

Aineiston LTFS-nauhavarmennus Musiikkiarkistossa

Documents related to the management of the Music Archive Finland's resources

- 1 "Aineiston järjestäminen ja nimeäminen Musiikkiarkistossa" (Organizing and naming of resources in the Music Archive Finland)
- 2 "Aineiston LTFS-nauhavarmennus Musiikkiarkistossa" (LTFS tape backup of resources in the Music Archive Finland)
- 3 "Aineiston hallinta Musiikkiarkistossa" (Management of resources in the Music Archive Finland).

This document introduces the general principles of tape verification of the Music Archive Finland's resources, as well as the software and hardware required by the functions and their settings. The processed material (file, object) must comply with the specifications of the document "Aineiston järjestäminen ja nimeäminen Musiikkiarkistossa" (Organizing and naming of resources in the Music Archive Finland).

The document presents the general principles of the activities. The exact workflows of the processes, together with the functions to be performed, are described in the appendices.

Tape verification is performed on LTO-6 tapes using LTFS technology. All information about objects relevant for long-term storage is stored in the Skaala database maintained by the Music Archive Finland.

The backup copy is performed on a Mac computer specially reserved for it, to which access is blocked from both the intranet and the Internet. A tape drive HPE StoreEver LTO-6 Ultrium 6250 Ext is connected to the computer with an ATTO Thunderlink™ SH 2068 adapter. The material to be backed up is copied to tapes from external USB 3 hard drives.

It is possible to copy approximately 2.375 TB of material onto one LTO-6 tape. Each tape and the objects to be copied to the tapes are given unique identifiers, which are stored in the Skaala database. The strain on the tapes is minimized by writing the tape completely in one go, after which the disk's physical write protection is turned on. The tapes are stored in stable archive conditions for the duration of their lifetime and no additional stress is applied to them: the tapes are primarily backups of material copies in active use, which are located in RAID hard disk packs and on individual external hard disks.

Before copying to tape, the technical metadata of the files is captured with the ExifTool program and stored in the electronic archive of the Music Archive Finland. The filenames of objects are checked, and their MD5 and SHA-1 hashes are calculated. After successful copying to tape, the material is copied back to the hard disk. The MD5 and SHA-1 hashes of the returned objects are calculated, after which the calculated hashes before and after the tape write are compared. If the hashes match, the backup has been successful, and the object hashes with tape locations are stored in Skaala. In addition, Skaala stores actions on the verification tape with log data.

The LTO tapes are refreshed and the material stored on them is inspected approximately every five years, unless the material is under the custody of the Digital Preservation Service (DPS) maintained by the IT Center for Science (CSC) (see the document "Aineiston hallinta Musiikkiarkistossa"; Management of resources in the Music Archive).

Sisällys

1 Sisällys	3
2 Johdanto	4
3 Nauhavarmennuslaitteisto ja käytettävät ohjelmistot	4
3.1 Laitteisto	4
3.2 Ohjelmistot ja niiden asetukset	5
3.3 Tietokoneen käynnistäminen	6
3.4 Toimintojen aikatauluttaminen	6
4 Nauhavarmennus Musiikkiarkistossa	6
4.1 Nauhavarmennuksen yleiset periaatteet	6
4.2 Tietokannan päivittäminen ja lokitietojen luominen	7
4.3 Aineiston järjestäminen ja valmisteleminen nauhavarmennusta varten	7
4.4 Tiedostonimien tarkistaminen	7
4.5 Tiedostojen validointi	8
4.6 Ylimääräisten tiedostojen poistaminen	8
4.7 Objektien metatietojen ja nauhaindeksien luominen	9
4.8 Nauhan kirjoittaminen ja palautustarkistus	9
5 Liitteet	11
5.1 Liite 1. Varmuuskopioinnin työnkulun yleinen kuvaus	11
5.2 Liite 2. LTFS-toimintojen lokitemplate. (Musiikkiarkiston sisäinen dokumentti)	12
5.3 Liite 3. LTFS-varmennuksen työnkulku Pääte-komentoineen (Musiikkiarkiston sisäinen dokumentti)	12
5.4 Liite 4. Tiivistetiedostojen tietokantatarkistus (Musiikkiarkiston sisäinen dokumentti)	12
5.5 Liite 5. Koodistot (Musiikkiarkiston sisäinen dokumentti)	12
5.6 Liite 6. Huomiot tekstissä mainituista dokumenteista	12

Johdanto

Musiikkiarkiston aineiston hallintaan liittyvät dokumentit:

1 Aineiston järjestäminen ja nimeäminen Musiikkiarkistossa

2 Aineiston LTF5-nauhavarmennus Musiikkiarkistossa

3 Aineiston hallinta Musiikkiarkistossa

Tässä dokumentissa esitellään Musiikkiarkiston aineiston nauhavarmennuksen yleiset periaatteet sekä toimintojen vaatimat ohjelmistot ja laitteistot asetuksineen. Aineistolla tarkoitetaan digitaalisessa muodossa olevia objekteja (tiedostoja), jollei toisin mainita.

Musiikkiarkistossa käsiteltävän aineiston on oltava dokumentin "Aineiston järjestäminen ja nimeäminen Musiikkiarkistossa" määritysten mukaista.

Nauhavarmennuslaitteisto ja käytettävät ohjelmistot

Nauhavarmennus suoritetaan LTO-6-nauhoille LTF5-teknologiaa käyttäen. Käytössä oleva laitteisto ja ohjelmistot sekä niiden mahdollisesti erikseen määritetyt asetukset on mainittu alla. Varmuuskopiointi onnistuu myös monin muin tavoin, ja ohjelmistoja sekä laitteistoa kannattaa ajoittain päivittääkin – joskin päivittäessä on huolehdittava toiminnan luotettavuudesta sekä vakiintuneiden käytänteiden jatkuvuudesta.

Dokumentissa mainittu tietokanta tarkoittaa Musiikkiarkiston Skaalaa, johon kaikki pitkäaikaisäilytyksen kannalta oleellinen tieto kirjataan. Lisäksi tietoa talletetaan tekstitiedostoissa muun muassa Musiikkiarkiston verkkolevylle, jossa lisäksi sijaitsevat tässä dokumentissa mainitut dokumentit ja ohjelmistojen asennustiedostot. Varmuuskopioitavaan aineistoon viitataan digitaalisina tai sähköisinä objekteina tai tiedostoina.

Laitteisto

Nauhalaitteisto tukeutuu Thunderbolt 2- sekä SAS-liitäntöihin.

- Tietokone: Mac Mini (Late 2014)
- Nauha-asema: HPE StoreEver LTO-6 Ultrium 6250 Ext
 - Firmware: 35GD/Standalone (Ver 019.725, Date 2016/10/06-11:06)
 - Asema on liitetty adapterin avulla suoraan tietokoneeseen, johon ei oletuksena ole pääsyä sisäverkon kautta; tietokonetta ja nauha-asemaa käyttää vain yksi henkilö kerrallaan
 - LTF5-nauhoja ei ole tarkoitettu aktiivikäytössä olevaksi levytilaksi: se ei ole ns. *random-access drive*, vaan lineaarinen media, kuten termit LTO (*Linear Tape Open*) ja LTF5 (*Linear Tape File System*) ilmaisevat
- Thunderbolt 2 – SAS -adapteri: ATTO Thunderlink™ SH 2068
 - ThunderLink™ SH 2068 Flash Bundle Version 170828 (osx_fls_tlsashbab_170828.dmg)
 - YoYotta-dokumenteissa kehoitetaan käyttämään versiota ThunderLink™ SH 2068 Flash Bundle Version 160129 (osx_fls_tlsashbab_160129.dmg) TLR-tuen puutteen vuoksi (Transport Layer Retries, jota ATTO ei enää tue lukuvaiheessa vanhemmille nauha-asemille tai adaptereille?)
 - Koskee todennäköisesti YoYottan omaa varmennusohjelmistoa. Käytetään uutta versiota ja palataan vanhaan, jos ongelmia ilmenee

- Dokumentin julkaisun jälkeen on tullut yksi uusi ajuriversio lisää, joka nyt on käytössä (20210315), mutta edelleen YoYottan sivuilla nähtävissä ole dokumentti kehottaa käyttämään Jan 2016 -versiota
- ThunderLink™ SH 2068 Release v1.40 (osx_drv_tlsashba_1400f1.dmg)
- Kovalevyt: useita USB 3 -liitäntäisiä levyjä
 - Levyt liitetään suorituskyvyn maksimoimiseksi suoraan tietokoneeseen eikä esimerkiksi käytetä yhteiskäytössä olevia verkkolevyjä. Esimerkiksi Gigabitin Ethernetin nopeus (100 Mt/s) on LTO-6:n nopeutta hitaampi (160 Mt/s)
 - LTO-6 nauhan täysi kirjoittaminen/lukeminen kestää laskennallisesti noin 4 t 10 min (2,4 Tt nopeudella 160 Mt/s). Käytännössä nauhan lukemisen tai kirjoittamisen voi olettaa kestävän 5-6 tuntia siitä hetkestä, kun kirjoitettavan aineistohakemiston siirtää *Finderilla* kohteeseen

Ohjelmistot ja niiden asetukset

Asetukset on dokumentoitu hyvin yksinkertaisesti.

- Tietokoneen käyttöjärjestelmä: *OS X 10.13.6 High Sierra*
 - Estetään kovalevyjen ja tietokoneen automaattinen nukkuminen, jotta työt eivät keskeydy suunnittelematta
 - Estetään ulkoisten kovalevyjen *Spotlight*-indeksointi
- Nauha-aseman työkalu: *HPE Library and Tape Tools 4.28 for Mac*. Komennot suoritetaan *Päätteessä*
 - Käynnistyskomento: `hpe_ltt`
 - Tarvitaan vain nauha-aseman firmwaren päivityksessä yms. ylläpitotoiminnoissa ja mahdollisissa ongelmatapauksissa
 - Unix-versio 4.30 saatavana, ei kokeiltu asentuisiko Maciin
- Nauhaohjelmisto: *HPE StoreOpen 3.4.2 for Mac*
 - Indeksien päivitys hyvä olla 60 minuuttia tai enemmän. Kirjoitetaan lähtökohtaisesti koko nauha yhdellä kirjoituskerralla täyteen
- ATTO Thunderlink™ SH 2068:n määrittäsohjelmisto: *ATTO Configuration Tool v4.34*
 - Tarvitaan vain adapterin firmwaren tai ajurien päivityksessä yms. ylläpitotoiminnoissa sekä mahdollisissa ongelmatapauksissa
 - Network – localhost – ThunderLink SH 2068 – Channel 1 – NVRAM-välilehti – Phy Speed 3.0 Gb/s
- Tiedostonimien käsittely: *A Better Finder Rename 11*
- Metatietojen käsittely: *ExifTool* by Phil Harvey
 - Komennot suoritetaan *Päätteessä*
 - Ohjelmaa päivitetään aktiivisesti. LTFS-varmennukset aloitettiin versiolla 12.21
- Laajennusattribuuttien (*extended attributes*) käsittely: *xattred*, *Pääte*
- Tiedostojen vertaaminen: *BEdit*
 - Aloitettu versiolla 12.6.7
- OS X:n perusasennukseen sisältyvät ohjelmat: *Finder*, *Pääte*
 - *Päätteessä* suoritetaan edellä mainittujen toimintojen lisäksi esimerkiksi ylimääräisten tiedostojen poistaminen, hakemistoindeksien luominen sekä tiivisteiden laskeminen
 - *Finderin* esitekatselusarake pidetään pois näkyvistä ja kuvake-esikatselutoiminnot pidetään poissa päältä, jotta vältetään turhaa prosessointia sekä mahdollisesti uusien laajennusattribuuttien luominen työn eri vaiheissa.
 - Edellä mainitun lisäksi yhden nauhahakemiston luominen ja siirtäminen nauhalle tai nauhalta *Finderilla* edesauttaa myös sitä, että nauhaa ei lueta turhaan, eikä *Finder* esimerkiksi yritä järjestää nauhalla sijaitsevia lineaarisesti (*linear vs. random-access*) kirjoitettuja tiedostoja kovalevyillä sijaitsevien tiedostojen tavoin tai luoda niille esikatselukuvia ja -ikoneita yms.
 - Kaiken *Finderin* suorittaman prosessoinnin voi ohittaa käyttämällä varmuuskopiointiin *Päätettä* tai mahdollisesti jotain muuta tiedostojen käsittelyyn tarkoitettua ohjelmaa, esimerkiksi
 - *Commander One*, *Path Finder*, *Forklift*, *XtraFinder*. HUOM! Mainittuja ohjelmia ei ole testattu

Tietokoneen käynnistäminen

Nauha-asema tulee olla liitettyä ATTO Thunderlink™ SH 2068 -adapteriin ennen tietokoneen käynnistämistä. Nauha-asema käynnistetään ensimmäisenä. Tietokone käynnistetään, kun nauha-asema on valmis (*Ready*-valo etupaneelissa on päällä jatkuvasti). ATTO Thunderlink™ SH 2068 -adapteri käynnistyy tietokoneen myötä. Kovalevyt voidaan käynnistää tai liittää tietokoneeseen tarpeen mukaan.

Toimintojen aikatauluttaminen

Suurten datamäärien kopiointi ja käsittely vie aikaa. Käytännössä yhden varmennusnauhan luominen tarkastuksineen kestää useita työpäiviä siitä hetkestä alkaen, kun noin 2,4 teratavua aineistoa sisältävä nauhahakemisto on saatu luotua.

Toimintojen kesto riippuu laitteistosta: muun muassa muistin määrästä, kovalevyn kunnosta ja kapasiteetista sekä prosessorin nopeudesta. Varsinkin tiivisteiden laskeminen *Päätteen openss*:llä on hidasta, sillä se hyödyntää vain yhtä prosessoriydintä. Vaihtoehtoja monisäikeiseen laskemiseen lienee olemassa, mutta niitä ei ole toistaiseksi saatu onnistuneesti toimimaan (20210319).

Edellä on mainittu aikaavievimpien toimintojen arvioituja kestoja, ja niiden mukaan kannattaa työtä yrittää rytmittää siten, että esimerkiksi nauhalle tai nauhalta kopioinnin (kesto noin 5–7 t) tai tiivisteiden laskemisen (kesto noin 17–18 t) aloittaa työpäivän päätteeksi, jotta seuraavana päivänä voi jatkaa työtä mahdollisimman sujuvasti.

Töitä voi myös jakaa tehtäväksi useammalla koneella mikäli varmuuskopioitavaa tai virkistettävää aineistoa on paljon ja sen on jakanut useille ulkoisille kovalevyille. Esimerkiksi LTO-koneen suorittaessa nauhatoimenpiteitä, voi toisaalla tarkastaa tiedostonimiä, laskea tai tarkastaa tiivisteitä sekä luoda indeksi- tai metatietotiedostoja.

Nauhavarmennus Musiikkiarkistossa

Toiminnot kuvataan alla yleisellä tasolla. Tarkat työnkulkukuvaukset suoritettavine komentoineen ovat liitteissä.

Nauhavarmennuksen yleiset periaatteet

Jokaiselle nauhalla annetaan yksilöllinen tunnusnumero (*id*), joka LTO-6/LTFS-nauhojen kohdalla aloitetaan numerosta 96 aikaisempien numeroiden ollessa varattuja edeltävien sukupolvien varmennusnauhoille. *Id*-numero kirjoitetaan nauhaan ja sitä käytetään nauhan yksilöllisenä tunnisteenä. Lisäksi nauhalle annetaan nimi, joka on niinkään yksilöllinen koostuen käytetystä teknologias- ta sekä nauhan *id*:stä: ”{teknologia+id}”, esimerkiksi ”ltfs96” tai ”lto101”. Nauhan nimeen ei saa sisältyä mitään erikois- tai tyhjämerkkejä.

Nauhoja säilytetään pystyasennossa mahdollisimman vakaisissa arkisto- tai toimisto-olosuhteissa (16–32 °C, suhteellinen ilmankosteus 20–80 %, ei pikaisia olosuhdemuutoksia), eikä niihin kohdisteta tarpeetonta räsitusta millään tavalla. Ennen ja jälkeen nauhaan kohdistuvaa toimenpidettä, nauhaa säilytetään samassa tilassa muun laitteiston kanssa useita tunteja (24 t), jotta olosuhteet ovat toimenpiteen alusta asti sekä nauhoille että laitteistolle mahdollisimman yhteneväiset.

Nauhalle kirjoitettava aineisto siirretään nauhanimen mukaiseen hakemistoon (esimerkiksi hakemiston ”ltfs96” sisältö kirjoitetaan nauhalle, jonka nimi on ”ltfs96” ja *id* ”96”) ja nauha kirjoitetaan

lähtökohtaisesti täyteen yhdellä kirjoituskerralla, jonka jälkeen laitetaan nauhan fyysinen kirjoitus-suoja päälle. Näin saadaan optimoitua nauhalle varmennettavan aineiston määrä minimoiden nauhaan kohdistuva rasitus.

Nauhalle tehdään lukutoimenpiteitä vain tarpeen mukaan aineistoa palauttaessa, mutta kuitenkin vähintään yhden kerran varmennuskirjoituksen jälkeen varmistettaessa varmuuskopiointiin onnistuminen.

Tietokannan päivittäminen ja lokitietojen luominen

Varmennettavien tiedostojen (LTFS-hakemistojen) käsittelystä tallennetaan lokitiedosto, johon lisätään tehdyt toiminnot ja niihin liittyvät huomiot. Tiedostot tallennetaan verkkolevylle, jonka tiedot kopioidaan myös tietokantaan nauhatoimenpidekohtaisiksi lokitiedoiksi.

Aineiston järjestäminen ja valmisteleminen nauhavarmennusta varten

Varmuuskopioitava aineisto kopioidaan nauhanimen mukaiseen hakemostoon, jonka koko ei saa ylittää 2,375 teratavua. Raja määritetään tarvittaessa pienemmäksi, jos varmuuskopiointit epäonnistuvat liiallisen datamäärän vuoksi. YoYotta-dokumenteissa mainitaan LTO-6:n kapasiteetiksi 2,3 Tt:

"LTFS creates a separate partition for indexes that uses around 5% of the tape. Also YoYotta reserves about 100GB to ensure the tapes does not over fill."

Täten 2,39 Tt lienee suoraa Finder-siirtoa käytettäessä validi raja-arvo: 5 % 2,5 Tt:n kapasiteetista on 125 Gt → 2,375 Tt. Huomaa, että mainittu 5 % on likiarvo. HPE Best Practises -dokumentti:

"As a very rough rule of thumb, each file stored in an LTFS volume will add about 1 kB to the index size, and so a volume containing 1,000,000 files is likely to require approximately 1 GB for the index."

Pitkäaikaisesti varmennetaan vain järjestettyä aineistoa, joka on nimetty Musiikkiarkiston nimeämiskäytänteiden mukaisesti, ks. dokumentti "Aineiston järjestäminen ja nimeäminen Musiikkiarkistossa". Täten mille tahansa nauhalle voidaan kirjoittaa mitä tahansa aineistoa arkistosta tai sarjasta riippumatta. Tiedostonimiä voidaan muuttaa aineiston uudelleenjärjestämisen yhteydessä tai virheelliset nimet korjata tarvittaessa nauhavarmennuksen jälkeen, mutta tällöin on talletettava tietokannan korjauslokiin tiedoston vanha sekä uusi nimi, jolloin esimerkiksi tiedostojen mahdollisesti aikaisemmin lasketut tiivisteet voidaan liittää uudelleennimettyihin tiedostoihin, ks. dokumentti "Aineiston hallinta Musiikkiarkistossa".

Päätettä käytettäessä indeksien ja metadatatiedostojen luontiin siirrytään varmuuskopioitavan hakemiston ylähakemistoon ja tarvittavat toiminnot kohdistetaan kirjoitettavaan hakemistoon koennolla "./{kansionimi}". {Kansionimi} viittaa aina toimenpiteiden kohteena olevaan kovalevylle sijaitsevaan hakemistoon, joka vastaa nauhan nimeä. {Nauhanimi}-termiä käytetään, kun viitataan toimenpiteiden kohteena olevaan nauhaan tai sillä sijaitsevaan aineistoon.

Tiedostonimien tarkistaminen

Ennen nauhavarmennusta tiedostonimet tarkistetaan *A Better Finder Rename* -ohjelmaa käyttäen. Eri tehtäville on luotu toiminnot, jotka valitaan suoritettavaksi ohjelman kohdassa *Action preset*. Suoritettavia toimintoja ovat esimerkiksi

- diakriittisten merkkien poistaminen
 - Toiminnolla saadaan helposti korvattua kielletyt erikoismerkit vastaavilla sallituilla ASCII-merkeillä
 - Sallitut ASCII-merkit, ks. dokumentti "Aineiston järjestäminen ja nimeäminen Musiikkiarkistossa"
- määritettyjen erikoismerkkien korvaaminen
 - Väilyönnit ja muut kielletyt erikoismerkit poistetaan tai korvataan sallituilla merkeillä

Lisäksi on *Päätteellä* suoritettava lisätarkastuksia, sillä edellä mainitut muutokset suoritetaan ottamatta huomioon tiedostopäätteitä. Esimerkiksi Dicientian arkistossa (tunnus: 361) on useita hakemistonimiä, joissa on pisteitä sekä tiedostonimiä, joissa on useita tai ei yhtäkään pistettä. Useissa tapauksissa nimet myös päättyvät välilyöntiin, mikä on sotkenut *ABFR*-ohjelmalla tehtyjä tarkastuksia. *Päätteellä* tarkastetaan täten *ABFR*-tarkastuksien jälkeen kaikki hakemisto- tai tiedostonimet, joissa mahdollisesti on välilyönti, tai jotka päättyvät pisteeseen. Lisäksi tarkastetaan erikseen hakemistonimet, jotka sisältävät pisteitä. Tiedostonimiä, jotka eivät sisällä pisteitä ei tarvitse tarkastaa, sillä siitä ei ole mitään hyötyä, mikäli tiedostomuotoa ei samalla yritetä tunnistaa ja antaa oikeaa tiedostopäätettä.

Mahdollisuuksien mukaan tarkastetaan myös hakemistohierarkia, esimerkiksi hakemistossa "7Vk1" ei saa olla tiedostoja, jotka alkavat merkkijonolla "7Vk2", "7X1" tai "8Vk". Erinäisiä hakemistonimitarkastuksia tehdään uusia objekteja vietäessä Skaalaan PAS-kirjanpitoon, mutta jo tässä vaiheessa olisi hyvä saada aineisto mahdollisimman oikeaan järjestykseen. Myöhemmässä vaiheessa esimerkiksi tiedostonimikorjauksien tekeminen on suoritettava jälkikäteen jo kirjoitetuille varmennusnauhoille, sekä nauhojen indeksi- ja metatietotiedostoihin.

Ennen toiminnon suorittamista otetaan talteen muutettavien tiedostonimien vanha sekä uusi versio valikosta *File – Save File List...* Huomaa, että osa tiedostonimistä on muutettava manuaalisesti, esimerkiksi tilanteissa, joissa piste korvataan yhdysviivalla poistamisen sijaan. Tällöin tiedostonimikorjauslistaan muutos tehdään toteutuneen muutoksen mukaiseksi.

Tiedostonimi on objektin ensisijainen ainutlaatuinen tunniste. Tästä eteen päin pysyvästi säilytetävien sähköisten objektien nimeämistä ei saa muuttaa ilman, että muutos kirjataan Skaalaan PREMIS-tapahtumana, ks. "Aineiston hallinta Musiikkiarkistossa", "Aineiston hallinta Musiikkiarkistossa - Liite 3 - PAS-objektien luominen Skaalaan" (Musiikkiarkiston sisäinen dokumentti) ja "Aineiston hallinta Musiikkiarkistossa - Liite 4 - PAS-objektien ylläpito Skaalassa, PREMIS-tapahtumat" (Musiikkiarkiston sisäinen dokumentti).

HUOM! Nauhalla sijaitsevia tiedostoja ei saa nimetä uudestaan Finderissa, vaikka se lieneekin mahdollista! Nauhan omat indeksit eivät päivitty uuden tiedostonimen mukaisiksi, sillä indeksit on jo kirjoitettu – lineaarisesti – nauhalle tiedostojen kirjoitusvaiheessa.

Tiedostojen validointi

PAS-paketointityökalu ja PAS-palvelun vastaanotto validoivat tiedostot siirtopaketteja luodessa ja edelleen PAS-palveluun lähettäessä, ks. "Aineiston hallinta Musiikkiarkistossa". LTF5-nauhavarmennus koskee lähtökohtaisesti vain PAS-palveluun lähetettävää aineistoa, joten aineistojen validointi voidaan ohittaa nauhavarmennusvaiheessa.

Ylimääräisten tiedostojen poistaminen

Kaikki piilotiedostot sekä -hakemistot tarkastetaan ja tarvittaessa poistetaan *Päätteessä*. Komen-toja suoritettaessa on huomioitava, että lähtökohtaisesti kirjainkoko on merkitsevä. Luotavien tiedostojen nimissä merkkijono "__db0", että tiedoston sisältöä ei kopioida Skaalaan PAS-kirjanpitoon pysyvästi säilytettäväksi. Merkkijono "__db1" puolestaan, että tiedoston sisältö talletetaan tietokantaan pysyvästi säilytettäväksi.

Poistetaan käyttöjärjestelmien luomat ylimääräiset tiedostot, esimerkiksi

- Macintosh-käyttöjärjestelmien luomat tiedostot: .DS_Store
 - Lisäksi tarkastetaan onko hakemistoissa muita piilotiedostoja, joiden tiedostonimi alkaa pisteellä, ja poistetaan ne harkinnan mukaan
- Windows-käyttöjärjestelmien luomat tiedostot: Thumbs.db / THUMBS.DB, Desktop.ini / DESKTOP.INI

Lisäksi eri ohjelmistojen pitkäaikaissäilytyksen kannalta turhia tiedostoja voidaan poistaa, esimerkiksi Sound Forgen luomat SFK-tiedostot (.sfk).

Täydellinen hakemistoindeksi luodaan komennolla, jolla nähdään tiedostoihin liittyvät UNIX-kaltaisten-käyttöjärjestelmien luomat laajennusattribuutit (*extended attributes*). Ylimääräiset tiedostot on jo poistettu – niitä ei dokumentoida laajennusattribuuttien tavoin. Attribuutit saattavat sisältää

esimerkiksi valokuvien esikatselukuvan, tiedoston latausosoitteen internetistä, *Logic Pro* -projektitiedoston nimen sekä tiedostolle tai hakemistolle *Finderissa* annetun värikoodin. Laajennusattribuutit poistetaan indeksin talteen ottamisen jälkeen. Mikään ei toistaiseksi (20210311) viittaa siihen, että attribuutit poistamalla menetettäisiin mitään oleellista metatietoa. Joka tapauksessa osa attribuuteista häviää nauhavarmennuksessa ja myös tiedostoja siirrettäessä eri tiedostojärjestelmien välillä, sillä eri järjestelmien käyttämät teknologiat tukevat attribuutteja eri tavoin.

Objektien metatietojen ja nauhaindeksien luominen

Seuraavaksi luotavat XLM-formaatin metatiedot ovat pysyvästi säilytettäviä, eivätkä sisällä tietoja laajennusattribuuteista. Tiedot talletetaan tietokantaan – mikäli tiedostonimessä on merkkijono ”__db1” –, josta ne ovat mahdollisimman helposti haettavissa.

Metatiedot otetaan talteen *Päätteessä ExifTool*-ohjelmaa käyttäen. Metatietotiedoston luonti voi kestää 1–10 minuuttia.

Nauhojen tiedostoindeksit sekä tiivistetiedostot luodaan seuraavaksi. Tiivisteiden laskukomento luo ensin SHA-1-tiivistetiedoston ja sen jälkeen MD5-tiivistetiedoston. Yhden tiivistetiedoston luonti voi kestää 4–10 tuntia – joskus jopa parikymmentä tuntia.

Tiivisteiden mukana talletetaan tietokantaan toiminnon suoritus-aika: ensimmäinen laskettu tiiviste on tiedoston alkuperäinen tiiviste, johon myöhemmin laskettuja tiivisteitä verrataan. Tiivisteitä voidaan laskea aineistosta myös satunnaisotannalla, ja tallettaa niitä tietokantaan milloin vaan, jolloin voidaan seurata ja dokumentoida objektien eheyttä.

Nauhan kirjoittaminen ja palautustarkistus

Nauhatoiminnot suoritetaan *HPE StoreOpen* -ohjelmaa sekä *Finderia* käyttäen. Nauha alustetaan, nimetään ja liitetään tietokoneeseen *mount*-komennolla *HPE StoreOpenilla* ja *Finderin* ikkunoissa yksinkertaisesti siirretään sijainnista toiseen kirjoitettava/palautettava hakemisto tavallisen tiedostosiirron tavoin. Toimintojen lopuksi nauha poistetaan koneesta *unmount*-komennolla.

Jokaisen nauhaoperaation jälkeen *HPE StoreOpenin* luomat sessiotiedot otetaan talteen *Diagnostics*-välilehdeltä ja talletetaan lokitiedostoon ja myöhemmin tietokantaan. Sessiotietoihin liitetään kopioidun hakemiston perustiedot, jotka otetaan talteen *Finderin* Info-toiminnolla: hakemiston tavukoko sekä kohteiden määrä. Tässä vaiheessa nauhahakemiston koko saattaa poiketa alkuperäisestä: piilotiedostot ja laajennusattribuutit, jotka on poistettu nauhahakemiston luomisen jälkeen ovat mahdollisesti vaikuttaneet alkuperäisen hakemiston kokoon. Juuri ennen nauhalle kopiointia otettujen hakemistotietojen tulee kuitenkin jatkossa vastata nauhalta kopiointin jälkeen otettujen hakemistotietojen kanssa. Luonnollisesti myös nauhalle kopioitujen sekä nauhalta kopioitujen tiedostojen tiivisteiden tulee vastata toisiaan.

- Tiedostonimi: `{kansio-/nauhanimi}---logFull__db1__$$(date + "%Y-%m-%d").txt`
 - *Päätteen* date-komento on vain esimerkkinä, tiedosto on luotava ja nimettävä käsin esimerkiksi *Texturilla*

Varmennuskirjoituksen jälkeen nauha poistetaan laitteesta ja sen fyysinen kirjoitussuoja laitetaan päälle.

Palautustarkistus tehdään mahdollisimman pian kirjoittamisen jälkeen. Nauha liitetään jälleen tietokoneeseen *mount*-komennolla *HPE StoreOpenilla* ja suoritetaan tiedostosiirto nauhalta kovalevylle: Palautuksen voi kirjoittaa samalle kovalevylle, kuin miltä aineisto on nauhalle kirjoitettu vain, jos levyllä on tilaa. Nauhalle kirjoitettua aineistohakemistota ei poisteta ennen kuin palautettu aineisto on tarkastettu.

Kopiointitoiminnon tavoin otetaan *HPE StoreOpenin* luomat sessiotiedot talteen ja talletetaan aikaisemmin luotuun lokitiedostoon ja myöhemmin tietokantaan *Finderilla* talteen otettujen hakemistotietojen kanssa.

Palautetuista tiedostoista lasketaan jälleen tiivisteet, joita verrataan ennen varmennusta tehtyihin tietoihin. Tiivisteet luodaan *Päätteessä* komennolla, joka luo ensin SHA-1-tiivistetiedoston ja sen jälkeen MD5-tiivistetiedoston. Ennen ja jälkeen nauhavarmennuksen lasketut tiivisteet on verratta-

va toisiinsa suorittamalla tiivisteiden tietokantatarkistus, ks. "Aineiston LTFS-nauhavarmennus Musiikkiarkistossa - Liite 4 - Tiivistetiedostojen tietokantatarkistus" (Musiikkiarkiston sisäinen dokumentti). Varmuuskopiointi on onnistunut, mikäli tiivisteet vastaavat toisiaan.

Tietokantaan talletetaan ennen nauhakihoitusta luodut MD5- ja SHA-1-tiivisteet. Palautustarkistusta varten luotuja tiivisteitä ei lisätä tietokantaan pysyvään säilytykseen. Tiivisteiden on joka tapauksessa vastattava toisiaan, jotta varmuuskopiointi voidaan hyväksyä, joten toisteisten tiivisteiden tallentaminen tietokantaan tässä vaiheessa on turhaa. Mikäli varmuuskopiointi on epäonnistunut, tilanne korjataan tapauskohtaisesti. Lähtökohtaisesti mikään toiminto ei saa epäonnistua pysyvästi säilytettäviä objekteja luotaessa, sillä tässä vaiheessa luodaan PAS-kirjanpidon lähtökohta: säilytettävien tiedostojen ensimmäinen versio.

Mikäli varmennus on onnistunut, kopioidaan nauhalle varmuuskopioitu tai palautustarkistettu aineistohakemisto Musiikkiarkiston verkkolevyille RAID-levypakkaan. Musiikkiarkistossa säilytetään pysyvästi säilytettävästä sähköisestä aineistosta yksi kovalevykopio nauhakopion lisäksi. Varsinainen sähköisen aineiston pitkäaikaissäilytys on kansallisen Kulttuuriperintö-PAS-palvelun vastuulla sen jälkeen, kun aineisto on hyväksytysti lähetetty PAS-palveluun, ks. dokumentti "Sähköisen aineiston hallinta Musiikkiarkistossa".

Varmennusnauhat virkistetään 5–10 vuoden välein, joskaan tämä ei ole välttämätöntä PAS-palvelussa säilytettävänä olevan aineiston kohdalla. Aineiston päivittämisen sekä tiedostomigraatioiden työnkulut ja käytänteet luodaan ja dokumentoidaan myöhemmin niiden tullessa ajankohtaisiksi. Migraatiot voi suorittaa itsenäisesti joko Musiikkiarkistossa tai yhteistyössä PAS-palvelun kanssa.

Liitteet

Liite 1. Varmuuskopioinnin työnkulun yleinen kuvaus

Objektien hallinta pitkäaikaissäilytykseen liittyvine toimintoineen, ks. dokumentti "Aineiston hallinta Musiikkiarkistossa".

- Valitse nauha ja anna sille käyttöönoton yhteydessä *id*
 - *Id* pidetään muuttumattomana koko nauhan elinkaaren ajan
- Alusta nauha ja nimeä se: $\{nauhanimi\} = \{teknologia+id\}$
 - Nauhalle ei anneta Sarjanumeroa tai viivakoodia. Musiikkiarkiston nauhat eivät tue viivakoodeja
- Luo varmuuskopioitavalle sähköiselle aineistolle hakemisto, jonka nimi $\{kansionimi\}$ vastaa nauhan nimeä $\{nauhanimi\}$
- Kopioi sähköinen aineisto nauhahakemistoon (max 2,375 Tt)
 - Säilytä hakemistohierarkia, mikäli se on käytännöllistä
- Tarkista hakemistohierarkia ja objektien nimeäminen
- Validoi tiedostot, mikäli mahdollista
- Poista ylimääräiset tiedostot
- Tarkasta tiedostojen laajennusattribuutit (*extended attributes*)
 - Luo täydellinen hakemistoindeksi (ei talleteta pysyvästi tietokantaan)
 - (Luo laajennusattribuuttien hakemistoindeksi (ei talleteta pysyvästi tietokantaan))
 - (Luo täydelliset metatiedot XML-tiedostoon (ei talleteta pysyvästi tietokantaan))
 - Poista tiedostojen laajennusattribuutit
- Ota objektien metadata talteen ja talleta ne tietokantaan
- Luo nauhahakemiston indeksit ja talleta ne tietokantaan
- Laske objektien MD5- ja SHA-1-tiivisteet ja talleta ne tietokantaan
- Kopioi nauhahakemisto nauhalle ja talleta kirjoitetun hakemiston sekä nauhaoperaation tiedot tietokantaan
- Poista nauha nauha-asemasta ja laita sen fyysinen kirjoitussuojaus päälle
- Tarkista varmuuskopioidut objektit
 - Palauta nauhan sisältö kovalevylle ja tallenna palautetun hakemiston sekä nauhaoperaation tiedot tietokantaan
 - Poista nauha nauha-asemasta ja arkistoi se vakaisiin säilytysolosuhteisiin pystyasennossa
 - Laske palautettujen objektien MD5- ja SHA-1-tiivisteet (ei talleteta pysyvästi tietokantaan)
 - Vertaa ennen kirjoitusta laskettuihin tiivisteisiin
 - Varmuuskopiointi on onnistunut, mikäli tiivistetiedostot vastaavat toisiaan
 - Ks. "Aineiston LTFS-nauhavarmennus Musiikkiarkistossa - Liite 4 - Tiivistetiedostojen tietokantatarkistus" (Musiikkiarkiston sisäinen dokumentti)
 - Luo tarvittaessa palautetun nauhahakemiston indeksit ja vertaa niitä ennen kirjoitusta luotuihin indekseihin

Liite 2. LTFS-toimintojen lokitemplate. (Musiikkiarkiston sisäinen dokumentti)

Erillisessä tekstitiedostossa:

- Aineiston LTFS-nauhavarmennus Musiikkiarkistossa - Liite 2 - LTFS-toimintojen lokitemplate.txt

Liite 3. LTFS-varmennuksen työnkulku Pääte-komentoineen (Musiikkiarkiston sisäinen dokumentti)

Erillisessä tekstitiedostossa:

- Aineiston LTFS-nauhavarmennus Musiikkiarkistossa - Liite 3 - LTFS-varmennuksen työnkulku Pääte-komentoineen.txt

Sisältää myös aineiston käsittelyosuuden, jonka toimintoja voidaan suorittaa nauhavarmennuksesta riippumatta.

Liite 4. Tiivistetiedostojen tietokantatarkistus (Musiikkiarkiston sisäinen dokumentti)

Erillisessä tekstitiedostossa:

- Aineiston LTFS-nauhavarmennus Musiikkiarkistossa - Liite 4 - Tiivistetiedostojen tietokantatarkistus.mysql

Liite 5. Koodistot (Musiikkiarkiston sisäinen dokumentti)

Erillisessä tekstitiedostossa:

- Aineiston LTFS-nauhavarmennus Musiikkiarkistossa - Liite 5 - Koodistot.txt

Liite 6. Huomiot tekstissä mainituista dokumenteista

YoYotta ja LTFS-dokumentit ovat verkkolevyllä tallessa. YoYotta-dokumenteissa on paljon hyvää tietoa, kunhan pitää mielessä, että ne on tarkoitettu YoYottan oman ohjelmiston käyttämisen tueksi.

Laitteiden ja ohjelmistojen asennustiedostot sekä niihin liittyvät ohjeet ovat myös verkkolevyllä tallessa.

Lisäksi tallessa on sekalaisia nettitulosteita esimerkiksi laajennusattribuuteista (*extended attributes*) ja LTO-/LTFS-teknologioihin liittyvistä artikkeleista.