsi-Savo URL: <http://www.lansi-savo.fi>

Etelä-Suomen Sanomat URL: <http://www.ess.fi>

Anygraaf Oy URL: [http://www.anygraaf.fi/fin/eng\\_frontpage](http://www.anygraaf.fi/fin/eng_frontpage)

## **Further information**

Project Manager Matti Hosio, National Library of Finland, Centre for Preservation and Digitisation,  
[matti.hosio\(at\)helsinki.fi](mailto:matti.hosio(at)helsinki.fi)