

Digitoinnin työpaja 3/4:
äänitteiden digitoinnin perusteet 26.3.2010

Äänitedigitointi Kotuksessa 1999–2010

Varmistaminen ja pitkäaikaissäilytys
(PAS)

Jussi Rusi



Äänitedigitointi Kotuksessa

- 1998 työryhmän esitys SKNA:n äänitteiden digitointisuunnitelmaksi (mukana myös tietotekniikkaosaajia):
 - avoin digitointityöasematarkaisu (PC-mikrotyöasema)
 - digitointi pyritään tekemään mahdollisimman laadukkaasti (aineisto digitoidaan vain kerran):
 - häviötön standardinmukainen korkeatasoinen äänitiedostoformaatti (PCM- koodattu Microsoft Wave)
 - häviölliset kuuntelukopiot (mp3)
 - alkuperäisaineiston häiriöitä tai virheitä ei korjailta eikä tasojta normalisoida
 - kahdennettu ja riittävän suuri varmistuskapasiteetti, varmistusten hajautettu säilytys
 - digitointi reaaliaikaista ja tehtävä itse (1-2 digitointilinjaa)
 - osin ostopalveluna (v. 2001 →)
- tutustumiskäynnit mm. Tampereen yliopiston Kansanperinteen laitos, Tartossa
- Rahvaluule arhiivissa, useiden laitevalmistajien, maahantuojien ja myyjien luona
- tärkeimpinä konsultteina Stig Hedlund (HF Media Solutions Ab) ja Mika Pellinen (Diginord Oy)

Digitoinnin tietotekninen laiteympäristö

- digitointilaitteisto:
 - laadukkaalla äänikortilla varustettu tehokas PC-mikro (Windows/Linux-käyttöjärjestelmät) digitointi- ja varmistustyöasemaksi
→ erillinen varmistustyöasema → osaksi Kotuksen IT-järjestelmän varmistusta
 - erillinen ulkoinen A/D-muunnin (44,1 kHz/ 16-bit), jossa tarvittavat liitännät
→ 48 kHz/ 24-bit ja 96 kHz/ 24-bit
 - mikseri, kaiuttimet, kaapelointi
- digitointiohjelma(t):
 - SoundForge v. 4.5, NoiseReduction (plug-in)
→ SoundForge v. 9.0, BatchConverter (plug-in)
- varmistuslaite ja ohjelmat:
 - SLR50-varmistusnauhuri ja -media (10 kpl 25 GB, n. 75 t /mono)
→ Ultrium LTO-3 (n. 30 kpl 400 GB, n. 800 t /mono) → Ultrium LTO-4
 - varmistusohjelmistot Seagate Backup (Windows) ja tar (Linux)
→ Veritas Backup Exec ja Tar98 (Windows) → Retrospect
- Laitteisto- ja ohjelmistokustannukset aloitettavana lähes 16000 €
 - varmistusnauhuri 4000 €
 - A/D-muunnin 3000 €
 - PC-mikro 2500 €
 - kaiuttimet 1500 €
 - varmistusmedia (kokonaiskapasiteettiä 500 GB) 1200 €

Osaamisen kehittäminen ja ylläpitäminen sekä tiedonkulku

- jäsenyys IASA:ssa (2000 →)
 - hyödyllisiä seminaareja, mm. Yleisradio, Itävallan tiedeakatemian Phonogrammarchiv (perusfilosofia: ”ikuinen” digitaalinen äänitearkisto)
 - Kotuksen digitointihankkeen toteutustapa kiinnostanut mm. Gambian, Hollannin ja Ruotsin edustajia
- SKNA:n Digitointihankkeen tukiryhmä (2000-2006)
 - digitointijärjestys
 - tietotekniset ratkaisut
- Tieteellisten äänitearkistojen digitointi ja tiedonhallinta –yhteenliittymä (2000-) digitoinnin jouduttamiseksi lisärahoituksella
- → Digitoinnin tekninen työryhmä (2001-2007) → EU:n TAPE-projekti (Training for Audiovisual Preservation in Europe) (2004-2008) → DigiWiki- (2007-) ja Äänistö-hankkeet
 - mm. tietokantatyö, workshopit ja seminaarit

Digitoinnin haasteita

- tekniikan kehitys
 - ammattikäyttöön tarkoitettujen analogisten laitteiden ja varaosien saatavuus (valmistus lopetettu tai ovat kalliita) → laitehuolto
 - varsinkin tietotekniikan lyhyt elinkaari ja tuotetuki (käyttöjärjestelmän ja sovellusohjelmien yhteensopivuus, laitesukupolvet, media)
- tehdäkö itse vai hankkiako ostopalveluna?
 - itse tehtynä kolme työtuntia tuottaa yhden tunnin digitoitua ääntä
 - sopivat digitointiolosuhteet (Kotuksen vanha digipaja: akustolevyseinät, mutta toimimaton ilmasto → kesällä ikkuna auki)
 - ostopalvelu säästää myös omaa analogista laitteistoa tai hankintakustannuksia kun käytettävissä ei ole riittävän laadukasta laitteistoa
 - rappeutuneet ja uhanalaiset tallenteet:
 - esim. vahaliერიöt, asetaattinauhat, ohuet nauhat, C-kasetit, Minidisc-levyt
 - tarvitaan ehkä kallista restaurointikäsitelyä (kosteuden poistaminen väliaikaisesti, ”uunissa paistaminen”)

Lisää digitoinnin haasteita

- yhteistyökumppanien valinta (ostopalvelu)
 - mistä tunnistaa laatukriteerit täyttävä digitointi
 - selvitys käytettävästä laitteistosta (sekä analoginen että digitaalinen)
 - referenssit, nettisivut
 - pienen koe-erän digitointi harkitusti valituista äänitteistä (esim. digitoinnin eri vaiheita ja laitteiston testausta varten tehty nauhoite)
 - prosessien kontrollointi ja laadunvalvonta
 - digitoinnissa syntyviä virheitä, esim.
 - heikkolaatuinen A/D-muunnos
 - signaalin häiriöt
 - tasojen säätö (yli- tai aliohjautuvuus)
 - riittämättömät järjestelmäresurssit (katkoksia, äänitekohtien kahdesti toistumista ja staattisia purkauksia)
 - varmistuksessa syntyviä virheitä, esim.
 - mediavirheet → tiedoston tarkistusluku (MD5 hash)
 - arkistointiin kelpaamaton varmistusmedia: R-DAT-nauha, optinen media (mm. CD-R, DVD-R) tai ulkoiset (siirto)kiintolevyt
- <http://unesdoc.unesco.org/images/0014/001477/147782E.pdf>

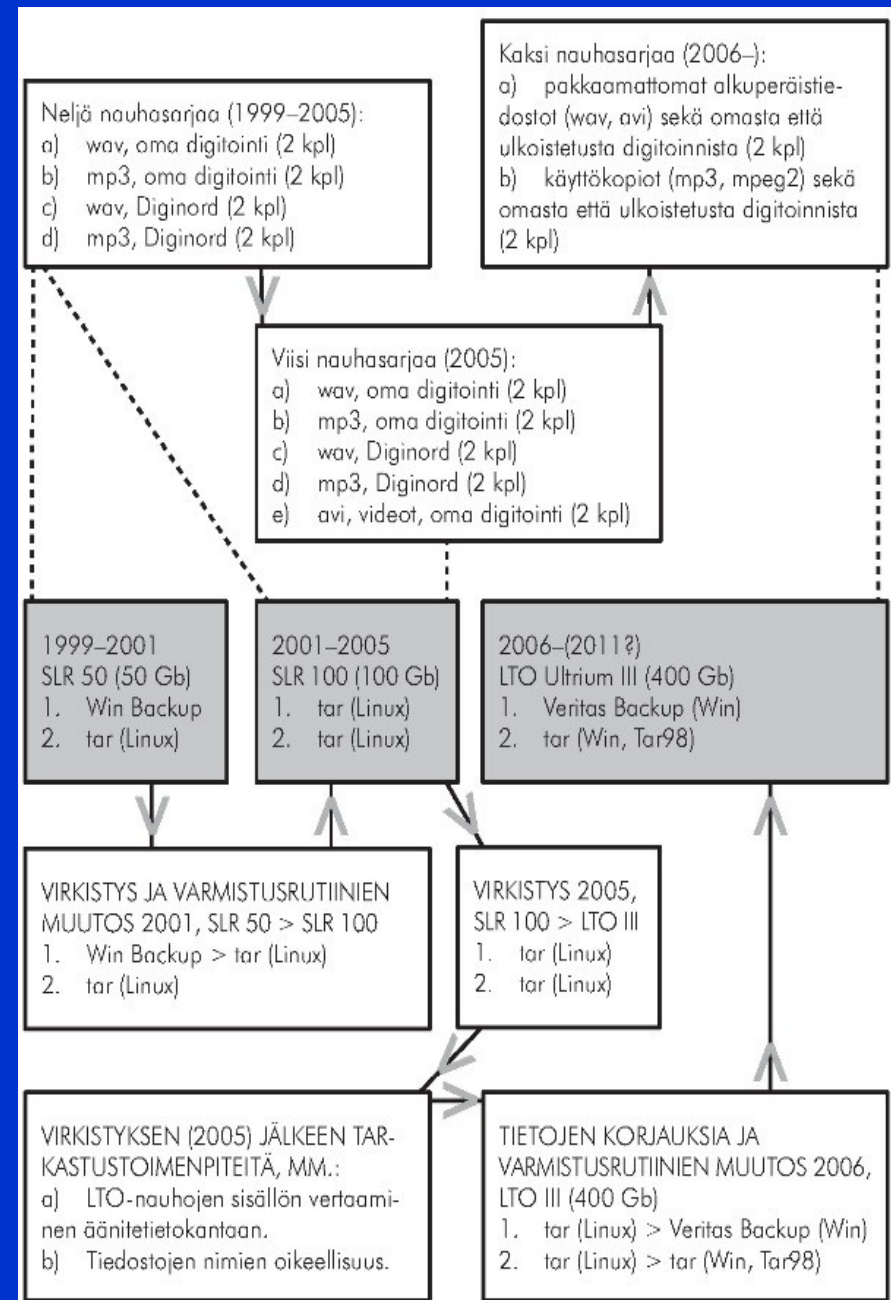
Varmistamisesta ja pitkäaikaissäilytyksestä

- suosituksia varmistamiseen ja varmistusten pitkäaikaissäilytykseen:
 - kahdennettu varmistus (arkisto- ja käyttökopio, jopa eri varmistusohjelmalla)
 - varmistusten siirto uuteen järjestelmään n. 5-7 v. välein
 - suositeltavaa varmistaa myös varmistusohjelman nauha- tai varmistuskohtainen kirjanpito sekä tiedostojen tarkistusluku
 - myös varmistusnauhojen pitkä säilyvyys edellyttää kelvollista säilytystä (huom! oikea säilytysasento)
 - arkistotila: lämpötila- ja kosteussuositukset (5-18° C, 30 %, ei suurta vaihtelua), metallihyllyt, sammutusjärjestelmä, sijainti

Kotuksen digitoinnin varmistuksista

- varmistusnauhojen virkistys käytännössä n. 5 - 7 v. välein kun ohjelmistoja ja/tai nauhureita vaihdetaan uusiin, vaikka data säilyykin datakasetilla 20 - 30 v.
- tallennusmedian ja -laitesukupolvien yhteensopivuus: esim. uusi laite lukee kahta edellistä nauhatyyppiä, mutta kirjoittaa vain yhtä edellistä

Julkaistu	DLT	SLR	LTO
1996	DLT IV (35 GB)		
1997		SLR50 (25 GB)	
1998	SDLT I (160 GB)		
1999	DLT IV (40 GB)		
2000		SLR100 (50 GB)	LTO-1 (50 GB)
2003	VS1 (80 GB)	SLR-140 (70 GB)	LTO-2 (200 GB)
2004	SDLT II (300 GB)		
2005	VS1 (160 GB)		LTO-3 (400 GB)
2006	S4 (800 GB)		
2007			LTO-4 (800 GB)
2010			LTO-5 (1.6 TB)
TBA		SLR-200 (100 GB)	
TBA		SLR-400 (200 GB)	LTO-6 (3.2 TB)



Pitkäaikaissäilytyksestä (PAS)

- Digitaalisen aineiston tuottamiseen tulee aina sisältyä aineistoa koskevan metatiedon kerääminen pitkäaikaissäilytyksen tarpeita silmällä pitäen
- Aineistoa ja sen elinkaarta koskevat metatiedot säilytetään yhtä huolellisesti kuin itse aineisto
 - esim. tietokanta, bwf-äänitiedostoformaatti (wav-formaatin laajennus, jota digitointiohjelmat jo tukevat)
 - uniikki tunniste
 - tietoa alkuperäisestä äänitteestä, digitointiprosessista sekä digitaalisesta tallenteesta (mm. laitteet, tallennusformaatit, näytteenottotaajuus, erottelukyky, digitointiohjelma ja sen versio)
 - aineiston omistajuus ja sitä koskevat käyttörajoitukset ym. hallinnolliset tiedot
 - aineistoon kohdistuvat säilytys- ja muut toimenpiteet
- OpM:n 2008 aloitettu Kansallinen digitaalinen kirjasto -hanke
 - tavoite v. 2016 on tarjota muistiorganisaatioille sekä muille sellaisille organisaatioille, joilla on pitkäaikaisesti tai pysyvästi säilytettäviä kulttuuriperintöaineistoja luotettava ja käytettävyydeltään monipuolinen sähköisen pitkäaikaissäilyttämisen palvelu
- Tulevaisuus? → Migraatio myös toiseen äänitiedostoformaattiin?

1. TALLENNE

11980

signuminaiku

2. ÄÄNITYS

2

signuminloppu

3. KIELENOPAS

15395

kielenoppaan ID-nro

Tallenteiden ja äänitysten ja kielenoppaiden tiedot ovat tietokannassa vain kerran. Tämä näkymä yhdistää tiedot toisiinsa.

media avokelanauha
 audiovideo äänite
 kartuntavuosi
 tallenteen kopio
 tallenteen tietoja

sukunimi Rusi
 etunimi Ville
 omaasukua
 kuts.nimi
 elämänavaih. Elänyt synnyinkylässään vuoteen 1939, jolloin lähti talvisotaan (työkomppaniaan).

kieli1	suomi (murteet)	11980	2
kieli2		11981	1
kieliryhma	1. suomi	11982	1
nauhapaikka	B 0000 - loppu	11985	1

ammatti kalastaja sukup. mies
 synt.aika 1.2.1893 vuosik. 1890
 synt.paikk. Kuolemajärvi, Seivästö
 pitaja Kuolemajärvi
 kunta Luovutetulla alueella
 kyla Seivästö
 komment. Varsin hyvämurteinen ja vuolas kielenopas.

kokoelma	1. kielitieteellinen kokoelma	sisältö	
äänitteen kesto	0.3 äänitteenkestomin 30	tekn. kommentit	
nauhoitus aika	10.9.1974	äänityskomentit	
nauhoituspaikka	Mikkelin mlk., Ihastjärvi, Pulliala	muut kommentit	
haastattelija	Yli-Paavola, Jaakko	käyttörajoitus	
mono/stereo		rajoituskoodi	1 - vapaassa käytössä sääntöjen mukaisesti
litterointi		äänityksen kopio	
tallennuslupa		kopion formaatti	
digitointiformaatti	wav; 48000; 24	rahoittaja	Digitoinnin on rahoittanut Kuolemajärvi-säätiö.
digitoija	Diginord b		
digitointipäivä			

Yllä määriteltyyn äänitteeseen liitetyt kielenoppaat:

Kielenoppaan ID	sukunimi	etunimi	syntymäaika	pitäjä
15395	Rusi	Ville	1.2.1893	Kuolemajärvi

Digitaaliset tiedostot:

11980_2az.wav	LTO_09
---------------	--------

