

# AI in Libraries

Testcases from the National Library of Norway

Roger Jøsevold  
Deputy National Librarian  
National Library of Norway





- 220 staff
- Aquisition and cataloging
- Digitisation and digital library development
- Long term preservation

Arctic circle 66° 34'

Mo i Rana

- 210 staff
- Reading rooms
- Exhibitions
- Conferences and events

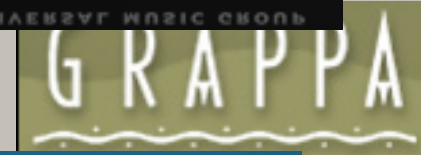
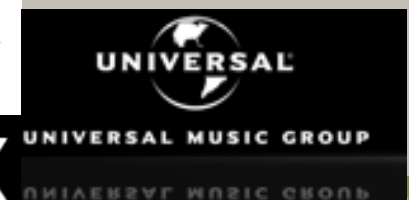
Oslo

# Legal deposit in Norway

- all published content from all media to The National Library

- Legal Deposit Act of 1989

- All printed material
- Music (sheet and sound)
- Broadcasting (radio and TV)
- Film and video
- Photographs
- Digital documents
- The .no domain



# Strategy for developing a Digital National Library for the 21. Century

- Developed and adopted 2004 to 2006
- To be relevant in the future we have to be on internett
- We will digitise the whole collection. Will take 15-30 years, depending on budgets.
- The digitising will be done in an industrial «conveyerbelt» manner, from A to Å without consideraton
- Exeptions will be made if material is needed for specific reasons



# Revised legal deposit act July 1 2017

- **Changes**

- Legal deposit of both the published format and the digital production-/print-files
  - metadata from production included
- Broad harvesting of the .no domain
  - Limited open access to the net-archive
  - Specific rules to privacy protection



# ...amendments July 1 2018

- **Access to legal deposit in digital format**
  - Four licences to be shared by University and College faculty and students
  - Two Licences to be shared by Public Libraries

Scale: - Total population 5 300 000  
- Faculty and students 300.000  
- Public libraries 650



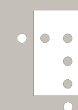
# Research institution and language bank

- Since 2010 The National Library have had the role of being The Norwegian Language Bank
  - Supply academia and the industry with relevant language resources
  - Developing text corpuses and tools for quantitative text analysis - Digital Humanities
  - Since this year: collaborate with «the Industry» to secure the use of Norwegian Languages in digital assistants and the like



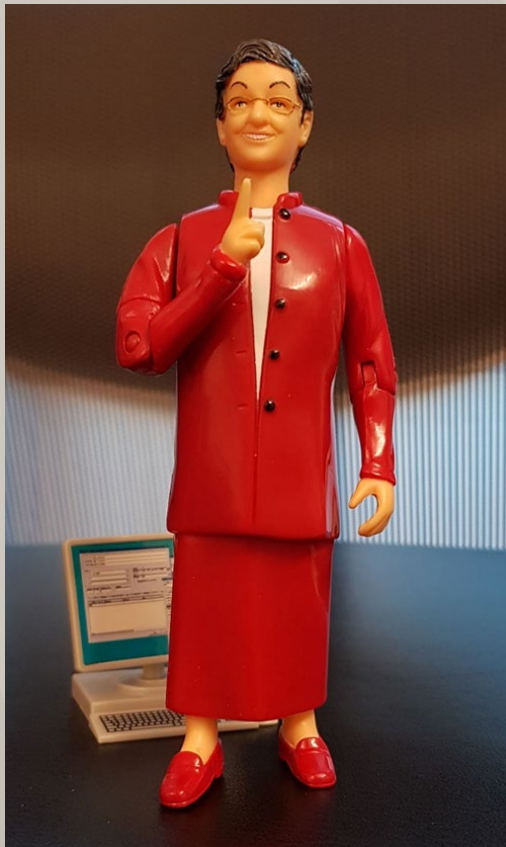
# The Collection

<b>Media</b>	<b>Analog Collection</b>	<b>Digital Collection</b>	<b>Years Digital coll.</b>
<b>Books (copies)</b>	3 200 000	550 000	1519-2019
<b>Newspaper (copies)</b>	5 600 000	2 154 000	1776-2019
<b>Journals</b>	4 815 000	28 000	
<b>Radio Broadcasting (hours)</b>	2 204 516	1 430 000	1930-2019
<b>TV Broadcasting (hours)</b>	1 269 000	993 000	1960-2019

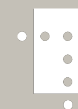




# «Nancy - the superlibrarian»



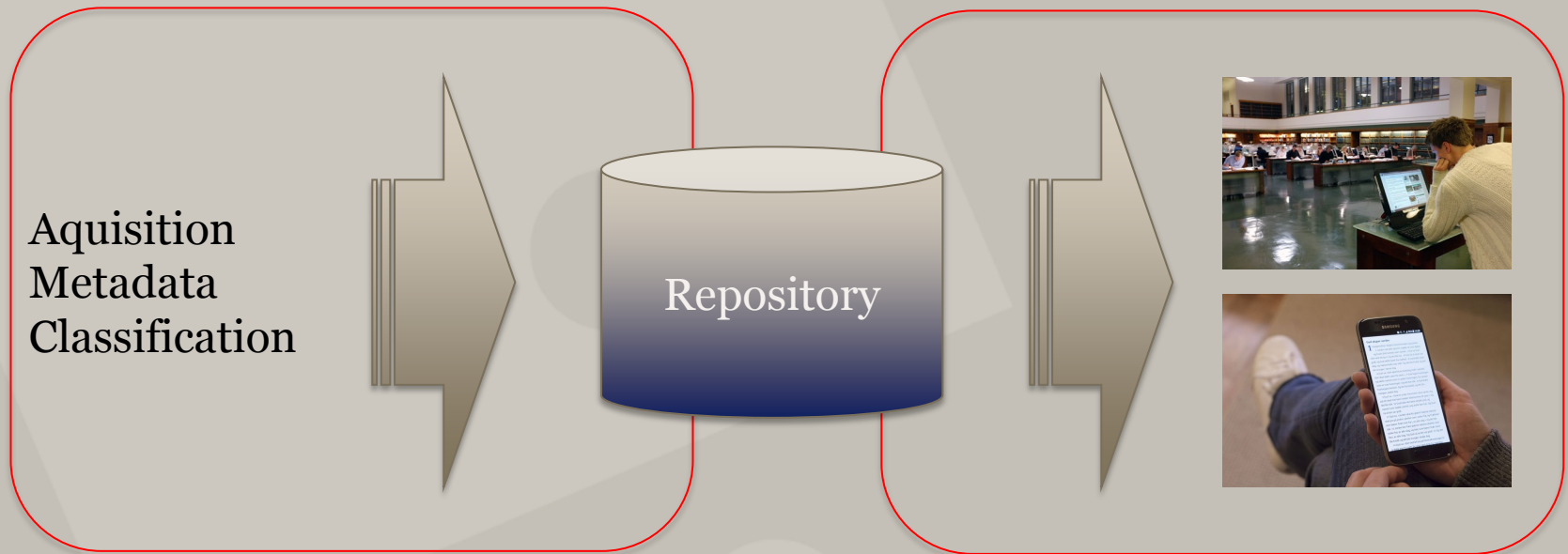
- Umbrella for AI testing at NLN
- Small scale projects and experiments with defined scope and running from weeks to months in volume
- Currently 3 full time staf+ some sourcing
- Main tool Google TensorFlow (open source software library)



# Two spheres for AI-testing

Production / ingest

Search / dissemination



# Production test 1: Separating groupes of litterature

- Fiction, school books, textbooks, academic literature
- Assign each books to only one group
- Result: accuracy 95%





# LIBRARY MAP

<http://bu.bl.ac.uk/info/ddc.html>

Mapped by [rwild@1-900-870-6235.com](mailto:rwild@1-900-870-6235.com)

## History of the world

- 920 Genealogy, names, insignia
- 940-2 the British Isles
- 950 Asia and the Far East
- 970 North America
- 990 other parts of the world
- 920 Biography
- 900 History: general resources
- 900 History: departments
- 909 World history
- 900 History: journals
- 910 Geography and travel
- 930 ancient world
- 960 Africa
- 940 Europe
- 980 South America
- 960 Spanish & Portuguese literature
- 860 French literature
- 800 German literature
- 800 Canadian literature
- 809 Poetry: general resources
- 808-899 Children's literature
- 808-899 Women writers
- 808-838 Science fiction & editorial techniques
- 808-88 Quotations
- 808-8 Literature: general collections
- 808-8 Authorship, writing & editorial techniques
- 808-83 Fiction
- 800 Literature: general resources

## 900 Geography & history

- 020 Library and information sciences
- 004-006 Computing and the Internet
- 003 Systems
- 002 The book
- 010 Bibliography
- 001 Knowledge, humanities & research
- 070 News media, journalism, publishing
- 030 General encyclopaedic works
- 080 General collections
- 060 General organisations & museology

## 800 Literature & rhetoric

- 800 Literature in other languages
- 870 Latin literature
- 850 Italian literature
- 820 English, Scottish & Irish literature
- 810 American literature in English
- 808 World-wide
- 708 in the US
- 708 in the UK
- 710 Chic and landscape art
- 720 Architecture
- 740 Drawing and decorative arts
- 760 Graphic arts: printmaking & prints
- 780 Music
- 790 Recreational & performing arts
- 770 Photography & photographs
- 750 Painting and paintings
- 730 Plastic arts: sculpture
- 709 History of art
- 701 philosophy
- 703 dictionaries & encyclopaedias
- 705 journals & magazines
- 706 organisations
- 707 art schools
- 640 Home economics & family living
- 660 Chemical engineering
- 670 Manufacture for specific uses

## Art galleries & museums

- 809 Literary study & criticism
- 808-83 Fiction
- 808 General rhetoric
- 800 Literature: general resources
- 700 The arts
- 700 The arts: general resources
- 700 artists
- 700 general resources
- 620 Engineering & allied operations
- 630 Agriculture & related technologies
- 650 Management & auxiliary services
- 660 Chemical engineering
- 670 Manufacture for specific uses

## 700 The arts

## Fine & decorative arts

## 600 Technology (applied sciences)

## 500 Natural sciences & mathematics

## 400 Language

## 300 Social sciences

## 200 Religion

## 100 Philosophy & psychology

## 000 Generalities

## Philosophy

- 100 societies
- 100 journals
- 100 departments
- 100 general resources
- 090 Manuscripts, rare books, other rare printed materials
- 070 News media, journalism, publishing
- 030 General encyclopaedic works
- 010 Bibliography
- 001 Knowledge, humanities & research
- 070 News media, journalism, publishing
- 030 General encyclopaedic works
- 010 Bibliography
- 001 Knowledge, humanities & research

- 120 Epistemology, causation, humankind
- 130 Paranormal phenomena
- 150 Psychology
- 180 Ancient, medieval, oriental philosophy
- 190 Modern Western philosophy
- 170 Ethics (moral philosophy)
- 160 Logic
- 140 Specific philosophical schools and viewpoints
- 110 Metaphysics
- 200 Religion: education & research
- 230 Christianity: general resources
- 230 Christianity: journals
- 236 Christian creeds & catechisms
- 242 Christian writings
- 268 Christian education
- 280 Christian denominations
- 294.4 Jainism
- 294.6 Sikhism
- 296 Judaism
- 299.5 14 Tadaim
- 297.93 Baha'i Faith
- 299 Other religions
- 299 Islam
- 299 Zoroastrianism
- 294.5 Hinduism
- 294.3 Buddhism
- 270 History of Christianity
- 253 Texts of sermons
- 246 Christian art
- 234.161 Baptism
- 220 Bible
- 210 Philosophy & theory of religion
- 200 Religion: general resources
- 300 Social sciences: general resources
- 310 Collections of general statistics
- 320 Political science
- 340 Law
- 360 Social problems & services: associations
- 380 Commerce, communications transportation
- 355 Military science
- 370 Education
- 390 Customs, etiquette, folklore

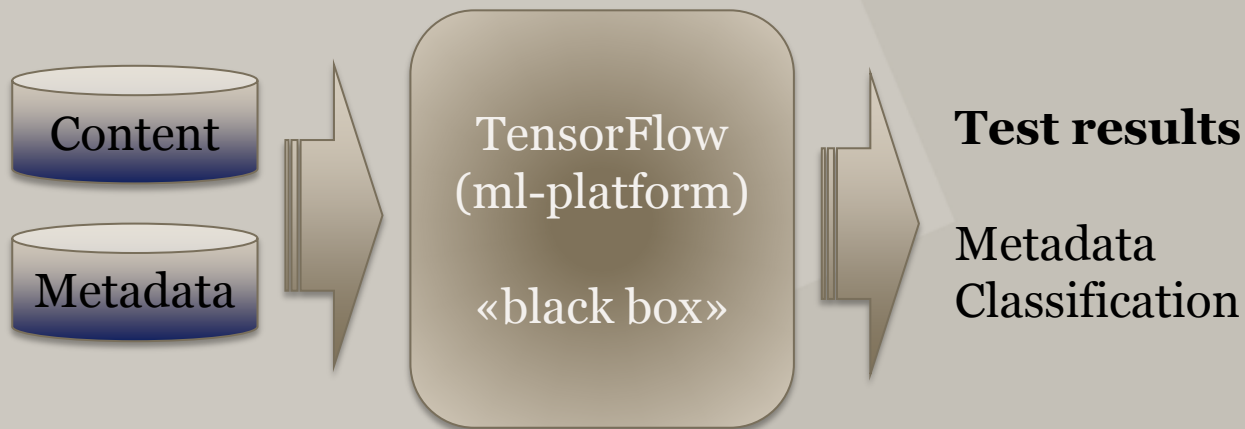


Contact Us

Help

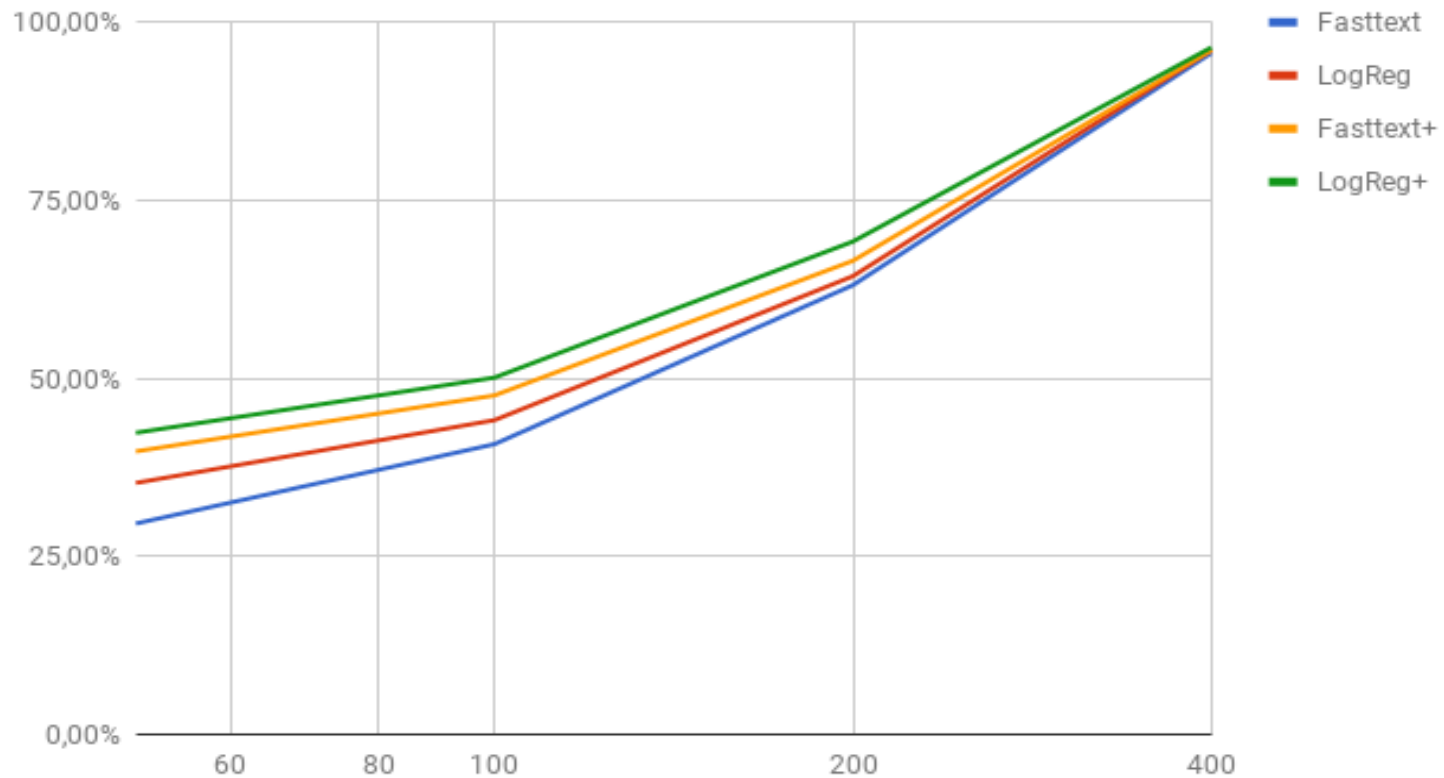
Feedback

# Production test 2: Dewey classification



# What have we experienced?

Precision vs size of training set



# Use of a pre-trained service

Google's ml platform «TensorFlow»

- Still Image analysis
- Speech to text conversion
- Natural language processing
- Video analysis



Labels

Web

Text

Document

Properties

Safe Search

JSON



trikk2.jpeg

Car 98%

Motor Vehicle 98%

Transport 96%

Vehicle 95%

Black And White 88%

Mode Of Transport 88%

Vintage Car 83%

Snapshot 82%

Automotive Design 78%





Labels

Web

Text

Document

Properties

Safe Search

JSON



trikk2.jpeg

+Page 1

+Block 1

TORGET

+Block 2

VAL HAGEBY

+Block 3

TORGET

+Block 4

ULLEVÅL HAGEBY



# A digital library-service based on machine learning only?

Content from legal deposit for January 2011

- **Newspapers:** (250 titles, 3.000 copies, 4.400.000 articles, 209.000 images)
- **Radio:** two national network channels (742 hrs audio)
- **TV:** one national channel (100 hrs video)



# Design principles

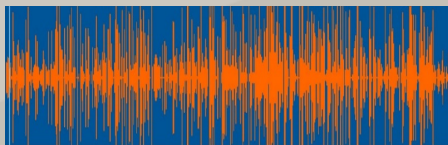
News-  
papers



TV



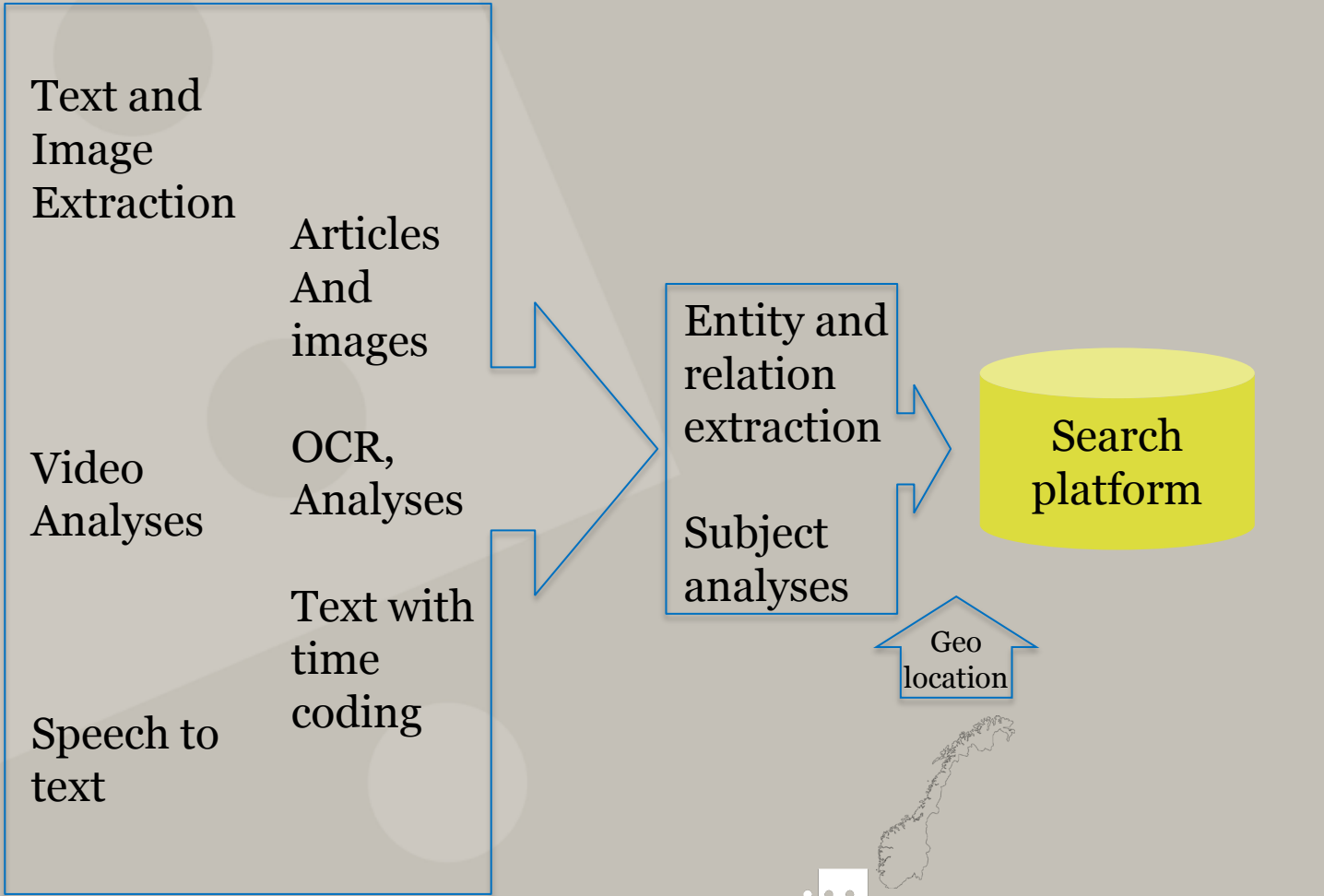
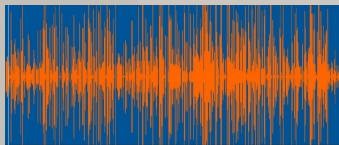
Radio



- Persons (names)
- Places (geography)
- Organizations
- Time (calender)
- Relations
- Subject classification

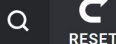


# Design principles





jens stoltenberg



LOCATIONS

- norge
- oslo
- sverige
- russland
- europa
- lofoten

ORGANIZATIONS

- ap
- sv
- stortinget
- ntb
- arbeiderpartiet
- høyre

PEOPLE

- maria amelie
- kristin halvorsen

maria amelie

70.27,-14.71 - 54.42,35.92

01.01.2011 - 31.01.2011

SHOW MAP

All Newspapers TV Radio

**jens stoltenberg + maria amelie**  
... was mentioned in agderposten avis on Friday 14th January 2011  
norge skjønnerlitteratur  
norsk kriminalitet sverige  
[Les avisen](#)

**jens stoltenberg + maria amelie**  
... was mentioned in agderposten avisbilde on Friday 14th January 2011  
norge skjønnerlitteratur  
norsk kriminalitet sverige  
[Les avisen](#)

**jens stoltenberg + maria amelie**  
... was mentioned in helgelandarbeiderbla avisbilde on Monday 17th January 2011  
norge norway refugees  
norsk innvandrere  
[Les avisen](#)

**jens stoltenberg + maria amelie**  
... was mentioned in helgelandarbeiderbla avis on Tuesday 18th January 2011  
politikk historie  
skjønnerlitteratur geografi  
juss  
[Les avisen](#)

**jens stoltenberg + maria amelie**  
... was mentioned in helgelandarbeiderbla avisbilde on Tuesday 18th January 2011  
politikk historie  
skjønnerlitteratur geografi  
juss  
[Les avisen](#)

**jens stoltenberg + maria amelie**  
... was mentioned in agderposten avis on Saturday 22nd January 2011  
skjønnerlitteratur norsk  
fortellinger ungdomslitteratur  
norske  
[Les avisen](#)

**jens stoltenberg + maria amelie**  
... was mentioned in adresseavisen avis on Saturday 22nd January 2011  
kultur politikk norge  
journalistikk fjernsyn  
[Les avisen](#)

**jens stoltenberg + maria amelie**  
... was mentioned in adresseavisen avisbilde on Saturday 22nd January 2011  
kultur politikk norge  
journalistikk fjernsyn  
[Les avisen](#)

**jens stoltenberg + maria amelie**  
... was mentioned in adresseavisen avis on Monday 24th January 2011  
nyhet norge  
statsforvaltning politikk  
offentlig  
[Les avisen](#)

# Considerations, what's next?

1. Content?
2. Legal issues?
3. National Library's role and mandate?
4. A new paradigm?
5. From test to production?



# Content?

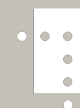
- Continue to digitise the whole National Library collection
- Extra funding to include digitising from The National Archive, libraries and museums in general.
- «Full text» digital legal deposit for all media by 2022
- Harvesting the .no domain
  
- A unique National digital Cultural Heritage repository





# Legal issues – copyright?

- Extended collective licencing?
  - The bookshelf ?
  - Flexible rights-clearing ?
- On-line access in accordance with The Legal Deposit Act
- Text mining – corpus building
- ?



# Legal issues - personal information protection (new legal deposit act)?

- The right to comment, conceal or delete personal information that is made available by mistake, or.....

if published by a site without a responsible editor those concerned may demand the information to be concealed or deleted

- A request for deletion that is disputed by The National Library may be appealed to a separate board



# National Library?

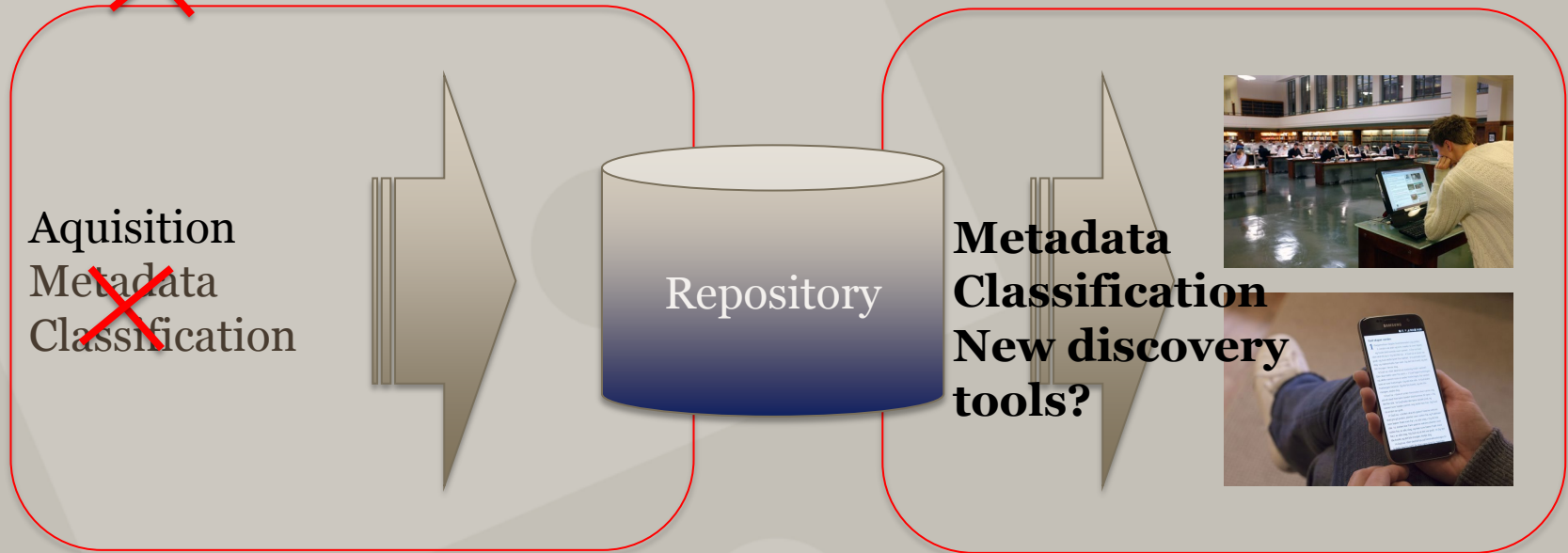
- Virtual national cultural heritage universe?  
(digital cultural heritage «one stop shopping»)
- Institutional upheaval and restructuring?  
(National Library? - National Archives? –  
National Museums?)



# From test to production?

~~Production / ingest~~

Search / dissemination

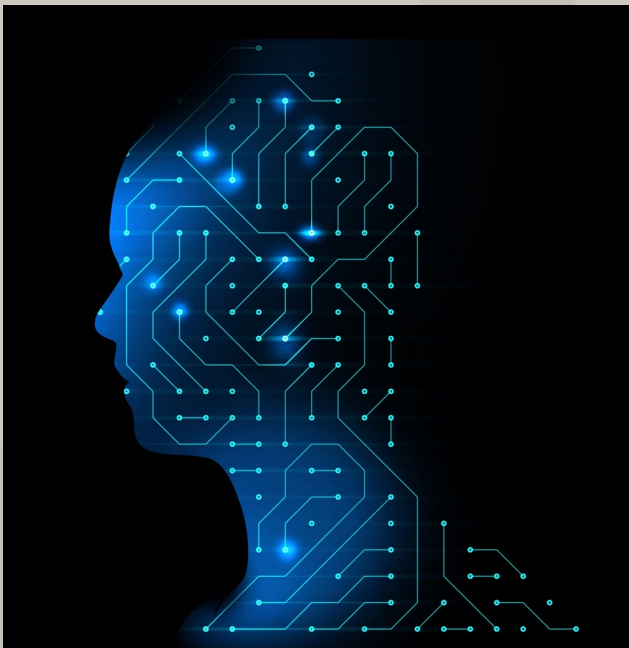


# From test to production

Collection	Analog Collection	Metadata	Digital Collection
Books (copies)	3 200 000	very rich	550 000
Journals	4 815 000	rich	35 000
Newspaper (copies)	5 600 000	limited	2 154 000
Radio Broadcasting (hours)	2 204 516	limited	1 430 000
TV Broadcasting (hours)	1 269 000	limited	993 000
Photography (numbers)	80 millions (?)	limited/non existing	1 530 000
Ephemeras (items)	2 700 000	archival/very limited	0
Private collections (pages)	>16 000 000	limited/non existing	740 000



# Conclusion I



- Keep testing
- Apply in beta nb.no?
- Colaboration with Stanford University Library
- Fantastic Futures conferences and work-shops
- Desember 2018 in Oslo
- Desember 2019 at Stanford



# Conclusion II

Maneuvering through the “Bermuda Triangle” called decision making

*Acuracy / good-enough?*



Do we need to?

Do we dear to?

