

# Yes Please, Both!

## Large-Scale Digitization and Legal Deposit of Newspaper in Norway

IFLA Pre-Conference, 14th-15th august 2013,  
Singapore

Newspapers to the People

Mona Løkås



# I will talk about...

- Our Digitization Program
- Legal framework
- How we do it
- Key principles and experience
- Examples



# Digitization Program

## **2005:**

The National Library of Norway (NLN) decided to start a massive digitization program:

- Digitize the complete analog collection

## **2013:**

- Digital legal deposit (newspaper, radio, TV)
- Digital long term preservation (established, ca 4 petabyte)
- Digital library user experience ([www.nb.no](http://www.nb.no))



# Legal framework

- NLN has rights to perform digitization for in-house long-term preservation
- NLN has rights for in-house user access to the digitized collection
- For newspapers especially: agreements with publishers will give access to all the Norwegian libraries



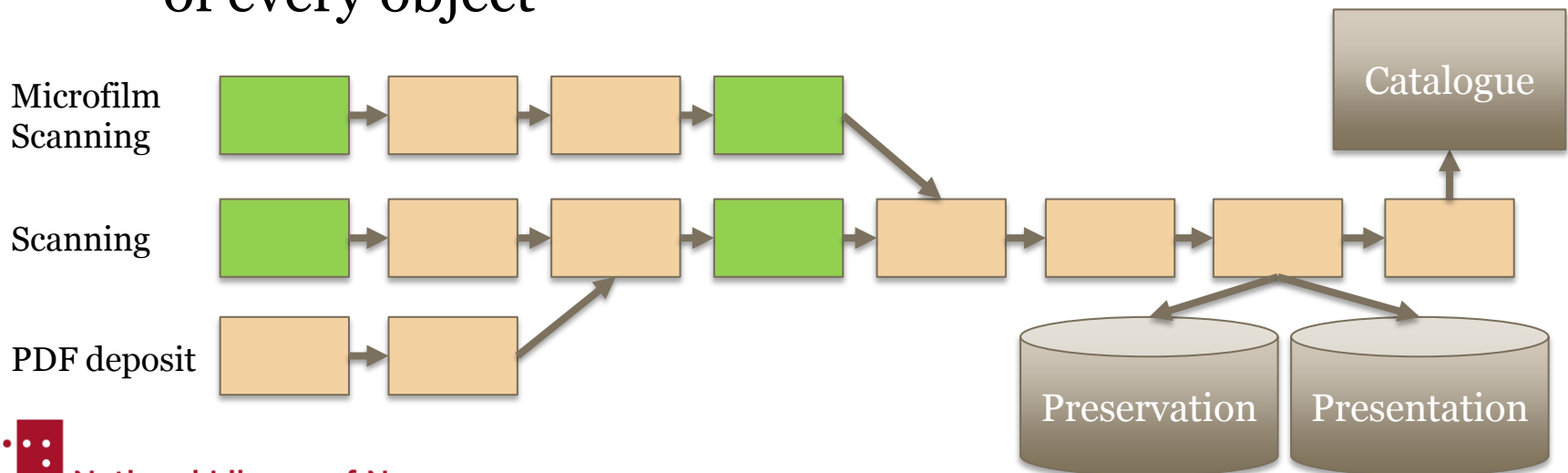
# Digitization principles

- **Uniformity:** All objects are subjects to a minimum of QA and processing to ensure a uniform format, naming and structure
  - Metadata: Descriptive, Technical, Preservation, Structure
  - Unique Identifiers: URN
  - Format: JPG2000, XML, PDF
  - Quality: 400dpi, 24 bit colour
  - Content: Images, text
- **Automation:** All steps (as far as possible) are automated to ensure uniform and consistent handling
- **Preservation:** All objects are stored in a generic preservation infrastructure in a lossless quality
- **Presentation:** All objects are indexed in a search engine and made accessible on nb.no according to access policy



# Production lines

- **Automated steps** designed to take care of specific tasks.
  - Validation, processing, metadata generation, OCR
- **Manual steps** for scanning, QA and error handling
- **Production database** holds information and status of every object



# Standard for naming of files

The filenames of the digital newspaper follow a home-made standard:

Aftenposten\_Morgen\_1\_20111011\_152\_391\_1.zip

Aftenposten\_Morgen\_1\_20111011\_152\_391\_2.zip

This naming standard makes it possible to automatically transfer all the files to its own metadata system and without no adding of further metadata.





Dag og Tid 2002.12.07 - Nasjonalbiblioteket - Windows Internet Explorer

http://www.nb.no/nl

Dag og Tid 2002.12...

Google

< Previous issue Next issue >

Page 1 / 32

# DAG OG TID

FRI VEKEAVIS FOR KULTUR OG POLITIKK

Nr. 49, laurdag 7. desember 2002 41. årgang Laussal 20 kroner

**Oogabooga Bush**  
Bestefar til George W. Bush plyndra grava til hovding Geronimo og spona nazistene. Gamle Bush (bilet), sønn og sønnesønn var med i broerskapen Skull and Bones.

**Dei nyreaksjonære**  
Finst det ein ny klasse av reaksjonære intellektuelle som har sprunge ut frå venstresida? Det er ein debatt som går i Frankrike med utgangspunkt i boka *Attende til lov og orden*, som tek eit oppgjør med «dei nye reaksjonære».

**Uvener i ørkenen**  
Antropologen Malinowski og for taren Witkiewicz var barndomsvenner. I 1914 skilde dei lag for godt i dei australske ørkenene, eit emne for i romanen til Agneta Plejel (bilet).

BAKSIDA SIDE 18-19 SIDA

**Trakassert og prisnor**

**Døds-skvadronar**

Dag og Tid 1966.03.26 - Nasjonalbiblioteket - Windows Internet Explorer

http://www.nb.no/nl

Dag og Tid 1966.03...

Google

< Previous issue Next issue >

Page 1 / 16

UNIVERSITETSBIBLIOTEKET  
DRAMMENSVN. 42 B  
OSLO 2

HOV KONTORMØBLER  
SYSTEM HALVPULT

# DAG OG TID

NR. 10 - 5. ÅRGANG - OSLO 26. MARS 1966 - PRIS KR. 1,50

RIK KOM Partipol

## Journalisten som var «kjel

i TROMSØ miste

Dag og Tid 2013.06.28 - Nasjonalbiblioteket - Windows Internet Explorer

http://www.nb.no/nl

Dag og Tid 2013.06...

Google

< Previous issue

Page 1 / 40

**SJAKK** Carlsen er god på dårlege dagar òg. | SIDE 19

**USA** | Optimismen på Key West er utan ende. | SIDE 15-17

FRI VEKEAVIS FOR KULTUR OG POLITIKK

**A-AVIS**

**KOMMENTAR** | SIDE 2  
Eg hentar fram ei gammal kveldsbøn. Kjære tog! Talk for alt som eg har fått, kunne du nå berre ha gått, skriv Sigrun Slagstad om erfaringane sine med NSB.

# DAG OG TID

NR. 26  
28. JUNI - 4. JULI 2013  
52. ÅRGANG  
KR 45,-

**UTANRIKS** | SIDE 6-7  
**Folketelt.** Nelson Mandela var folketeltar i kirseid. Men heller ikkje han kunne løysa dei djuppe problema i Sør-Afrika, skriv Kaj Skagen.

**INTERVJU** | SIDE 3  
**Nivåfall.** Det store råfallet av studentar lygn av for dårleg grunnutdanning i skulen, meiner Teina-president Lise L. Randeberg.

**LYRIKK** | SIDE 21-23  
«Legg deg hjå meg i mi breie

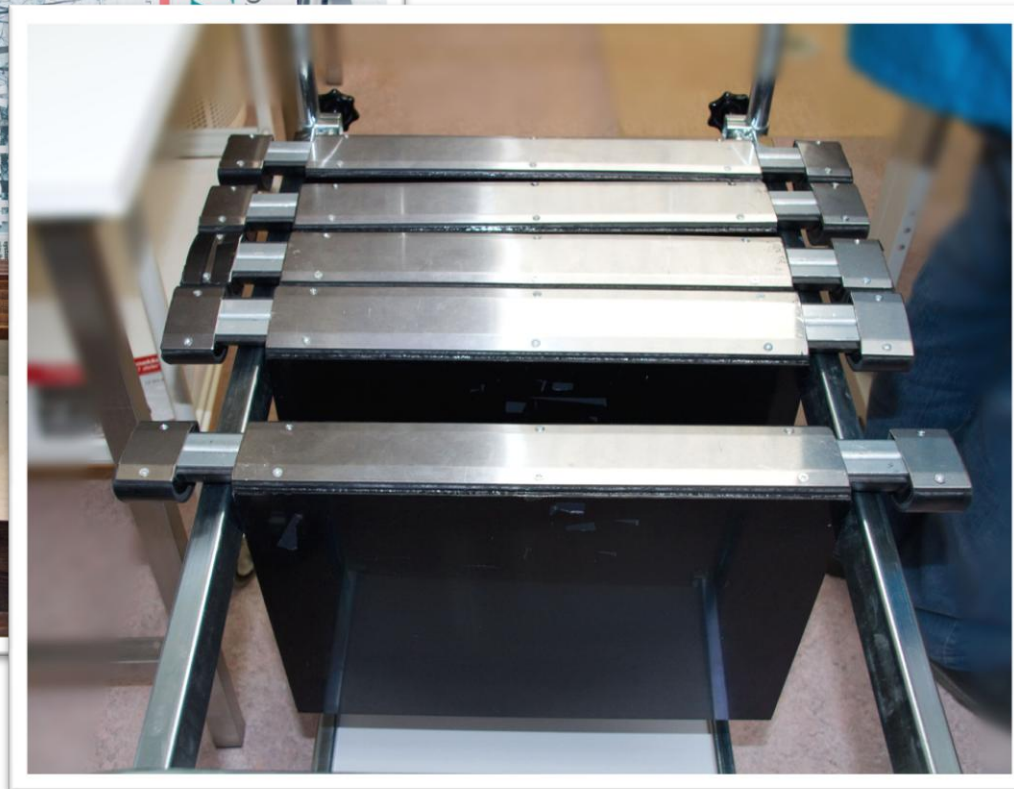
# Syndebukken

**PANIKK:** Norske aviser slår alarm om gift i oppdrettslaksen, sjølv om han er reinare enn villaks og makrell.

**MILJØ** | SIDE 4-5



# Scanning



# Scanning



# Search for digital content

Nasjonalbiblioteket

Tilbud Besøk Hva skjer? Om NB Bibliotekutvikling Presse Kontakt

Søk i digitalt innhold  
avisar, bøker, bilder, håndskrifter, radio og film

Søk **Søk**

Søk i mer enn 100 000 norske bøker i bokhylla

« **August 2013** »

Man	Tir	Ons	Tor	Fre	Lør	Søn
			01	02	03	04
05	06	07	08	09	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	

**Denne uken:**

Omvisning i Nasjonalbiblioteket  
7. aug kl. 14

Arrangementskalender

<http://www.nb.no/Presse>

Ivar Aasen  
Den 5. august i år er det 200 år sidan Ivar Aasen - skaparen av det nynorske skriftspråket - vart fødd. Nasjonalbiblioteket har mykje materiale av og om Ivar Aasen.  
Les mer.

# Advanced search

The screenshot shows a web browser window titled "Avansert søk - Nasjonalbiblioteket - Windows Internet Explorer". The address bar contains the URL "http://www.nb.no/nbsok/advanced\_search?lang=en". The page header includes the National Library of Norway logo and the text "The National Library of Norway". In the top right corner, there are links for "Norsk English", "a a a | Change contrast | Login", and "About us".

## Advanced search - Newspapers

The search interface features a horizontal menu with tabs for "All", "Newspapers", "Images", "Posters", "Books", "Film", "Music", "Radio", "Sound rec.", "Music manuscr.", "Private arch.", "Maps", "Music books", and "Journals". The "Newspapers" tab is currently selected.

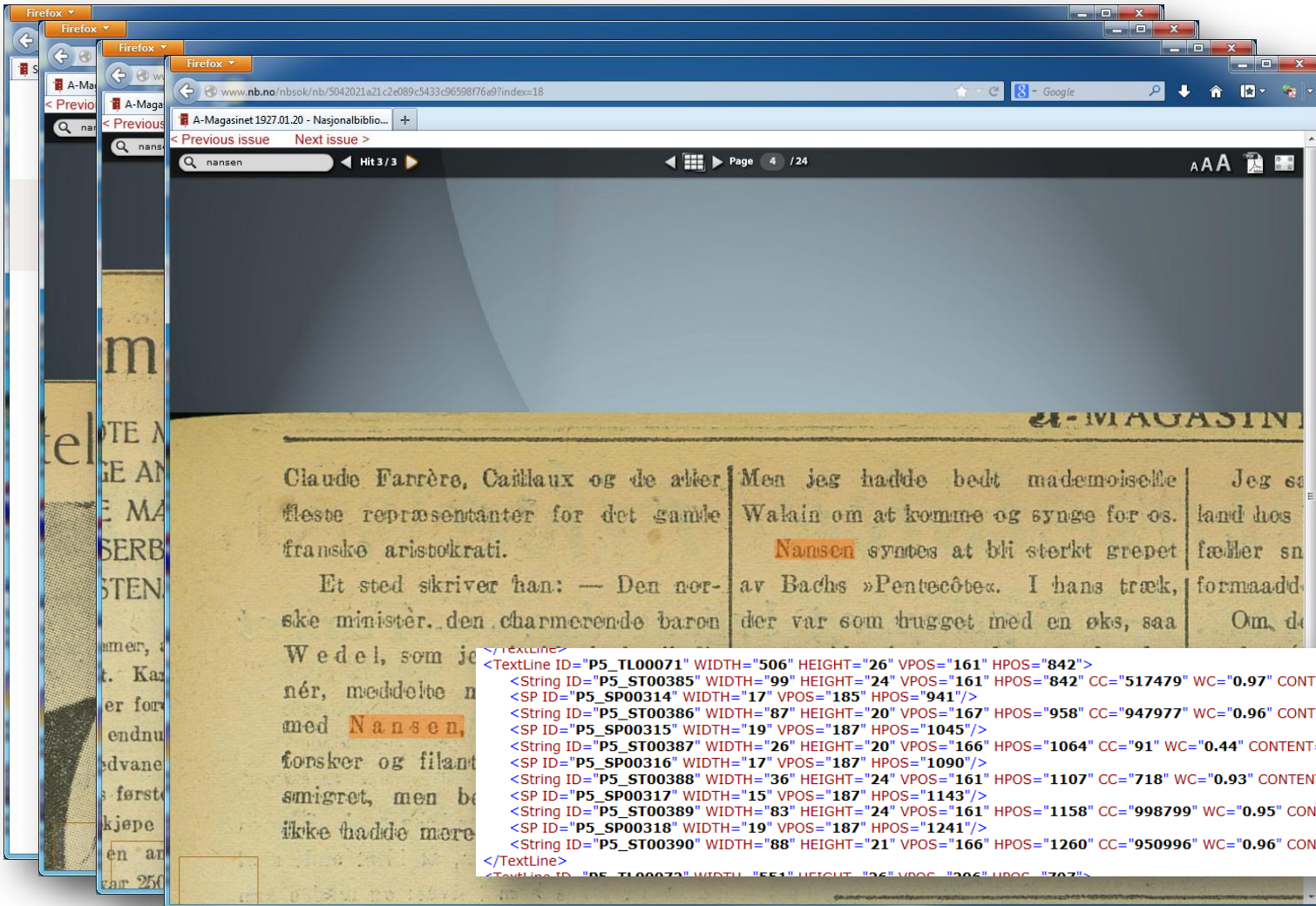
The search form includes the following fields and options:

- Search string:** A text input field.
- none of these words:** A text input field with a note: "Put exact words in quotes: 'Knut Hamsun'".
- any of these words:** A text input field with a note: "Put exact words in quotes: 'Knut Hamsun'".
- Title:** A text input field with a dropdown arrow.
- From date:** A date input field with a calendar icon.
- To date:** A date input field with a calendar icon.
- Filter results:** A dropdown menu currently set to "Digitally available".

At the bottom of the search form are two buttons: "Search" and "Cancel". A link "More about digital newspapers in the National Library" with the subtext "Liste over digitaliserte aviser" is located to the right of the "Title" field.



# Search



```
<TextLine ID="P5_TL00071" WIDTH="506" HEIGHT="26" VPOS="161" HPOS="842">
  <String ID="P5_ST00385" WIDTH="99" HEIGHT="24" VPOS="161" HPOS="842" CC="517479" WC="0.97" CONTENT="Nansen"/>
  <SP ID="P5_SP00314" WIDTH="17" VPOS="185" HPOS="941"/>
  <String ID="P5_ST00386" WIDTH="87" HEIGHT="20" VPOS="167" HPOS="958" CC="947977" WC="0.96" CONTENT="syntes"/>
  <SP ID="P5_SP00315" WIDTH="19" VPOS="187" HPOS="1045"/>
  <String ID="P5_ST00387" WIDTH="26" HEIGHT="20" VPOS="166" HPOS="1064" CC="91" WC="0.44" CONTENT="at"/>
  <SP ID="P5_SP00316" WIDTH="17" VPOS="187" HPOS="1090"/>
  <String ID="P5_ST00388" WIDTH="36" HEIGHT="24" VPOS="161" HPOS="1107" CC="718" WC="0.93" CONTENT="bli"/>
  <SP ID="P5_SP00317" WIDTH="15" VPOS="187" HPOS="1143"/>
  <String ID="P5_ST00389" WIDTH="83" HEIGHT="24" VPOS="161" HPOS="1158" CC="998799" WC="0.95" CONTENT="sterkt"/>
  <SP ID="P5_SP00318" WIDTH="19" VPOS="187" HPOS="1241"/>
  <String ID="P5_ST00390" WIDTH="88" HEIGHT="21" VPOS="166" HPOS="1260" CC="950996" WC="0.96" CONTENT="grepet"/>
</TextLine>
```

```
<TextLine ID="P5_TL00072" WIDTH="551" HEIGHT="26" VPOS="166" HPOS="707">
  <SP ID="P5_SP00287" WIDTH="20" VPOS="452" HPOS="320"/>
  <String ID="P5_ST00354" WIDTH="32" HEIGHT="18" VPOS="433" HPOS="340" CC="50" WC="0.72" CONTENT="og"/>
  <TextLine ID="P4_TL01257" WIDTH="144" HEIGHT="34" VPOS="344" HPOS="2199" CC="01004" WC="0.88" CONTENT="M.END."/>
  </TextLine>
  <TextLine ID="P4_TL00215" WIDTH="358" HEIGHT="34" VPOS="403" HPOS="2060">
    <String ID="P4_ST01258" WIDTH="279" HEIGHT="33" VPOS="404" HPOS="2060" CC="64055647760" WC="0.50" CONTENT="PARISERBREV"/>
    <SP ID="P4_SP01042" WIDTH="22" VPOS="437" HPOS="2339"/>
    <String ID="P4_ST01259" WIDTH="57" HEIGHT="33" VPOS="403" HPOS="2361" CC="360" WC="0.96" CONTENT="TIL"/>
  </TextLine>
  <TextLine ID="P4_TL00216" WIDTH="555" HEIGHT="37" VPOS="460" HPOS="1959">
    <String ID="P4_ST01260" WIDTH="319" HEIGHT="35" VPOS="462" HPOS="1959" CC="799705655706" WC="0.98" CONTENT="AFTENPOSTENS"/>
    <SP ID="P4_SP01043" WIDTH="34" VPOS="497" HPOS="2278"/>
    <String ID="P4_ST01261" WIDTH="202" HEIGHT="35" VPOS="460" HPOS="2312" CC="04366704" WC="0.98" CONTENT="MAGASIN."/>
  </TextLine>
</TextBlock>
```



# Norske Intelligenz-Seddeler.

No. I.

Onsdagen den 25 Maji 1763.

Da mange længe har forlangt og ønsket at see nyttige Efterretninger Norge betreffende, men man ogsaa tilligemed ville forsikre sig om en bestandig Forraad af endeel hertil fornødne Materier; saa forevises dette Stykke, som en Prøve, der viser hvortil samme kand tiene, og skal Fortsættelsen herefter udkomme hver Torsdag. Dog bygges dette Lovte og denne heele nyttige Indretning fornemmelig paa alle Patriots Godhed og Villighed, i at indsende til Bogtrykkeriet alt hvad dertil eragtes tjenligt.

Deconomiiske o  
Efterretning

Om Citroner at forbr  
over i Nor

Der ere vel mange Maa  
vare Citroner paa i Norge  
I et hvert Huus bruger er  
nu een, en anden forsøger  
den Maade.

Somme Stæder plejer  
at tage en Del reen hvid  
den vel, og med den at tild  
der skal giemmes, efterat  
Træk-Papir Stykke-viis  
i en godt glasseret Steen-  
Laug for Laug, hvorime  
Sand, efterat den er ble  
kommes, saa at Sander  
over de nedlagde Citroner

## So far ...

- About 17 % of the newspapers are scanned and available from **nb.no**
- 15 newspapers are deposited as print-quality (within 4-5 years, legal deposit established for all 254 Norwegian titles)
- Revision of the legal deposit law







Mona Løkås  
mona.lokas@nb.no

