

Digitalizace zvukových dokumentů, dlouhodobá archivace digitálních zvukových dokumentů

Natálie Ostráková

Národní knihovna ČR



Digitalizace zvukových dokumentů

•Digitalizace

- Záchrana obsahu degradujících nosičů
- Zpřístupnění teď i pro další generace

•Aktuálně

- gramofonové desky a fonografické válečky

•Výhledově

- magnetické pásky
- magnetofonové kazety
- DAT kazety
- CD

Digitalizace zvukových dokumentů

• Digitalizační linka

- Přehrávač
- Předzesilovač
- Zvuková karta, A/D převodník
- Počítač a odpovídající software



• Pro digitalizaci fonováleček

- Archéophone5 - využívá moderní přenosky
- Endpoint – přehrává fonoválečky bezkontaktně pomocí laseru (České museum hudby, Indiana University)

Digitalizace desek

•Pro digitalizaci desek

• Vývoj gramofonů dodnes

System optického čtení záznamů Irene8, může skenovat i rozbité disky



(zdroj: <https://hal.archives-ouvertes.fr/hal-01885324/document>)

Digitalizace zvukových dokumentů

• Digitalizace zvukových nosičů

• Parametry digitalizace

- 24bitová hloubka, vzorkovací frekvence 96 kHz
- hodnoty vysoko nad hranicí lidského sluchu, umožňují zachytit informace, které sice člověk slyšet nemůže, ale zvukový dokument je přesto obsahuje
- Pro jiné nosiče a jiný zvukový obsah se bitová hloubka a vzorkovací frekvence může lišit (např. mluvené slove, nosiče DAT), liší se i praxe institucí

•

• Digitalizace doprovodných materiálů

- Deska, fonováleček
- Obaly
- Booklety
- Formát JPEG2000

Digitalizace nosiče a doprovodných materiálů



(zdroj: tinfoil.com)



(zdroj: <https://www.hippopx.com/en/his-masters-voice-shellac-shellac-disc-78rpm-album-cover-gramophone-plate-label-135080>)

Digitalizace a dlouhodobá archivace

- **Digitalizací ochrana obsahu nekončí**

- Digitální dokumenty jsou též předmětem ochrany
- Digitální dokumenty jsou též ohroženy ztrátou obsahu

- **Jak zajistit přístup k obsahu zvukových dokumentů i za desítky let?**

Digitální archivace

- Long term preservation
- Je třeba již produkci digitálních dokumentů řídit s ohledem na jejich trvalou udržitelnost
- Opatření= standardizace
 - Umožňuje perfektní znalost dat
 - Umožňuje efektivní správu dat, vyhledávání, řazení, provádění ochranných opatření (migrace)
 - Dává pokyny pro vytvoření dat v podobě, která je dle současného stavu vědomostí archivovatelná

Standard NDK

- Metadata
 - Souborové formáty
 - Pravidla popisu
 - Identifikátory
-
- Účel: Standardizace produkce digitálních dokumentů
 - Vychází z mezinárodních standardů a doporučení

Metadata

- Cíl: zachytit co nejvíce existujících informací
- Popisná metadata**
 - MODS, DC, volitelně katalogizační záznam v MARCXML
 - U desky se popisuje: deska nebo skupina desek, jedna strana desky (jedna nahrávka), jednotlivé zvukové stopy a doprovodné obrazové materiály
 - U fonoválečku se popisuje: fonováleček (většinou na jednom nosiči jedno dílo) a skladba

•Technická metadata a archivační metadata

- AES57, PREMIS, MIX (obrazy)

- Zaznamenává se: délka nahrávky, velikost souboru, bitová hloubka, vzorkovací frekvence, formát souboru, druh komprese apod.
- Možnost popsat i analogový nosič—materiály, drážku apod.
- Popis digitalizačního procesu- HW a SW použité při digitalizaci
- Evidence životního cyklu souborů
- Některá technická metadata extrahují charakterizační nástroje (FITS, Jhove)

•Strukturální metadata

- vyjadřují, jak spolu jednotlivé objekty souvisejí, a umožňují tak zobrazení dokumentu ve smysluplné podobě

NTK

50°6'14.083"N, 14°23'26.365"E
Národní technická knihovna
National Technical Library

210 mm



Popisná metadata

 Národní muzeum

Hledat v digitální knihovně ...

SbírkyProcházetInformaceEnglish



Auch ich war ein Jüngling
Lortzing, Albert

SKLADBA	ČÁST ALBA
II zvuková stopa	Auch ich war ein Jüngling
zvuková stopa	Kirchenscene

Auch ich war ein Jüngling
aus "Waffenschmied"

Autor
Lortzing, Albert
Gounod, Charles

Nakladatelské údaje
1905-1920

Typ dokumentu
Zvukové nahrávky

Jazyk
Němčina

Místo uložení
Národní muzeum - České muzeum hudby
Signatura: MD 573

Fyzický popis
Rozsah: 1 gramofonová deska

Poznámky
Lortzing, Kirchenscene : aus "Margarethe"
Faust / Gounod
Wilhelm Hesch, zpěv

NTK

50°6'14.083"N, 14°23'26.365"E
Národní technická knihovna
National Technical Library

Techická metadata



```
< mets: mets LABEL="Arie z Komediantů, [mezi 1900 a 1915?]" TYPE="sound recording" xmlns: mets="http://www.loc.gov/METS/" xmlns: dc="http://purl.org/dc/elements/1.1/" xmlns: mods="http://www.1
< mets: metsHdr CREATEDATE="2019-07-04T22:32:49" LASTMODDATE="2019-07-05T10:39:56">
  < mets: agent ROLE="CREATOR" TYPE="ORGANIZATION">
    < mets: name>ABAO01< mets: name>
  </ mets: agent>
  < mets: agent ROLE="ARCHIVIST" TYPE="ORGANIZATION">
    < mets: name>ABAO01< mets: name>
  </ mets: agent>
</ mets: metsHdr>
< mets: amdSec ID="SOUND_0001">
  < mets: techMD ID="AES57_001">
    < mets: mdWrap MDTYPE="NISOIMG" MIMETYPE="text/xml">
      < mets: xmlData>
        < aes: aes xmlns: aes="http://www.aes.org/audioObject">
          < aes: audioObject ID="1" title="to be filled in" disposition="" analogDigitalFlag="FILE_DIGITAL" schemaVersion="1.0.0">
            < aes: format>Waveform Audio</ aes: format>
            < aes: audioDataEncoding>PCM</ aes: audioDataEncoding>
            < aes: byteOrder>LITTLE_ENDIAN</ aes: byteOrder>
            < aes: firstSampleOffset>304</ aes: firstSampleOffset>
            < aes: audioDataBlockSize>4</ aes: audioDataBlockSize>
            < aes: use useType="OTHER">PRESERVATION_MASTER</ aes: use>
            < aes: primaryIdentifier identifierType="FILE_NAME">sa_mb15c210-d3f7-11e6-a1de-001018b3eb4c_0001.wav</ aes: primaryIdentifier>
            < aes: face audioObjectRef="1" label="face_1" ID="FACE_60bb2a68-626f-43b3-b5fb-43cd357bad58" direction="NONE">
              < aes: timeline>
                < aes: startTime editRate="96000">0</ aes: startTime>
                < aes: duration editRate="96000">5448408</ aes: duration>
              </ aes: timeline>
              < aes: region faceRef="FACE_60bb2a68-626f-43b3-b5fb-43cd357bad58" formatRef="FORMAT_REGION_f63ce54e-4109-477d-bd2f-90e07dd2e183" ID="REGION_0e236548-c26d-4013-b54a-a0c7f16408c1">
                < aes: timeRange>
                  < aes: startTime editRate="96000">0</ aes: startTime>
                  < aes: duration editRate="96000">5448408</ aes: duration>
                </ aes: timeRange>
                < aes: numChannels>2</ aes: numChannels>
                < aes: stream ID="STREAM_177d5c12-58cc-474b-b6bf-e7350cdcc77b" label="stream_0" faceRegionRef="REGION_0e236548-c26d-4013-b54a-a0c7f16408c1">
                  < aes: channelAssignment leftRightPosition="0.0" channelNum="0" frontRearPosition="0.0"/>
                </ aes: stream>
                < aes: stream ID="STREAM_1ee7b66a-9712-497f-94f7-66a5f66b13b9" label="stream_1" faceRegionRef="REGION_901d55ce-cc8c-4206-b75a-c1e34e2a4467">
                  < aes: channelAssignment leftRightPosition="0.0" channelNum="1" frontRearPosition="0.0" />
                </ aes: stream>
              </ aes: region>
            </ aes: face>
          < aes: formatList>
            < aes: formatRegion ownerRef="REGION_0e236548-c26d-4013-b54a-a0c7f16408c1" ID="FORMAT_REGION_f63ce54e-4109-477d-bd2f-90e07dd2e183" label="format_region_1">
              < aes: bitDepth>16</ aes: bitDepth>
              < aes: sampleRate>44100.0</ aes: sampleRate>
            </ aes: formatRegion>
          </ aes: formatList>
        </ aes: aes>
      </ mets: xmlData>
    </ mets: mdWrap>
  </ mets: techMD>
</ mets: amdSec>
```

Souborové formáty

- Celá řada formátů pro zvukové dokumenty
 - MP3, WAVE, AIFF, FLAC, ALAC, WMA, Monkey.....
- Úspěšné dlouhodobé uchovávání= využívání dokumentovaných, rozšířených formátů
 - WAVE, BWF
- Zpřístupnění= komprimované formáty, široce rozšířené
 - MP3



SIP balíček zvukového dokumentu



METS

popisná, právní a strukturální metadata



Checksum

kontrolní součty pro soubory v balíčku (MD5)



Info

informace o obsahu balíčku



catalog_entry

záznam z katalogu v MARCXML



source_audio

zvukové soubory WAV pro archivaci



mastercopy_audio

zvukové soubory WAV upravené



usercopy_audio

zvukové soubory MP3 pro zpřístupnění



mastercopy



usercopy

Zdigitalizované obaly, booklety, nosiče v archivní a uživatelské kvalitě



amdsec

technická a archivační metadata pro zvukové a obrazové soubory



alto



txt

výstup z OCR

NTK

50°6'14.083"N, 14°23'26.365"E

Národní technická knihovna
National Technical Library



210 mm

Děkuji za pozornost

Dotazy.....

Kontakt:

Natalie Ostráková natalie.ostrakova@nkp.cz