Experiences from the YLE Radio archive audio digitization project

Pekka Salosaari – Yle Radio Archives, BAAC Annual Conference Oct 4 2012 Helsinki
Outline

• Yle & the archives
• Clientele and environment
• Digitization, a strategic choice
• Workflows in the archival process
• Metro, Yle’s MAM system
• Lessons and future plans
2012

- Budget 440 million euros
- 3100 employees
- 4 + 1 TV channels, 6 radio channels, yle.fi and other Internet services
- 20 regional radio services in Finnish, 5 in Swedish and 1 in Sámi
- 8 regional TV news broadcasts
- 98% of Finns use Yle’s services yearly, more than 70% daily
- Radio production > 200 000 h /year
Who are we for?

Radio archives customers and users:

- Yle’s Radio and TV channels, journalists, programme production, news desk

- Independent production companies, joint productions

- Yle’s Elävä Arkisto
  ’Living Archive’ web-service
  yle.fi/elavaarkisto/

- Yle Sales -> private and commercial use

- Research community

= The Finns

Co-operation:
KAVA
IASA
DISMARC
Audio archives in Finland
Years in audio archiving

1926 Yleisradio - The Finnish Broadcasting Company
1934 first recording equipment purchased
1939 tape recorders
1955 Tape depository
1961 Audio archive
1987 online catalog
1989 digital production (DAT)
2001 digital archive system
2010 Integrated MAM-system Metro
The Radio archive collections:

radio programmes, news broadcast, music recordings, radio plays, coverage, sports…

5000 78 rpm shellac records (1935-1950)

250 000 ¼” analog open reel tapes (1950-2005)

50 000 DAT’s

35 000 CD-R’s

2000 MDI’s

80 000 sound effects on 20 000 analog reel tapes

1000 Multi-track 2” analog and DASH tapes

Paper documents
Digitization - strategic issues

Why digitize?

Archival paradigm: Preservation, access and usage, born-digital content

The choice between keeping everything and selecting something:

- no selection: may be cost effective at first, how about for years to come?
- selection: where are the decisions made, based on what assumptions?

Priorisation: where to start and what next?

- obsolete carriers in danger

- content value, production costs

- availability of playback technology

- on-demand service for programme production needs

- changing business environment
The ingest process

- Selection and retrieval of carriers
- Metadata creation and addition (applies for some material)
- Carrier logistics and preparations for playback
- Audio capture in mass digitization unit
- 24 bit/48 KHz BWF Metro import and final check
- Carrier storage after digitization
- Metadata addition
- Restauration (on demand or systematic)
- Scanning the paper documents and catalogues in the radio archive?
Tools of the trade

2 NOA units for mass digitization and ingest:

Unit 1:
Record: 4 Studer A807s
Medialector: 8 Sony PCM-R500s / Tascam MDCD1

Unit 2:
Record: 3 Studer A807s

In test: NOA CDLector

Sonic SoundBlade HD for audio restoration

WaveLab workstations for editing, copying and lightweight restoration
TV - interfacing with planning, production and delivery systems

- **MGX WHATS'On Scheduling**
- **Avid Interplay**
- **Subtitle Editor**
- **Avid iNEWS**
- **Pebble Beach Automation**
- **Cavena Subtitling Playout**
- **Omneon Playout Servers**
- **FP D.I/V Archive hi-res mgmt**

**D-Center**
- Live transmissions
- File-based Material
- Tape ingest

**External users**
- Press Preview Portal
- Translator Portal

**Programme Metadata**

**METRO**
- **MXF**
- **MPEG1 Subtitle files**
- **Rundown Metadata**
- **Subtitle files**

**Flash 8**

**Control (e.g. API)**
Number of users
- currently > 4 100 registered users
- ~ 3 000 users logged in at least once
- normally 150 - 200 concurrent users

Number of objects
- Video Objects ~ 600 000
- Audio Objects ~ 1 600 000
- Image Objects ~ 375 000

Amount of material 2 * 1,5 PB (peta bytes)
- hires video: 2 x 1,4 PB, browse video: ~25 TB
- hires audio: 2 x 100 TB
- 1,5 - 3,0 TB new material/working day - 0,5 TB/day in weekends
- ~ 50 TB new material each month

Number of processed workflows
- ~ 6 500 import/export/subworkflows processed daily

Status as 28.5.2012
Video objects
- programme 125 000 metadata of 477 000 programmes
- clip 61 000 raw material etc
- web clip 15 000 YLE Living Archive clips
- trailer 9 000 metadata of 15 000 trailers
- footage 43 000 EBU, Reuters and inhouse material

Audio Objects
- audio 1 200 000 radio programmes and audio clips
- music 335 000 Record Library’s music recordings
- sound effect 40 000 inhouse sound effects
- amount 200 000 h audio material
- metadata 3 000 000 entries in Radio catalog database

Image Objects
- image 330 000 metadata of 375 000 images
Lessons learned, plans for future

- hang on to the original carriers and related metadata
- before migration is over, prepare for the next
- double check everything after updates, both system and organization
- archivists think about the future

To do-list:

- large scale CDR-ingest, data model support
- multi-track recordings, both on tapes and hard disks
- offline metadata in tape binders, cataloging cards, publications etc.
- new cataloging & import system
Thank You! Questions?

Your Story