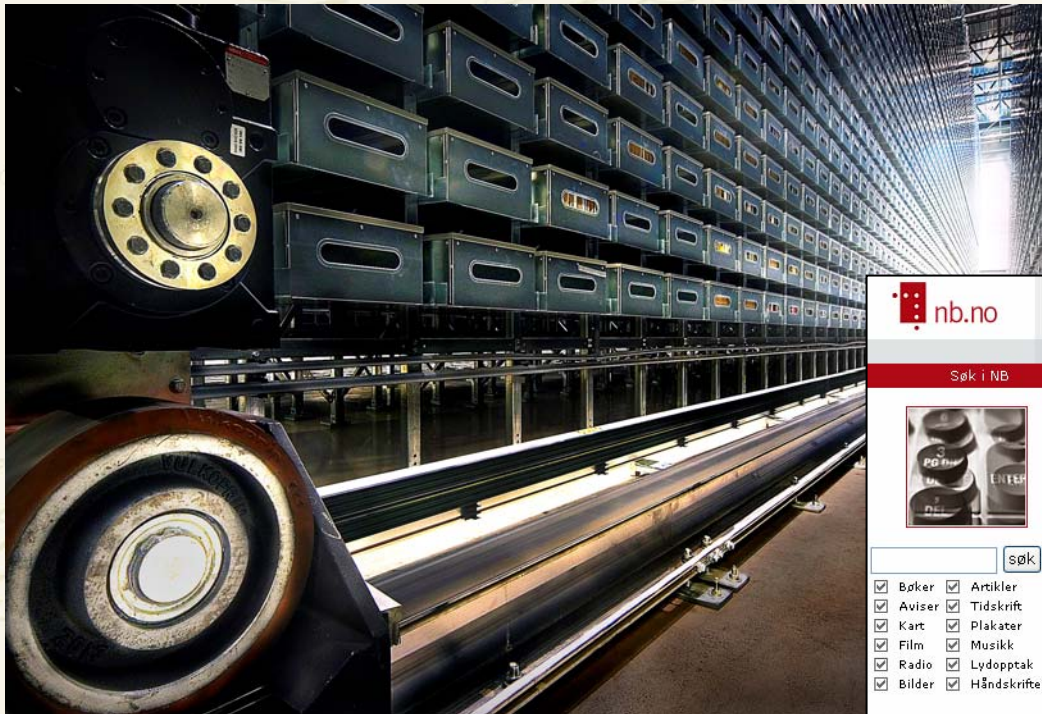


The Norwegian Digital Radio Archive - 8 years later, what happened?

Svein Arne Brygfjeld, National Library of Norway



Large-scale audio digitization



nb.no Nasjonalbiblioteket

Om Nasjonalbiblioteket Kontakt Logg inn English

Søk i NB Opplevelser Fag Aktuelt

- Bøker
- Aviser
- Kart
- Film
- Radio
- Bilder
- Artikler
- Tidsskrift
- Plakater
- Musikk
- Lyddopptak
- Håndskrifter

Søketips

Gå dit ...

Gå dit ...

Flere nyheter ...

Månedens gjenstand::
Sabotørspillet fra 1946
Sabotørspillet utkom høsten 1946. Det ble lansert som en spennende og morsom julegave, og kanskje brukte mange norske barn romjulsdagene dette året på å sette seg inn i brettspillets regler og roller.
Les mer ...

Kalender ::
lørdag 19 nov 2005
Mens vi venter på julen!
Les mer ...

torsdag 24 nov 2005
Liker De Brahms?
Les mer ...

Nasjonalbiblioteket - Postboks 2674 Solli, 0203 Oslo
telefon: 810 013 00 - send e-post
Copyright Nasjonalbiblioteket © 2005

Besøksadressen:
Oslo - Drammensveien 42
Mo i Rana - Finsetveien 2

Background

- The partner institutions

The Corp

- Pub

<http://net>

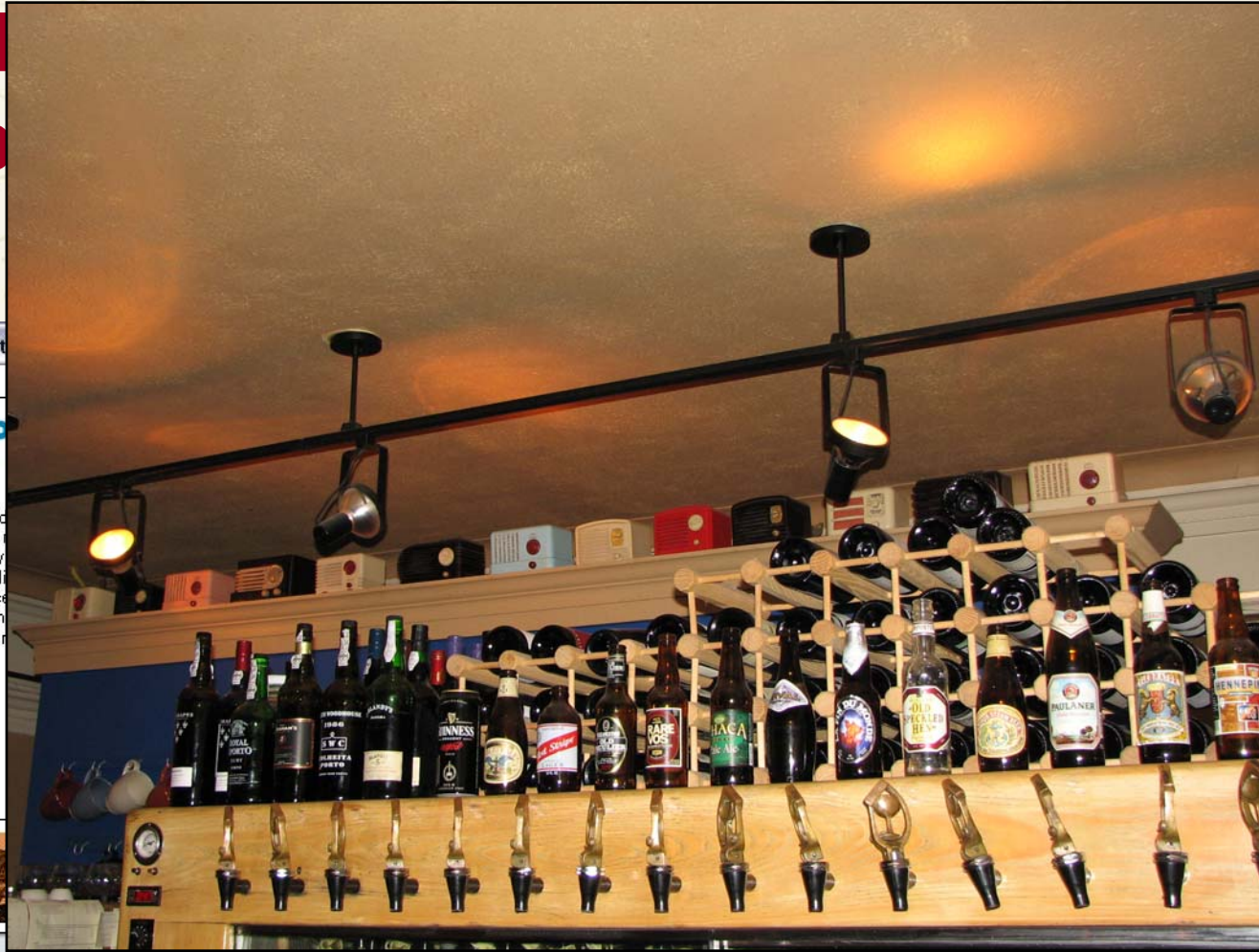
NRK P

NRK P1

NRK P1 - god
for over 1,4
hver dag. Ny
distriktssend
sport, service
underholdnin
voksen pop-

NRK P2
SANN
ER LIVET

Fullført





• Nat

IPRES@Cornell October 2006

Background cont.

- The project

Shared interest

They wanted

- re-use of archival recordings in their production
- To reduce the need for physical storage space
- To preserve the audio recordings, tapes were deteriorating
- To prepare for the digital domain
- To save money

We

- Are the Norwegian Memory
- Want to give long-term access to the recordings for a wide variety of users

Both

- Public institutions

The original archive

- Estimated to be approx 50.000 hrs recordings on $\frac{1}{4}$ inch analog tape
- Extensive daily
- Good, but not w level
- 5-10 archivists



1998

Starting point

- A vision
- Open mirror
- A pilot im

Goals

- Digitize th
- (>50.000
- Internet-k
- Permane

RLG DigiNews: Volume 2, Number 6, December 15, 1998 - Mozilla Firefox

http://www.rlg.org/legacy/preserv/diginevs/diginevs2-6.html

December 15, 1998, Volume 2, Number 6, ISSN 1093-5371

Table of Contents

- Feature Article
 - Digital Archiving: Approaches for Statistical Files, Moving Images, and Audio Recordings
 - [Introduction](#) by Oya Y. Rieger
 - [Archiving Statistical Data: The Data Archive at the University of Essex](#) by Simon Musgrave and Bridget Winstanley
 - [Universal Preservation Format \(UPF\): Conceptual Framework](#) by Thom Shepard
 - [Norwegian Digital Radio Archive Initiative](#) by Svein Arne Brygfield and Svein Arne Solbakk
- Technical Feature
 - [Selecting a Digital Camera: the Cornell Museum Online Project](#) by Peter B. Hirtle and Carol DeNatale
- [Highlighted Web Site](#)
- [Calendar of Events](#)
- [Announcements](#)
- [FAQs](#)
- [RLG News](#), including "Digital Preservation Needs and Requirements in RLG Member Institutions, Executive Summary" by Margaret Hedstrom and Sheon Montgomery
- [Hotlinks Included in This Issue](#)

Feature Article

Digital Archiving: Approaches for Statistical Files, Moving Images, and Audio Recordings

Introduction
Oya Y. Rieger, Co-Editor, *RLG DigiNews*
oyr1@cornell.edu

One of the features that makes digital archiving an overwhelming effort is the richness of the digital terrain. What we simply refer to as digital information is indeed a deep and complex combination of materials, including text, images, sound, video, and numeric and spatial files. The reason we group this wide range of materials under the umbrella of digital information is because they are all composed of "bits and bytes." As articulated in a recent British Library Research and Innovation Centre report, the type of material to be preserved is one of the key factors governing the choice of a preservation approach. (1) Although there is significant overlap in the issues faced with all digital materials, there are also problems unique to each format. For example, a simple ASCII text file is in a standard format with limited hardware and software dependencies. On the other hand, a numeric file that includes statistical data in a predetermined format, including macros for calculations, is much more challenging as a result of several interdependencies in its technical environment.

The October 1998 issue of *RLG DigiNews* included an article by Margaret Hedstrom reviewing national initiatives in digital preservation. This month, we offer readers a comparative digital archiving view from a different perspective, by type of digital material. We invited representatives from three digital archives, each with a different digital material type (statistical files, moving images, and audio), to describe the preservation issues they are addressing. We also asked them to articulate the attributes of these formats that represent unique challenges to the long-term preservation of their collections.

Fulltext

2006

A running archive
Trust
Cooperation
Internet base

Surprises!

The image shows a screenshot of a Mozilla Firefox browser window. The address bar displays the URL `http://www.nb.no/sesam/index.php?query=*8search_kat=1&mediatype%5BRadio%5B`. The page content includes the logo for `nb.no` and the text "Søk i Nasjonalbiblioteket". A search bar contains the text "Søk i NB" and a search button labeled "Søk". Below the search bar, there is a section for "Du er logget inn som: Sam All Sesamsen" and a "Søketips" field. A search result is displayed: "[STORTINGSPRESIDENT HAMBRO I ANLEDNING 25.09.1940] - Tale av Hambro von".

Overlaid on the browser window is a radio player interface. The address bar of the player shows `http://radioarkivet.nb.no - |||| radioarkiv...`. The player has a progress bar and a volume control. The "Clip Info" field contains "Historisk radiomateriale NRK". The player shows a bitrate of "40,0 Kbps", the format "mp3", and a duration of "01:23,4/09:51,0". There are buttons for "[0:00]" and "0:00]", and a "Lukk vindu" button. The player is labeled "Fullført" at the bottom.

Basic principles

- Digitization is supposed to be done once only
 - High quality (48 KHz, 16-bit stereo, no compression)
 - Standard (Broadcast Wave Format)
 - Original tapes preserved by the library
- Off-the-shelf technology
 - As little in-house development as possible
 - Open-source where applicable
 - General technology
- Everything except selection/priorities done by the library

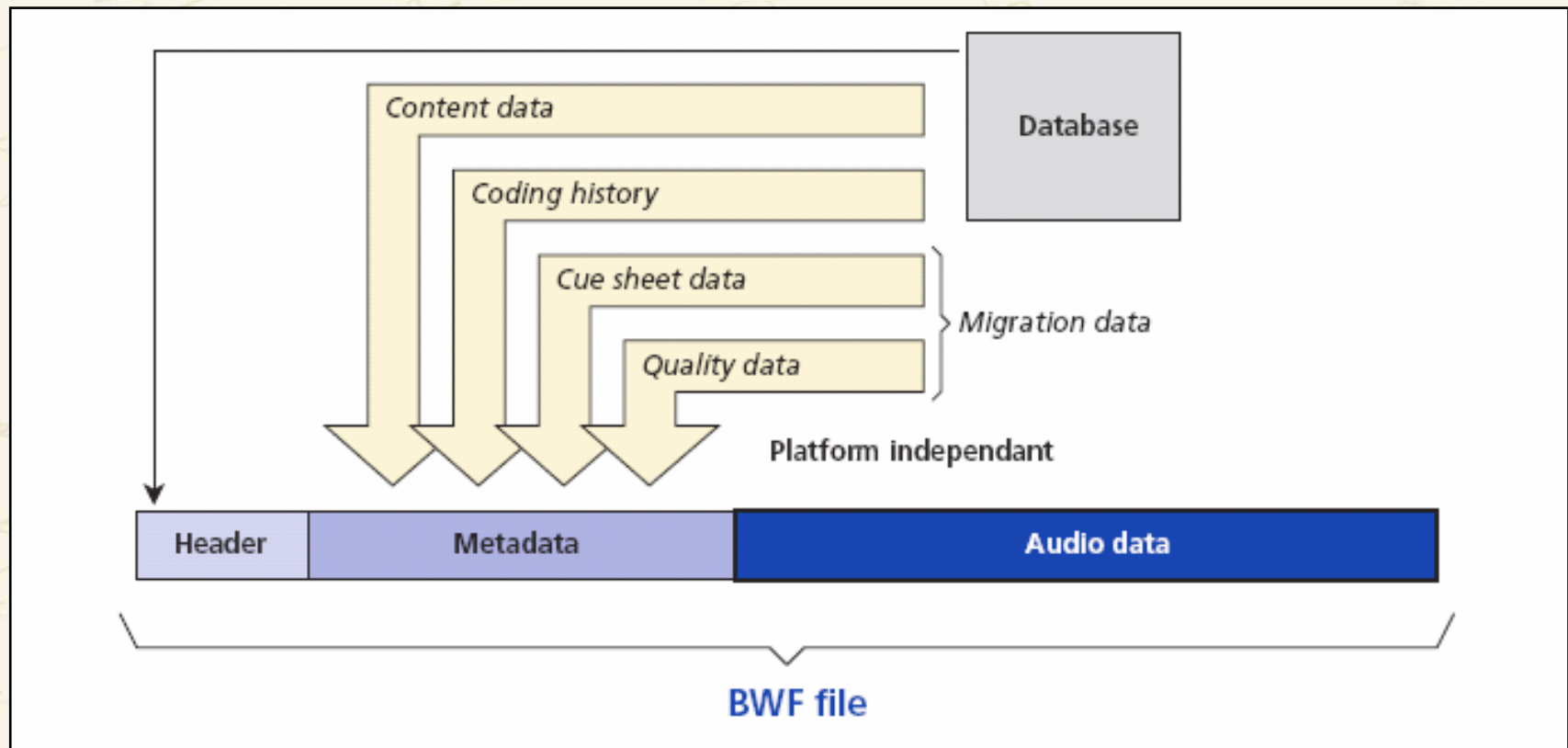
Technology (1998, remember 😊)

- As many ¼ inch tape players as we could get
- Three Unix work stations tuned to handle three (four) audio streams continuously, three tape players each
- Professional external high-quality A/D converters
- 1 TB RAID disk
- Some in-house developed software
- Repository solution made in-house
- Search based on web/oracle
- Delivery based on Real streaming and ftp

Technology (now)

- Digitization infrastructure unchanged
- Repository solution developed further for general use
- General infrastructure improved
- 1 TB RAID disk is now 500 TB
- Consolidation on Linux operating system
- Search based on general search engine

Audio formats



Access and use, Public and Research

- Limited access
 - Copyrights unclear in some cases
 - Limited amount open to the public
 - Everything open for research
 - Everything open in our buildings
- Research
 - Role Based Access Control
 - Researchers can log in using username/password from their institution
 - Daily use, different perspectives

Effects

- There was more
 - The tape archive contains more than estimated
 - The tape archive grew during the first years
- Trust
 - High level of trust between the partners
 - Door opener
- Good services
 - Popular service for the professionals, significant increase in use of archival material
 - Some parts available for all on the Internet
 - The complete archive available within the library and on the Internet or researchers

Effects cont.: Surprising

- Much of the archive was lost because of extensive re-use of magnetic tape
 - Employees rescued recordings by hiding tapes, building hidden/secret archives (drawers, home...)
 - (Many) these show up now
 - Significant amount
 - Find-a-tape campaign
- New archive-based radio channel
 - Active role for the library as well defining relevant content

Today: current audio input

- Digitization: 10.000 hrs/year
 - Unknown recordings still show up
 - Archive is larger than estimated
- Migration from digital tape: 60.000 hrs/year
 - QIC, tape robot/library
- Automated legal deposit: 35.000 hrs/year
 - 4 radio channels
 - Fully automated, includes metadata provided by producer

Needed now

- Audio pattern recognition
 - Support search for certain sounds and voices
- Audio to text conversion
 - Support content search and navigation

Lessons learned

- Massive use of off-the-shelf components works
- Pay attention to those steps done only once
 - Reading original, A/D-conversion
- Tuning of workflow, processing and logistics takes time
- Good practice establishes trust and trust is needed
- Long term use and re-use is a better argument than preservation
- We have learned to walk, *and now we start running:*

We will digitize our collection in 15 yrs

Digital now	Total	Type
210 000	4 700 000	newspapers
205 000	1 300 000	still images
100	450 000	books
570	250 000	hrs moving images
1 000	4 000 000	manuscripts
45	180 000	maps
4 000	80 000	hrs music
7 000	60 000	posters
80 000	1 000 000	hrs radio
100	850 000	journals
1 000	1 900 000	small prints

Thank you for listening

svein.arne.brygfjeld@nb.no

....information at your fingertips - always