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

Svein Arne Brygfjeld, National Library of Norway
Large-scale audio digitization
Background

- The partner institutions
The Norwegian Broadcasting Corporation

- Public service broadcasting
- Established in 1933, among the first in Europe (BBC in 1922)
- Extensive archives from the beginning until today
- Four radio channels and two TV channels now
- Everything also distributed on the Internet
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 ¼ inch analog tape
- Extensive daily use
- Good, but not well-structured metadata on program/cut level
- 5-10 archivists
1998

Starting point
- A vision
- Open minds
- A pilot implementation

Goals
- Digitize the complete historical radio recordings (>50,000 hrs)
- Internet-based service
- Permanent cooperation
2006

A running archive

Trust

Cooperation

Internet based services

Surprises!
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 workstations 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

- High quality
- Was: Linear BWF, mostly 48/16/2, some higher
- Is: Same

- Access/use
- Was: RealAudio 64 Kb/s, MPEG1 layer 2 384 Kb/s, Linear BWF
- Is: Various MPEG1 Layer 3 (MP3), MPEG1 layer 2 384 Kb/s, Linear BWF, Developing
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

<table>
<thead>
<tr>
<th>Type</th>
<th>Digital now</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>newspapers</td>
<td>210 000</td>
<td>4 700 000</td>
</tr>
<tr>
<td>still images</td>
<td>205 000</td>
<td>1 300 000</td>
</tr>
<tr>
<td>books</td>
<td>100</td>
<td>450 000</td>
</tr>
<tr>
<td>hrs moving images</td>
<td>570</td>
<td>250 000</td>
</tr>
<tr>
<td>manuscripts</td>
<td>1 000</td>
<td>4 000 000</td>
</tr>
<tr>
<td>maps</td>
<td>45</td>
<td>180 000</td>
</tr>
<tr>
<td>hrs music</td>
<td>4 000</td>
<td>80 000</td>
</tr>
<tr>
<td>posters</td>
<td>7 000</td>
<td>60 000</td>
</tr>
<tr>
<td>hrs radio</td>
<td>80 000</td>
<td>1 000 000</td>
</tr>
<tr>
<td>journals</td>
<td>100</td>
<td>850 000</td>
</tr>
<tr>
<td>small prints</td>
<td>1 000</td>
<td>1 900 000</td>
</tr>
</tbody>
</table>
Thank you for listening

svein.arne.brygfjeld@nb.no

….information at your fingertips - always