

2015 IFLA International News Media Conference

MUNIN - a monitoring system for e-legal deposit of newspapers

Abstract

The National Library of Norway (NLN) is going through a massive digitization program. This program aims at digitization of the entire analog collection in the library, digital legal deposit, digital long term preservation and digital library user experiences. This will take several years to complete and some tools are needed to make it effective and uncluttered.

With over 250 different publications of newspapers, varying from being issued once a month to twice a day NLN sees a real need for a monitoring system. The goal is for NLN to receive all active newspapers digitally so time could be spent on scanning the preserved historical material. PDF's in print quality are delivered to the library based on automatic file transfer during the night. By implementing a monitoring system NLN can effectively receive and supervise the PDF papers, as well as make us independent with the opportunity to correct technical faults directly and automatically notify publishers if errors occur.

A direct consequence of digital legal deposit is that Norwegian Libraries who subscribe to microfilm of some of the newspapers from NLN will get direct access to the digital newspaper via our online service, saving us both time and resources. At this moment 30 newspapers is delivered digital to NLN.

The presentation will present and demonstrate the system developed to support us with e-legal deposit newspapers.

Presenters:

Gretel Mari Westman, advisor (Periodical Section)

Gretel.westman@nb.no

phone:+47 75 12 11 06

Mona Løkås, head of Periodical Section

Mona.lokas@nb.no

phone:+47 75 12 12 09

MUNIN - a monitoring system for e-legal deposit of newspapers

1. Becoming a digital national library

The National Library of Norway (NLN) started the digitizing program in 2006. The main goal was to digitize the entire collection, not only the paper based media but also radio, broadcast, photo and so on. It is estimated that it will take 20–30 years for the entire collection to be digitized. The Norwegian Legal Deposit Act requires that all published content, in all media, be deposited with the National Library of Norway. The collection is also being expanded through purchases and gifts. The digital collection contains material dating from the Middle Ages up to the current day.

Becoming a national digital library means:

- Digitizing the non-digital collection
- Collecting the digital born masters of non-digital publications
- Collecting publications that only exist in digital formats
- Cooperation with copyright holders and publishers to be able to make the collection digital, and make the digital collection available to all the users

2. Newspapers in the National Library of Norway

NLN's collection of newspapers include about 55 000 binders, 25 000 boxes of single editions, 70 854 microfilms. In addition to the physical news media we have harvested the webpages of the 254 Norwegian newspapers daily.

Today all publishers of newspapers produce their publications in digital format. NLN strives to ensure that digital sources for all publications are deposited directly by the publisher, together with the publisher's metadata, such as title, number, edition and volume. This will permit the collection to be expanded to include high-quality digital content as well as increase the size of the physical collection. Metadata received directly from the publisher will also improve the efficiency of cataloguing of material and simplify retrieval.

Today we receive about 30 newspapers digitally. All these titles is received in the same format and naming standard. Resulting in an automatic workflow from the minute it's harvested from the publishers ftp-server to updating and being available in our digital library.

The National Library has also made agreements with newspaper publishers that make it possible to give access to newspaper titles in digital format in all the Norwegian libraries. Some titles are also available for everyone.

3. Digital delivery of newspapers

The Norwegian act of legal deposit came into force in 1990. The act states that NLN should receive all printed media as it is presented to the public. We require a higher quality on the pdf than the publishers use on their web pages and online services, and we therefore have to make contact and conclude an agreement. A renewed act is being considered and NLN is preparing for it to come in to effect. Until the new act is approved a digital delivery of newspapers starts with NLN initiating cooperation with the publishers or the publishers themselves make contact to hear more about this solution. There are many technical guidelines to be followed for the e-legal deposit to go automatic and smoothly. To keep the publishers committed we divide the process into four steps, from the first initiating contact to a done deal and regular delivery.

1. The publisher receive a written guideline and technical requirements sheet
2. The publisher makes a sample delivery of a newspaper in accordance with the technical requirements. This goes on until both NLN and the publisher is satisfied
3. The formal agreement is signed by both parties
4. A regular e-legal deposit is made as described in the formal agreement

Step number two is the most time consuming in the process, consisting with some trying and failing. The NLN are preserving the digital newspaper in a millennial perspective and for the information to survive several generations of converting to new formats it is important that we receive a high quality PDF.

The name standard of the delivery is also important during the automatic receiving of the files as well as being able to find the files amongst millions of files in the future. In a digital delivery of newspapers each page in the publication is an individual pdf. The files are compressed in a .zip og .rar folder with the same name standard.

PDF files:

<newspaper_subname_zone_date_volume_number_issue-
sequensnumber_pagenuumber_appendixname>.pdf

Example:

adressavisen_null_null_20060610_240_130_1-1_002_hovedavis.pdf

aftenposten_morgen_1_20060607_147_253_1-1_001_nyheter.pdf

Compressed folder:

<newspaper_subname_zone_date_volume_number_issue><.zip/.rar>

Example:

adressavisen_null_null_20060610_240_130_1.zip

aftenposten_morgen_1_20060607_147_253_1.rar

All the sections in the name have to be filled and follow these guidelines:

- Format on date is *yyyymmdd*
- Write *null* if the section does not apply to the publication
- If the newspaper has zones, this is a serial number starting with 1. (e.g. 1=Mo i Rana, 2=Oslo, 3=Bergen etc.)
- If the newspaper has more than one issue per day, this is a serial number starting with 1.

The compressed file containing all the PDFs is now put on an ftp server and is harvested automatically by NLN. It is preferable that the publisher manage their own ftp servers. By giving NLN access to their server with username and password we can automatically retrieve publications and effectively include them in a well-established production trail for digital received newspapers. We have experienced that some publishers do not have their own ftp server. This led us to making an alternative route for e-legal deposit where we assign space on our ftp server. The publisher will receive a username and password from us and NLN simply harvest the files from our own server.

Step by step an e-legal deposit transfer proceeds as follows:

- Export publication to high quality PDF (one file per page)
 - Name your files according to the standard specified in the requirements specification
 - Compress and extract files with .zip or .tar
 - Name the zipped file by the NLN name standard
 - Add zip. file in a folder named “delivery to NLN” situated firmly on your ftp server OR connect to our ftp server via an ftp client with the username/password you received from NLN
-
- It has to be room for at least 14 days with a zipped file on your ftp server
 - NLN retrieves the file
 - Naming and format checked
 - The file is removed from the ftp server

4. MUNIN

Approximately 30 newspapers are received digitally at the Norwegian National Library. The publications vary from being published monthly, weekly, six days a week to twice a day, just to mention some. With these differences as well as a constant growing title registry NLN saw a need for a tool, a monitoring system, to help us maintain control and an overview of what was coming in through e-legal deposit. The librarians of the periodical section are primarily the users of MUNIN. The system is designed to have an understandable graphical user interface, be easy to operate, and the single portal to get all the information a librarian need about an e-legal deposit of a newspaper.

To start with you can register a newspaper title in the system. This includes the title, a prefix and the publishers contact information. We also register how often the publication is issued and how many days the publisher has to deliver it to NLN before we automatically claim them via e-mail. How many days the publisher has from the publications release date to NLN receiving the publication can also be registered in the title registry. Based on the date, number and volume we can register what we expect the next issue to be named, and we will get a notification if the received issue does not answer to this expectation. This allows us to reveal faults and changes consecutively.

Titler uten avleverende institusjon ?

Adresseavisen

Titel/Kanal	Mediehus	Prefiks ?
<input type="text" value="Adresseavisen"/>	<input type="text" value=""/>	<input type="text" value="adresseavisen"/>
Kontaktperson navn	Kontaktperson epost ?	Avleveringsfrist ?
<input type="text" value="Ivar Erikssen"/>	<input type="text" value="mona.lokas@nb.no"/>	<input type="text" value="100"/>

Utgivelsesmønster ? **Innhøstingsdetaljer** ?

Man	Tir	Ons	Tor	Fre	Lør	Søn
<input type="text" value="1"/>	<input type="text" value="1"/>	<input type="text" value="1"/>	<input type="text" value="1"/>	<input type="text" value="1"/>	<input type="text" value="1"/>	<input type="text" value="0"/>

- Avis Nordland
- Dag og Tid
- Dagens Næringsliv
- Rana Blad
- Snåsningen
- Ytringen

Illustration 1: A screen shot of the title registry in MUNIN, showing title, pattern of issue and contact information.

Titler uten avleverende institusjon ?

Adresseavisen

Nyeste mottatte materiale ?

Dato: 23.07.2018
Nummer: 167
Årgang: 248

Justér forventning for neste utgivelse ?

Dato: 24.07.2018
Nummer:
Årgang:

- Avis Nordland
- Dag og Tid
- Dagens Næringsliv
- Rana Blad
- Snåsningen
- Ytringen

Illustration 2: A screen shot of the title registry in MUNIN, showing what date, number and volume we are expecting.

After the newspaper is registered in MUNIN we can monitor the newspapers in a calendar. You can choose if you want to see just one single title or all of them together. The calendar simply shows one of three symbols on each date. A green “check” indicates that all transfers have been completed successfully. The yellow sign is an indication that something is wrong with the files or the transfer, or it highlights that an e-mail has been sent to the publisher for not delivering the expected issue. The red sign means that the material is not accepted and needs to be managed manually. An example of the red sign appearing would be if we did not register the title before being received.

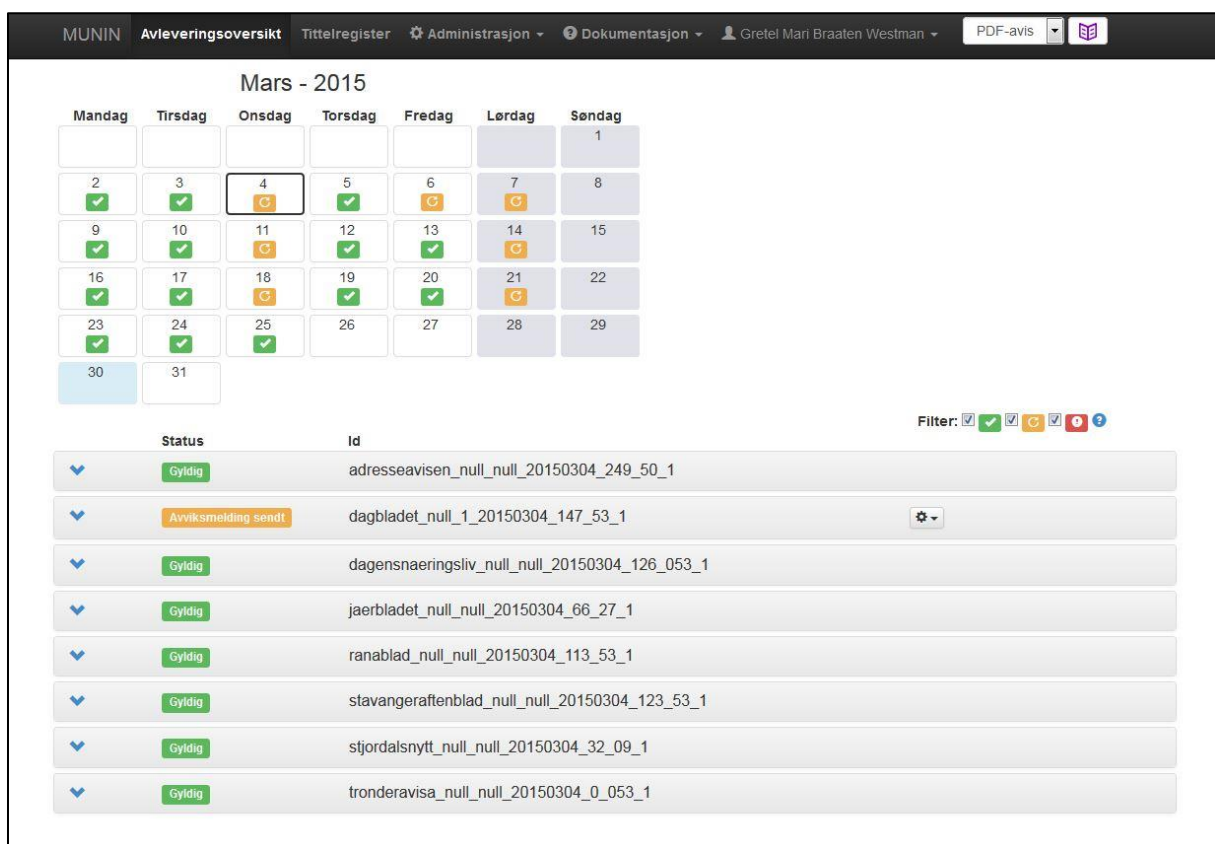


Illustration 3: A screen shot of the calendar in MUNIN.

The title registry and the calendar are the functions used on a daily basis, but MUNIN also have functions like statistics, a listing of all harvesters and their status and a listing of all claim e-mails sent to the publishers. Today MUNIN is customized to receiving newspapers and periodical material, but the technology can be adapted to other types of media. In the future MUNIN may be used to monitor e-legal deposit of broadcasted materials, books and posters. MUNIN gives a sense of control and makes librarians independent.