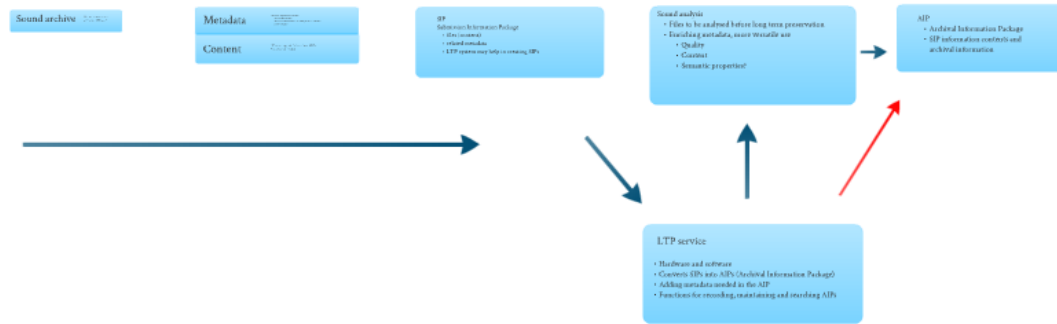


## Process steps



## Metadata

- Metadata now available (LTP minimum requirements):
- Descriptive metadata
  - Administrative metadata, including technical metadata
  - Rights metadata

LTP final report: "The more exhaustive metadata that SIPs include, the more versatile will be the use of preserved digital information in the future."

## Technical metadata

Technical characteristics of a file

1. Basic information, embedded in sound file
    - File type, size, length, channels, bit depth, resolution...
  2. Information embedded in BWF
    - Broadcast audio extension block
    - Description, Coding history
    - Quality block
    - Quality data
    - Cue sheet
- Problems:
- Some organisations produce technical metadata, not all
  - Not all files have LTP
  - Old Applications not always related to metadata
  - Where can we access this metadata?

If audio files could talk, what would they tell us?

## LTP project

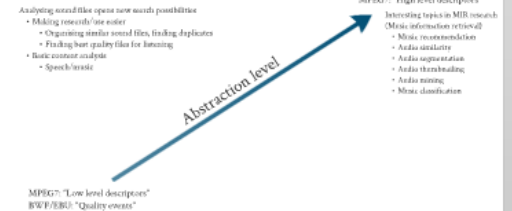
- LTP project
- NDL (National Digital Library)
  - One search portal
  - Access to material via organisations own portals
  - <https://www.kodi.fi/en/>
  - LTP is a sub-project of the (NDL) project
  - LTP-project website: <http://www.kodi.fi/en/long-term-preservation>
  - PDF: End report of the long-term preservation section (2010), main source of this presentation.
  - LTP in use earliest in 2016, probably later
  - Reliable long-term preservation of digital materials in Finland

Cooperation in long-term preservation

## Preparing sound files for long term preservation. A draft for a file receiving process.

Juha Korvenpää  
Centre for Digitisation and Preservation  
The National Library of Finland

## Sound analysis



LTP final report: "Limiting the amount of metadata produced by different organisations will increase both usability and preservation of the information in LTP."

## Finnish archives, museums, libraries and data preservation

- Sound files created and preserved in many Finnish sound archives (and museums and libraries)
- Obligation to preserve materials in digital format for a long period of time.
  - Currently each preserving their own digital collections.
  - Are organisations able to preserve access to digital materials in the future?



## Conclusion

- LTP-process
- a possibility to enrich audio related metadata
  - all files analysed controlled way
    - harmonising metadata
  - audio analysis, at least: quality aspects
  - Commercial products for sound quality analysis already on the market
- Co-operation
- Long term preservation experts
  - Digital library metadata professionals (METS)
  - Digital signal processing, technical universities
  - MIR research

Comments, feedback:  
[juha.korvenpaa@helsinki.fi](mailto:juha.korvenpaa@helsinki.fi)

## METS

- METS - Metadata Encoding and Transmission Standard
- "Metadata container"
  - Structural standard for SIP and AIPs
  - Generic Metadata profile is documented in LTP-project
  - Audio-METS
- Standards for audio materials for long term preservation?
- Audio-METS version 2.0 implementation
  - Audio Engineering Society metadata standards for audio materials AES33-2011 and AES36-2011
  - AES33/2011 co-operation

- METS includes:
- METS header
  - Descriptive Metadata
  - Administrative Metadata
  - File Section
  - Structural Map
  - Structural Links
  - Relations

- Problems
- Audio specific METS still to be defined, even for basic setting
  - More metadata may do the job/required
  - Cooperation needed to develop audio-METS
  - Can this metadata be accessed from NDL-portal?

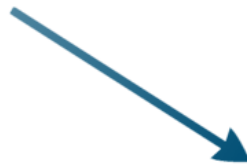
**Preparing sound files for  
long term preservation.  
A draft for a file  
receiving process.**

Juha Korvenpää  
Centre for Digitisation and Preservation  
The National Library of Finland

## Finnish archives, museums, libraries and data preservation

Sound files created and preserved in many Finnish sound archives (and museums and libraries)

- Obligation to preserve materials in digital format for a long period of time.
- Currently each preserving their own digital collections.
- Are organisations able to preserve access to digital materials in the future?



Centralised Long term preservation system

- cost reductions for organisations?
- more secure preservation of data?

# LTP project

## LTP project

- NDL (National Digital Library)
  - One search portal
  - Access to material via organisations own portals
  - <http://www.kdk.fi/en/>
- LTP is a sub-project of the (NDL) project
- LTP-project website: <http://www.kdk.fi/en/long-term-preservation>
  - PDF: End report of the long-term preservation section (2010), main source of this presentation.
- LTP in use earliest in 2016, probably later
- Reliable long-term preservation of digital materials in Finland

Cooperation in long-term preservation

# Sound archive

Contract between sound  
archive and LTP system

# Metadata

Minimum requirement metadata

- Descriptive metadata
- Administrative metadata, including technical metadata
- Rights metadata

# Content

LTP system required file format (wav, BWF)

Files converted if needed

## SIP

### Submission Information Package

- files (content)
- related metadata
- LTP system may help in creating SIPs

# LTP service

- Hardware and software
- Converts SIPs into AIPs (Archival Information Package)
- Adding metadata needed in the AIP
- Functions for recording, maintaining and searching AIPs



## AIP

- Archival Information Package
- SIP information contents and archival information

# Metadata

Metadata now available (LTP minimum requirements):

- Descriptive metadata
- Administrative metadata, including technical metadata
- Rights metadata

LTP final report: "The more exhaustive metadata that SIPs include, the more versatile will be the use of preserved digital information in the future."

# Technical metadata

Technical characteristics of a file

1. Basic information, extracted from sound files

- File type, size, length, channels, bit depth, resolution...

2. Information embedded in BWF

- Broadcast audio extension chunk
  - Description, Coding history
- Quality chunk
  - Quality data
  - Cue sheet

3. Automatic quality control in some digitisation systems

- System analysing incoming signal
- Signal to noise ratio, azimuth, sound level, peaks...
- Data written into BWF (EBU Tech 3285 Supplement 2)

Export to XML file

Problems

- Some organisations produce technical metadata, not all
- Not mandatory for LTP
- Old digitisations without any related metadata?
- How to get access to this metadata?

If audio files could talk, what would they tell us?

Sound analysis

# Sound analysis

Analysing sound files opens new search possibilities

- Making research/use easier
  - Organising similar sound files, finding duplicates
  - Finding best quality files for listening
- Basic content analysis
  - Speech/music

Abstraction level



MPEG7: "Low level descriptors"  
BWF/EBU: "Quality events"

MPEG7: "High level descriptors"

- Interesting topics in MIR research  
(Music information retrieval)
- Music recommendation
  - Audio similarity
  - Audio segmentation
  - Audio thumbnailing
  - Audio mining
  - Music classification

LTP final report: "Unifying the semantics of metadata produced by different organisations will increase both usability and preservation of the information in LTP."

ME

- '
- S

# METS

## METS - Metadata Encoding and Transmission Standard

- "Metadata container"
- Structural standard for SIPs and AIPs
- Generic Mets-profile is documented in LTP-project
- Audio-Mets?

## Standards for audio materials for long term preservation?

- AudioMD version 2.0 specification
- Audio Engineering Society metadata standards for audio materials: AES57-2011 and AES60-2011
- AES/EBU co-operation?

## METS sections

- METS Header
- Descriptive Metadata
- Administrative Metadata
- File Section
- Structural Map
- Structural Links
- Behavior

## Problems

- Audio specific METS still to be defined, even for basic setting
- More metadata may do this complicated
- Co-operation needed to develop audio-METS
- Can this metadata be accessed from NDL-portal?

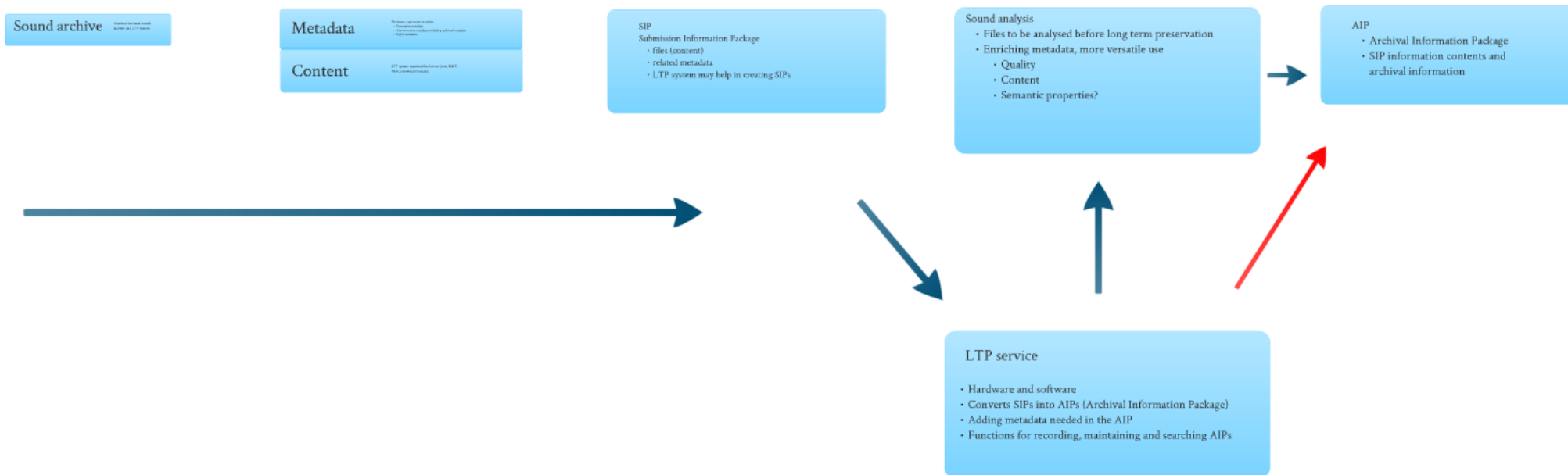
## Sound analysis

- Files to be analysed before long term preservation
- Enriching metadata, more versatile use
  - Quality
  - Content
  - Semantic properties?





## Process steps



# Conclusion

## LTP-process

- a possibility to enrich audio related metadata
- all files analysed controlled way
  - harmonising metadata
- audio analysis, at least quality aspects
  - Commercial products for sound quality analysis already on the market

## Co-operation

- Long term preservation experts
- Digital library metadata professionals (METS)
- Digital signal processing, technical universities
- MIR research

Comments, feedback:  
[juha.korvenpaa@helsinki.fi](mailto:juha.korvenpaa@helsinki.fi)