



# Mass digitisation perspective from YLE

Pekka Gronow  
Finnish Broadcasting Company  
(YLE)  
Unlocking Audio Conference  
26-27 Oct 1007

How long will it take to digitise 150,000 audio tapes and 40,000 DAT tapes? (That's just part of our collection)



It will take 3 men 10 years to do the job, working in shifts.  
6 men could do it in 5 years, but 30 men can't do it in a year



One man could do it in 30 years, but for many sound carriers, we cannot wait that long.

---



# The problem

---



- 250 000 **deteriorating** analog tapes
- 60 000 **unstable** digital carriers (DAT, CDR, MD)
- 100+ new **"born-digital"** sound files to be archived every day
- 1500 customers who want everything now

# The YLE archive in figures, 1

---



- 250 000 analog full-track tapes (19, 38, 76 cm/s)
- 200 000 vinyl discs (45, 33)
- 20 000 shellac discs (78)
- 10 000 analog tape cassettes
- 5 000 lacquer discs

## The YLE archive in figures, 2

---



- 100 000 CD:s (red book standard)
- 20 000 CD-R:s (orange book standard)
- 40 000 DAT tapes
- and cylinders, minidisks, multi-track analog and digital tapes, etc etc ...

# Content, 1

---



- the **archival collection**
  - **15 343** selected programs from 1935-1970, detailed documentation
  - the **music collection: 15 000+** music recordings owned by YLE (Radio Symphony orchestra etc, 1950-2007), detailed documentation

# Content, 2

---



- the sound effects collection
  - 80,000 sound effects on analog tape, with a card catalogue

# Content, 3

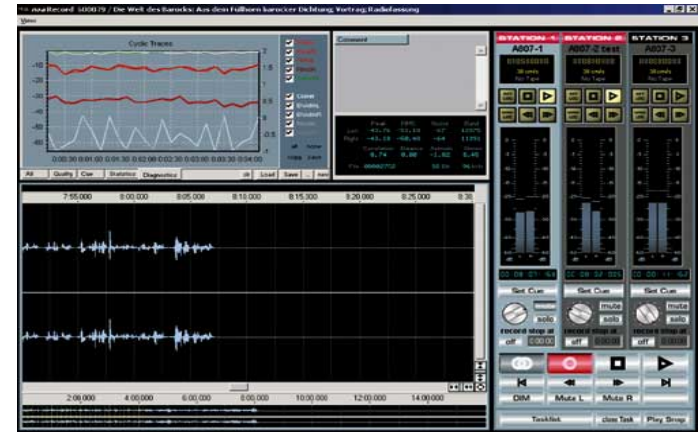
---



- the "working collection"
  - unselected material with minimum documentation

<b>1950 - 1954</b>	<b>218</b>
<b>1955 - 1959</b>	<b>2 666</b>
<b>1960 - 1964</b>	<b>3 802</b>
<b>1965 - 1970</b>	<b>15 329</b>
<b>1971 - 2007</b>	<b>200 000 +</b>

# Radio archive digitisation plan (1)



- the first stage, (1999- ) 2003-2005 (already finished)
- start with the oldest / most used / best documented tapes
  - digitise 15 343 tapes in **archival collection** with NOA
  - digitise 15 000 tapes in **music collection** with Quadriga
  - digitise 6 468 oldest tapes in **working collection (1950 - 1964)** with NOA
  - digitise and document 10 000+ sound effects
  - **CDs and 78s** digitised **selectively**

# Original plan published in 1999

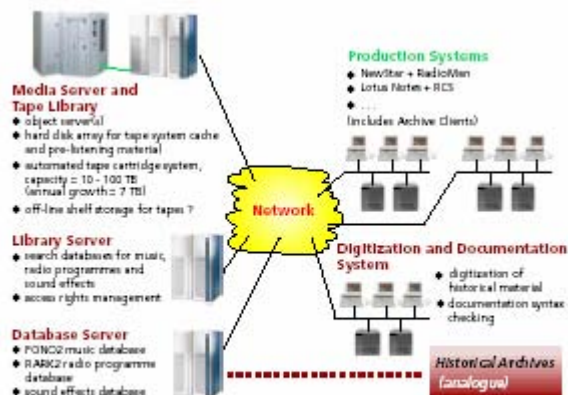


Figure 2  
System architecture.

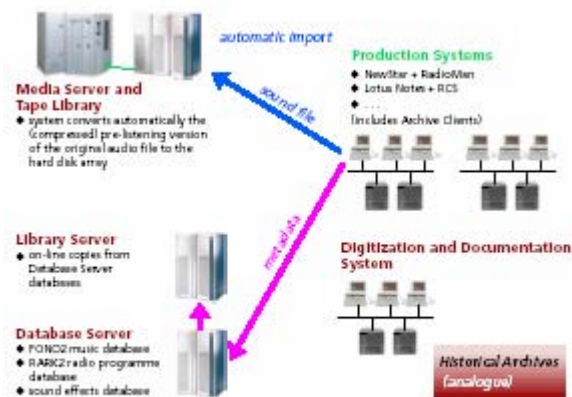


Figure 3  
Archiving from production systems.

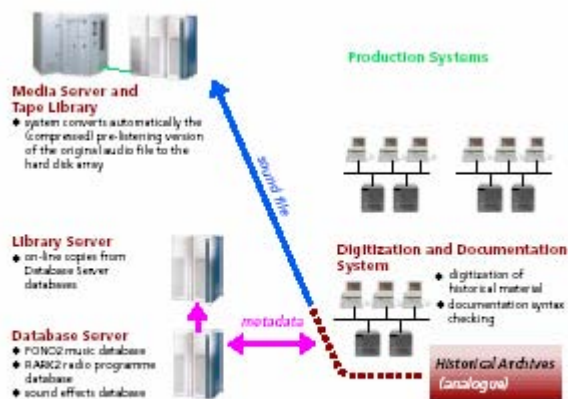


Figure 4  
Digitization of historical materials, and metadata checking.

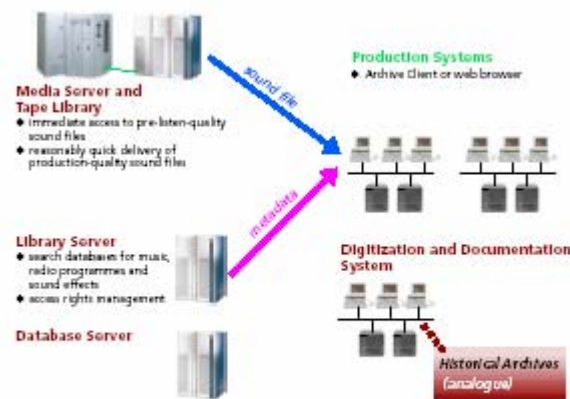


Figure 5  
Retrieval.

EBU Technical Review - Summer 1999

# A technical view...

———— audio files BWF = Broadcast Wave Format

..... metadata XML = Extended Markup Language

## Essence Management System

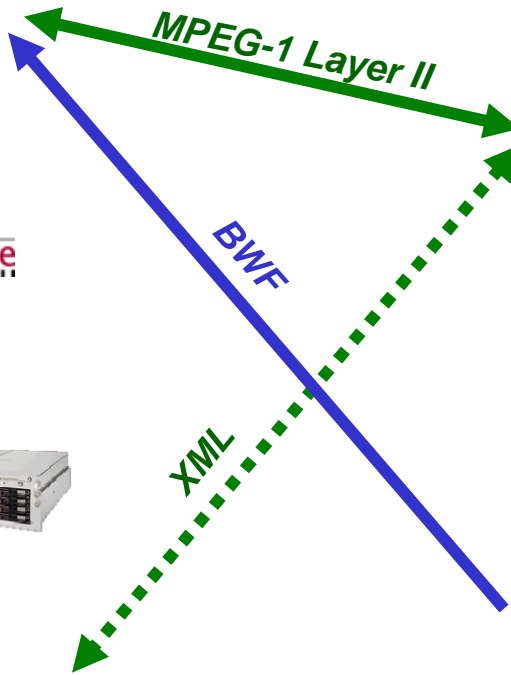
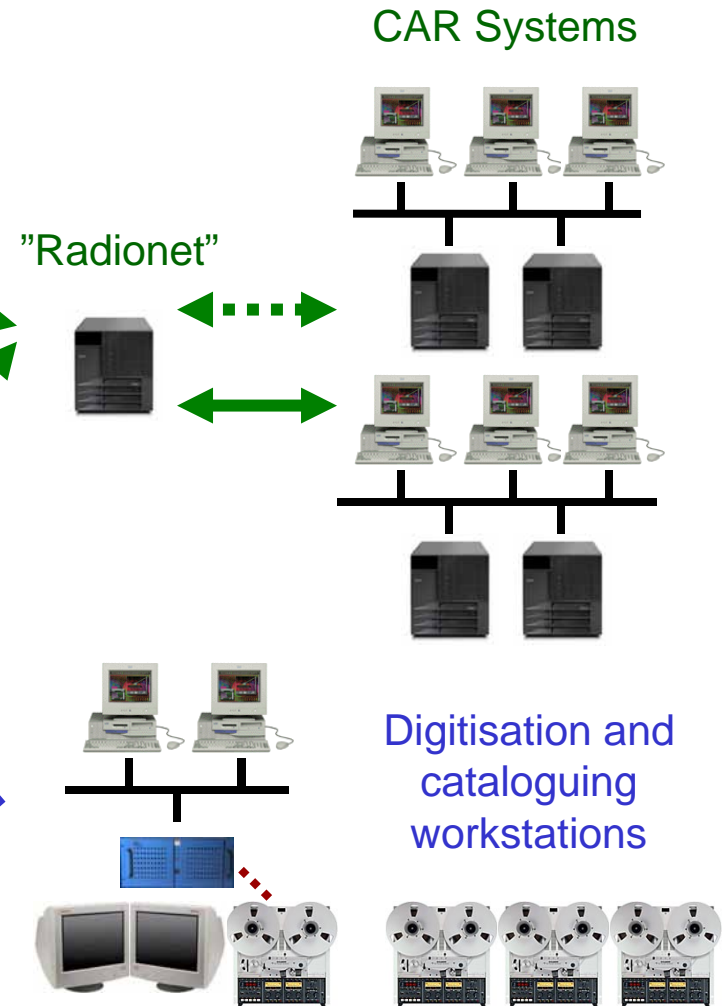
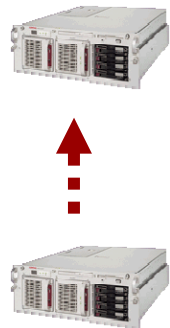


## Information Retrieval Tools



## Metadata Management

ORACLE®



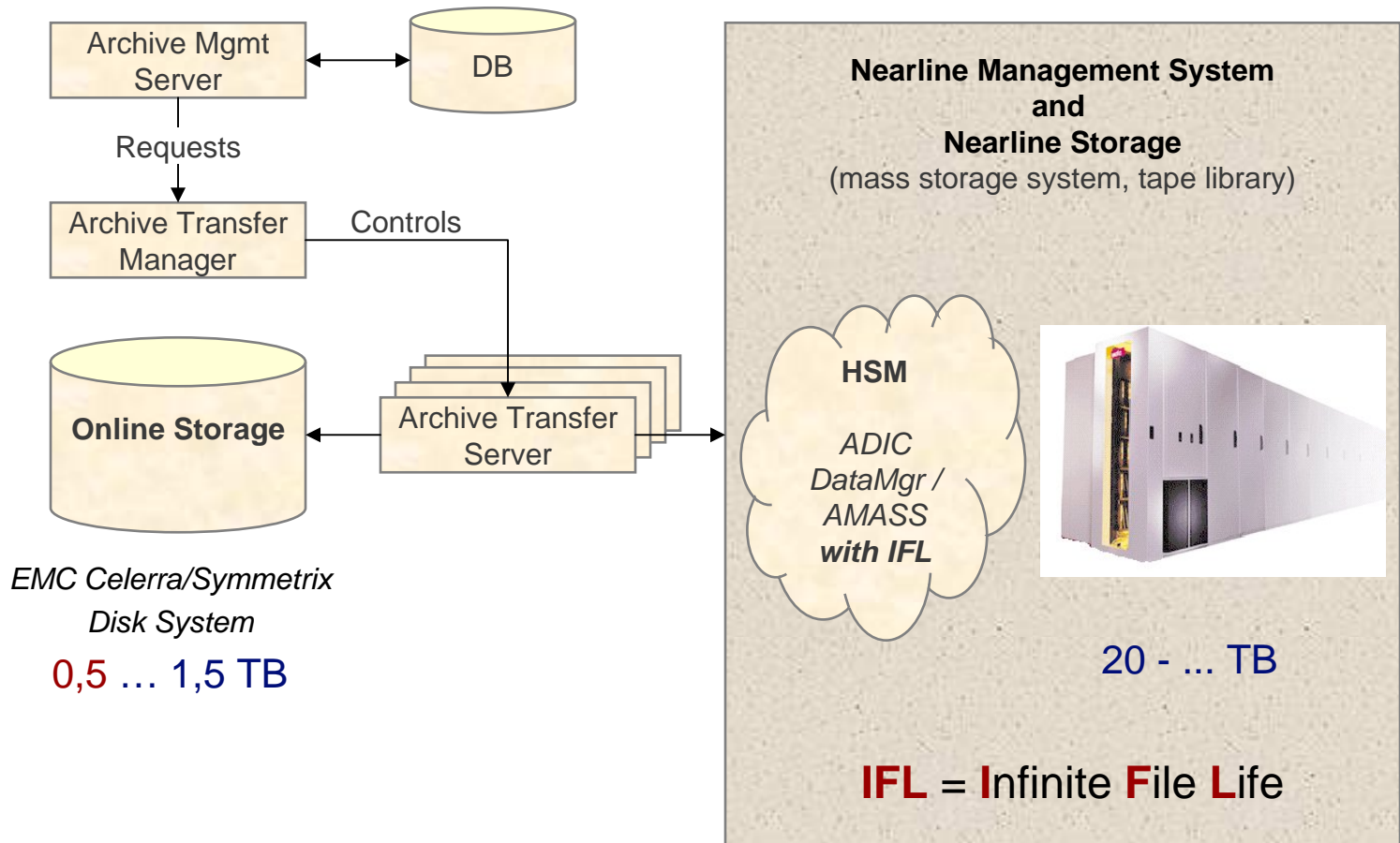
# Essence management, storage

---



- Archive Transfer Server
  - transfer between Online Storage and Nearline Storage
- Online Storage
  - caching the most recently stored or retrieved audio files
  - **all browse copies are kept on-line for fast access**
  - Nearline Management System
  - an HSM system **with media quality monitoring**
- Nearline Storage
  - **long term storage** of audio files in a tape library

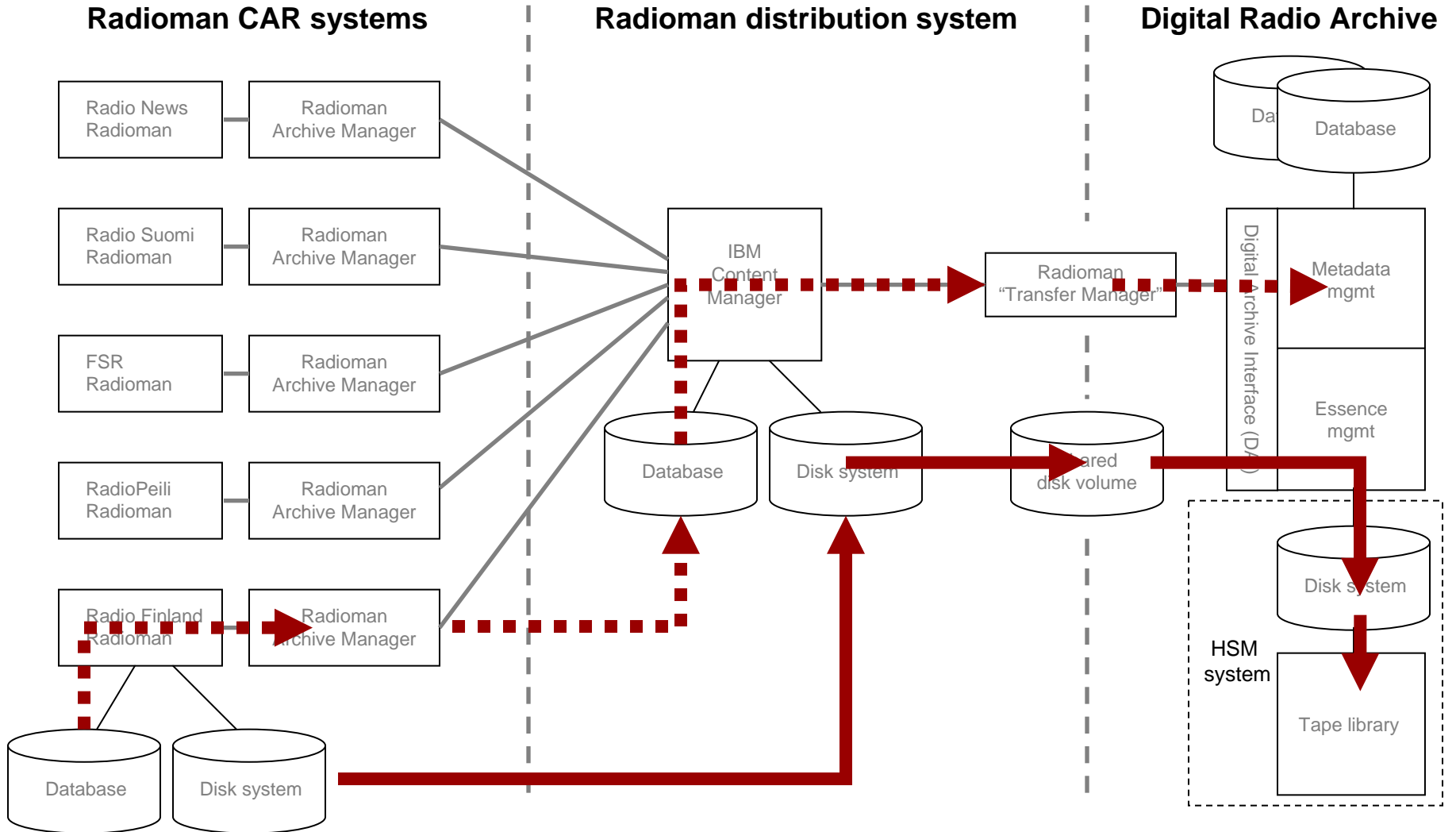
# Long term storage of audio files



Original diagram: Peter Thomas, Blue Order

# Archiving "born-digital" material from CAR systems (computer-aided radio production systems)

**———— audio files**  
**..... metadata**



# The user's viewpoint

---



- access through search and retrieval system
- user can search digital and/or analog materials
- browsing-quality sound streamed to all users
- production-quality files take longer
- format conversion on demand

# A look at the user interface

---

The users can access the digital archive through YLE's Intranet system and find the desired items in the catalogue

KART: CD-03545-06

KEST: 003:53

OTSI: YELLOW SUBMARINE

TEKI: Lennon, John [1940-1980] (säv, san). McCartney, Paul [1942- ] (säv, san).

TEOS: [All you need](#) is love

DLAA: Ura 06 [Esikuuntelu](#) [Laatuääni](#)

ESIT: [Beatles](#), The (yhtye).

KAUP: Parlophone / CDP 7464452

## Radio archive digitisation plan (2)

---



- the second stage
  - c. 2006 – 2015
  - digitise all DAT tapes by 2008
  - digitise the rest of the collection, beginning with radio drama, documentaries, sound effects ...

## We now have five years' practical experience with digital archiving. **First the good news...**

---

- IT CAN BE DONE. THERE IS NO ALTERNATIVE.
- Mass digitisation of sound recordings can be done on a large scale at a reasonable cost (but if you are still relying on a card catalogue, you are in big trouble)
- Metadata on new, born-digital radio programs can be harvested automatically
- Once the system is up and running, the demand for archival materials will increase
- In the long term, there will be considerable savings

## Then the bad news...

---

- Murphy's law is still valid in IT. Many things will go wrong. "Faith is good, control is better"
- The time window for some media, such as DAT, is closing sooner than expected
- Integration of radio production and archiving systems is still not yet fully completed
- Most of the savings we had hoped for were lost because of system down-time. Our mass storage system proved to be the wrong choice. A new system is now being installed, and we cross our fingers...

# What next?

---

- We now have a digital sound archive
- The pros and cons of being early
- From a preservation viewpoint, no major problems. One third of the job is done.
- Access problems will be solved with new mass storage system
- By 2015 the entire radio archive will be on-line; what shall we do with it?

# New services


---

- "The Living Archive", a web site for presenting historical radio and television programs
- The online shop, a site for purchasing sound files of radio programs (opening early 2008)
- Professional services for AV producers, such as a site selling sound effects

# Research

---

- cooperation with the national library (cataloguing of new CDs, digitisation of 78s)
- international cooperation with the DISMARC project: a European portal for sound archives (more info at [www.dismarc.org](http://www.dismarc.org))
- Finland will open a national radio and television archive in 2008, which could become a gateway to the YLE archives for the research community



Thank you for your interest  
contacts:  
[pekka.gronow@yle.fi](mailto:pekka.gronow@yle.fi)