

# **Aineiston hallinta 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 discusses the management of the resources permanently stored in the Music Archive Finland, by using the operations related to the Digital Preservation Service (DPS) maintained by the IT Center for Science (CSC). The processed resources (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).*

This document presents the general principles of the processes. The exact workflows of the processes, along with the operations to be performed, are described in the appendices.

At first, the two main functions of the Skaala database maintained by the Music Archive Finland are presented:

- Preparing the resources for long-term storage and recording related storage procedures (PRE-MIS events)
- Sending the resources to the DPS service and recording storage procedures (PREMIS events) for the resources

Thanks to the Skaala database, the Music Archive Finland has all the information about all the objects located in the DPS service. This information is independent of the interfaces of the DPS service.

At Skaala, objects are assigned unique identifiers (UUID, Universally Unique Identifier) at the beginning of their life cycle, and the integrity of the objects is monitored using MD5 and SHA-1 hashes. Objects are sent to the DPS service as Submission Information Packages (SIP) via SFTP interface. The packages are created using the tools created by CSC (Pre-Ingest Tool for creating Submission information packages). Metadata crucial to the long-term preservation is exported from Skaala and is included in the SIP package in two file formats:

- One or more XML files contain the descriptive and technical metadata of the resources and the SIP package.
- One CSV file contains a list of objects alongside a predefined set of administrative and technical metadata about the objects.

The DPS service assigns a unique identifier to each transfer which is recorded in Skaala. If the SIP is accepted for storage, an Archival Information Package (AIP) is formed, and the DPS service is officially responsible of the preservation of the objects. The AIP is also given a unique identifier, which is recorded in Skaala. In addition, detailed reports of the ingest process created by the DPS service are stored in the Music Archive Finland's electronic archive.

The DPS service monitors the technical development of the file formats it supports, and it may in the future recommend migrating a file format to an up-to-date format. If the migration is carried out – with the authorization by the Music Archive Finland – a PREMIS event will be recorded into Skaala, so that up-to-date information about the objects in the DPS can always be checked from Skaala.

The objects located in the DPS service can be retrieved by the Music Archive Finland via the REST interface provided by the service. The DPS service creates a Dissemination Information Package (DIP) based on the parameters determined by the Music Archive Finland. For example, the identifier of the transfer package, the name of the object (file name), the MD5 hash, the UUID of the object, or the free text sent in the descriptive XML metadata file, can be used as parameters.

# Sisällys

<b>1 Sisällys</b>	<b>3</b>
<b>2 Johdanto</b>	<b>4</b>
<b>3 Musiikkiarkiston sähköiset järjestelmät</b>	<b>4</b>
3.1 Musiikkiarkiston Skaala-tietokanta	4
3.2 Musiikkiarkiston sähköinen säilytysjärjestelmä	5
<b>4 Objektien käsitteleminen pitkäaikaissäilytystä varten</b>	<b>6</b>
4.1 Objektien tuominen Skaalan PAS-osioon	6
4.2 SIP-toiminnot: objektien vieminen Kulttuuriperintö-PAS-palveluun	6
4.2.1 SIP-hakemiston rakenne	6
4.2.2 Objektien kopioiminen SIP-hakemistoon	7
4.2.3 SIP:n paketointi ja lähettäminen PAS-palveluun	7
4.2.3.1 PAS-palvelun hyväksymä SIP	8
4.2.3.2 PAS-palvelun hylkäämä SIP	8
4.3 PAS-palvelun REST-rajapinnan toiminnot	9
4.4 PAS-palvelun säilytysvastuulla olevien objektien päivittäminen tai poistaminen	9
<b>5 Liitteet</b>	<b>10</b>
5.1 Liite 1. PAS-dokumentit ja -työkalut sekä PAS-palvelun käyttäminen	10
5.1.1 PAS-paketointiympäristön asetukset	10
5.1.1.1 SIP:n allekirjoittamisen vaatimat tiedostot	10
5.1.1.2 SSH-yhteyden tiedostot	11
5.1.1.3 Sip-compiler-konfigurointitiedosto	11
5.1.2 PAS-palvelun käyttämisen vaatimukset (Musiikkiarkiston sisäinen dokumentti)	11
5.2 Liite 2. Skaalan PAS-osion luomisen SQL-komennot (Musiikkiarkiston sisäinen dokumentti)	11
5.3 Liite 3. PAS-objektien luominen Skaalaan (Musiikkiarkiston sisäinen dokumentti)	11
5.4 Liite 4. PAS-objektien ylläpito Skaalassa, PREMIS-tapahtumat (Musiikkiarkiston sisäinen dokumentti)	11
5.5 Liite 5. SIP-toiminnot (Musiikkiarkiston sisäiset dokumentit)	12
5.5.1 SIP:n luominen	12
5.5.2 SIP:n lähettäminen	12
5.5.3 SIP:n tai AIP:n hakeminen sekä hallinnointi	12
5.6 Liite 6. SQL-kyselyt (Musiikkiarkiston sisäinen dokumentti)	12

# Johdanto

Musiikkiarkiston aineiston hallintaan liittyvät dokumentit:

- 1 Aineiston järjestäminen ja nimeäminen Musiikkiarkistossa
- 2 Aineiston LTFs-nauhavarmennus Musiikkiarkistossa
- 3 Aineiston hallinta Musiikkiarkistossa

Tässä dokumentissa käsitellään Musiikkiarkistossa pysyvästi säilytettävän aineiston hallinta Kulttuuriperintö-PAS-palveluun liittyvine toimintoineen. Aineistolla tarkoitetaan digitaalisessa muodossa olevia objekteja (tiedostoja), jollei toisin mainita.

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

## Musiikkiarkiston sähköiset järjestelmät

Aineistoa hallinnoidaan Musiikkiarkiston Skaala-tietokannassa ja säilytetään aktiivisessa käytössä olevilla verkkolevyillä. Pääsy verkkolevyille on rajattu Musiikkiarkiston sisäverkkoon.

## Musiikkiarkiston Skaala-tietokanta

Musiikkiarkiston Skaala-tietokannan PAS-osiossa ("Skaala") hallinnoidaan Musiikkiarkistossa pysyvästi säilytettäviä digitaalisia objekteja ("tiedosto", "objekti"). Säilytyksen varmistamiseksi objekteihin kohdistetaan elinkaartensa aikana toimenpiteitä, jotka kirjataan Skaalaan PREMIS-tapahtumina (*event*). Tavoitteena on siirtää objektit Tieteen tietotekniikan keskus CSC:n ylläpitämään pitkäaikaissäilytysjärjestelmään Kulttuuriperintö-PAS-palveluun ("PAS-palvelu").

Objektien hallinnointi Skaalassa jakautuu kahteen pääosaan:

- 1 Hallinnoidaan objekteja Musiikkiarkistossa ja valmistellaan ne PAS-palveluun viemistä varten
  - Annetaan objekteille ainutlaatuiset tunnisteet
  - Huolehditaan objektien eheydestä PAS-palveluun viemiseen asti
    - Skaalaan kirjattava PREMIS-tapahtuma: *message digest calculation*
  - Hallinnoidaan objekteja, esimerkiksi
    - Tiedostonimiä voidaan perustellusta syystä muokata koska tahansa
      - Skaalaan kirjattava PREMIS-tapahtuma: *filename change*
    - Vanhentuneet tiedostomuodot migroidaan tarvittaessa teknisesti ajantasaisiin muotoihin
      - Skaalaan kirjattava PREMIS-tapahtuma: *migration*
    - Aineistoa poistetaan säilytyksestä
      - Skaalaan kirjattava PREMIS-tapahtuma: *deletion*
  - Pidetään kirjaa objektien mahdollisesta LTFs-nauhasijainnista (offline-varmuuskopio)
    - Objektien sijaintia Musiikkiarkiston levyjärjestelmissä ei kirjata Skaalaan
- 2 Huolehditaan objektien lähettämisestä PAS-palveluun ja pidetään kirjaa PAS-palvelussa olevista objekteista
  - Objektit lähetetään Kulttuuriperintö-PAS-palveluun siirtopaketteina (SIP, *Submission Information Package*)
    - Skaalaan kirjattavat PREMIS-tapahtumat: *information package creation, transfer*
  - Pidetään kirjaa SIP:ien sisältämisestä objekteista, SIP:eihin kohdistuneista toimenpiteistä sekä Kulttuuriperintö-PAS-palvelun vastauksista (*response*) SIP:eihin liittyvistä toiminnoista
  - PAS-palvelu voi hyväksyä tai hylätä SIP:n
    - PAS-palvelun hyväksytyä SIP:n objektien säilytysvastuu on siirtynyt PAS-palvelulle, joka on muodostanut SIP:stä säilytyspaketin (AIP, *Archival Information Package*)

- PAS-palvelun hylättyä SIP:n suoritetaan tarvittavat toimenpiteet, jotta pysyvästi säilytettävät objektit saadaan hyväksytysti lähetettyä PAS-palveluun

Mikäli PAS-palvelun säilytysvastuulla oleviin objekteihin kohdistetaan PREMIS-tapahtumia, hoidaan säilytyksessä olevien objektien päivittäminen tapauskohtaisesti.

PAS-palveluun voitaisiin myös lähettää objekteja säilytettäväksi ilman niiden tallentamista Skaalaan. Tällöin objektien säilyminen olisi turvattu, mutta Musiikkiarkiston objektien hallinta olisi täysin riippuvainen PAS-palvelun rajapinnan tarjoamista ominaisuuksista. Esimerkiksi tiedostonimen muuttaminen aineiston uudelleenjärjestäminen yhteydessä johtaa tilanteeseen, jossa PAS-palvelussa oleva aineisto ei objektien nimiltä vastaa Musiikkiarkistossa olevaa aineistoa, mikäli muutoksista ei ole pidetty kirjaa, ja muutoksia päivitetty PAS-palveluun. PAS-palvelu on tältä osin alisteinen Skaalalle: kaikki objektien säilytykseen liittyvät päätökset tehdään Musiikkiarkistossa – itsenäisesti tai yhteistyössä CSC:n kanssa – ja suoritettujen toimenpiteiden kirjanpito tapahtuu soveltuvien osin Skaalassa.

## Musiikkiarkiston sähköinen säilytysjärjestelmä

Nauhavarmennuksen työnkulkua on tulevaisuudessa sovellettava siltä osin, kun objektit viedään Skaalaan ennen nauhavarmennusta. Nauhavarmennuksesta mahdollisesti myös luovutaan jossain vaiheessa, sillä tulevaisuudessa pysyvästi säilytettävien objektien säilytysvastuu pyritään siirtämään PAS-palvelulle heti, kun mahdollista. Tällöin objektien digitointi- ja vastaanottoprosesseissa ei tarvitse kerätä aineistoa LTFS-nauhakapasiteetin verran, vaan aineisto voidaan siirtää verkkolevyllä suoraan SIP-työjonohakemistoon odottamaan PAS-palveluun lähettämistä varten. Musiikkiarkistossa käytettävien LTO-6-nauhojen tallennuskapasiteetti on noin 2,375 Tt.

Mahdollisen LTFS-nauhavarmennuksen lisäksi pysyvästi säilytettävät objektit säilytetään myös RAID-varmennetuilla verkkolevyillä arkistonmuodostusta vastaavissa hakemistoissaan. Verkkolevyjen aineistohakemistot sijaitsevat päähakemistoissa, jotka on jaettu kolmeen kategoriaan:

**1** Objektit, joiden säilytysvastuu on PAS-palvelulla

**2** SIP-työjonossa sijaitsevat objektit, joita ollaan käsittelemässä PAS-palveluun lähettämistä varten, tai jotka on lähetetty PAS-palveluun

- PAS-palvelun hyväksyttyä SIP:n, aineisto siirretään kategorian 1 aineistohakemistoihin
- *HUOM! Työjonon varmuuskopiointista on huolehdittava erikseen*

**3** Objektit, joita ei ole lähetetty PAS-palveluun

- *HUOM! Varmuuskopiointista on huolehdittava erikseen, varsinkin tilanteissa, joissa objektit viedään Skaalaan välittömästi digitoinnin tai vastaanoton jälkeen odottamaan SIP:ien muodostamista*

Edellä oleva järjestely mahdollistaa SIP:ien luomisen sokkona sopivin aineistokokonaisuuksin, sillä objektien järjestyksen ollessa ajantasainen, ei ole vaaraa, että samaa aineistoa lähetettäisiin PAS-palveluun toistamiseen.

*HUOM! Mikäli objekteihin kohdistetaan PREMIS-tapahtumia, kuten migraatio, nimimuutos, tai poistaminen, on objektit välittömästi siirrettävä pois kansioista, joissa ne sijaitsevat aktiivisessa säilytysjärjestelmässä. On myös mietittävä, mihin ne siirretään ja miten päivitykset hoidetaan suhteessa PAS-palveluun ja LTFS-nauhavarmennukseen: SIP:ien päivittäminen, poistaminen, korvaaminen, LTFS-nauhojen korvaaminen, virkistys yms. Korjaavien toimenpiteiden jälkeen siirretään objektit jälleen aktiivisessa säilytysjärjestelmässä niiden statusta vastaaviin hakemistoihin.*

# Objektien käsitteleminen pitkäaikaissäilytystä varten

Objektit on ensin kirjattava Skaalaan PAS-kirjanpitoon, jotta ne voidaan lähettää PAS-palvelun säilytettäväksi.

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

## Objektien tuominen Skaalan PAS-osioon

Objektit tuodaan Skaalaan tiivistetiedostoja käyttäen, jotka on luotu joko LTFS-nauhavarmennuksen yhteydessä tai yksittäisenä erillisenä toimintona. Tiedostonimien tarkastaminen suoritettava erikseen ennen PAS-objektien luomista osana tiivisteiden laskemisen työnkulkua, jos nauhavarmennusta ei ole tehty, ja tiedostoja ei ole käsitelty perusteellisesti aineiston vastaanotossa tai järjestämisessä, tai analogisen aineiston digitoinnin yhteydessä.

Tuotaessa objektit Skaalaan on niille tallennettava vähintään yksi tiiviste, jotta objektien eheyden seuraaminen olisi mahdollista. Skaalassa on tuettuina MD5- ja SHA-1-tiivisteet. PAS-palvelu hyödyntää vastaanotossa vain MD5-tiivisteitä. Nauhavarmennuksen ohjeet, ks. dokumentti "Aineiston LTFS-nauhavarmennus Musiikkiarkistossa".

Tiivistetiedostoista otetaan lisäksi talteen tiedoston nimi, joka on Musiikkiarkistossa objektin ensisijainen ainutlaatuinen tunniste. Tuomisen yhteydessä tarkastetaan muun muassa tiedostonimen ainutlaatuisuus suhteessa muihin säilytettäviin tiedostoihin ja annetaan tuotaville objekteille pysyvät tunnistet (UUID, UUID-SHORT). Tiedostonimi voidaan jatkossa muuttaa – kunhan se pysyy ainutlaatuisena suhteessa muiden objektien tiedostonimiin ja muuttamisesta talletetaan PREMIS-tapahtuma Skaalaan. Objektien pysyvät tunnistet eivät muutu koskaan, ja myös niiden tulee olla ainutlaatuisia suhteessa muiden objektien tunnistetisiin.

Ennen objektien tuomista Skaalaan on suotavaa myös suorittaa objektien validointi, jotta tiedostojen yhteensopivuus KP-PAS-Palvelun kanssa olisi mahdollisimman aikaisessa vaiheessa tiedossa. Ks. Liite 5. SIP-toiminnot (Musiikkiarkiston sisäiset dokumentit).

## SIP-toiminnot: objektien vieminen Kulttuuriperintö-PAS-palveluun

Objektit viedään PAS-palveluun siirtopaketteina (SIP, *Submission Information Package*). Aluksi luodaan SIP-hakemisto, johon kopioidaan aineisto sekä luodaan paketoinnin vaatimat tiedostot paketoitua varten. Paketoinnin jälkeen SIP(-siirtopaketti) lähetetään PAS-palvelun vastaanoton käsiteltäväksi.

SIP:ien sisältö sekä SIP:eihin kohdistetut toiminnot kirjataan Skaalaan, ja lopulta PAS-palvelun vastaukset SIP-toimintoihin talletetaan Skaalaan.

## SIP-hakemiston rakenne

SIP-hakemisto nimetään luontipäivän mukaan, jonka perään lisätään juoksevan numeron tavoin käyttäytyvä suuraakkonen: mikä paketoitilinjasto on kyseessä, sekä juokseva numero: kuinka mones kyseisen paketoitilinjaston SIP-hakemisto on kyseessä saman luontipäivän sisällä. SIP-hakemistonimen muoto on "<vvvkkpp>-<vuo>-<nro>", esimerkiksi "20211010-A-1" ja "20211010-A-2". Mikäli samana päivänä olisi tehty SIP:jä kahdella tai useammalla paketoitilinjastolla, annetaan eri paketoitilinjastoille yksilölliset vuotunnukset, esimerkiksi "20221212-A-1", "20221212-A-2" ja "20221212-B-1".

SIP-hakemisto sisältää yhden alihakemiston (*Contents*), johon PAS-palveluun lähetettävä aineisto kopioidaan. Aineistohakemisto voi sisältää yhden tai useamman hakemiston tai objektin. Aineistohakemiston rinnalla, SIP-hakemiston juuressa, on oltava vähintään kaksi metadatatiedostoa: kuvailevan metadatan XML-tiedosto sekä PAS-vastaanoton vaatimat objektikohtaiset tiedot sisältävä CSV-tiedosto. Molemmat tiedostonimet nimetään alkaviksi SIP-hakemistonimellä ja ne päättyvät merkkijonoon "\_\_metadata.xml" ja "\_\_metadata.csv". XML-tiedostoja voi olla yksi tai useampi, joten XML-tiedoston SIP-hakemistonimeä vastaavan osuuden perään lisätään tarvittaessa yhdysviiva ja juokseva numero.

- <vvvkkpp>-<vuo>-<nro>-<[-nro]>\_\_metadata.xml
  - Kuvailevan metadatan tiedosto(je)n tulee olla PAS-palvelun tukemassa formaatissa, esimerkiksi MARC21, Dublin Core MODS tai EAD3
- <vvvkkpp>-<vuo>-<nro>\_\_metadata.csv
  - Sisältää esimerkiksi tiedostonimen, pysyvän tunnisteiden, MD5-tiivisteen sekä objekteihin kohdistetut PREMIS-tapahtumat
  - SIP-tunnukseksi määritetään PAS-vastaanotossa merkkijonoa "\_\_metadata.csv" edeltävä osuus, ja se vastaa täten SIP-hakemiston nimeä
    - SIP-tunnus on ainutlaatuinen tunniste, jonka avulla Skaalassa pidetään kirjaa SIP:ien sisältämistä objekteista ja niiden statuksista sekä SIP:eihin kohdistuneista toiminnoista
    - PAS-palvelu antaa lisäksi jokaiselle SIP:n siirrolle sekä mahdollisesti muodostettavalle säilytyspaketille (AIP, *Archival Information Package*) ainutlaatuiset tunnisteet, jotka talletetaan Skaalaan

Metadatatiedostot eivät ole pysyvästi säilytettäviä objekteja, eivätkä ne täten sellaisenaan sijaitse Skaalassa eivätkä PAS-palvelussa; Paketoinnin yhteydessä niiden sisältämät tiedot talletetaan PAS-palveluun METS-formaatissa, ja niiden sisältö on haettavissa PAS-palvelun REST-rajapinnan kautta, ks. "PAS-palvelun REST-rajapinnan toiminnot".

Kuvailevan metadatan tiedostot ovat hyvin yksinkertaisia vain vähän tietoa sisältäviä tiedostoja. Ainakin niin kauan, kunnes Musiikkiarkiston käyttämien kuvailevan metadatan järjestelmistä on mahdollisuus helposti liittää tiedot osaksi SIP:ejä. Tällöinkin on pidettävä mielessä, että kuvaileva metadata voi aika ajoin päivittyä, joten lähtökohtaisesti PAS-palvelu sisältää vain minim tiedot aineistosta, esimerkiksi tunnisteiden, jonka avulla ajantasaiset metatiedot voidaan hakea käytössä olevasta metatietojärjestelmästä.

## Objektien kopioiminen SIP-hakemistoon

PAS-palveluun lähetettävät tiedostot kopioidaan Musiikkiarkiston sähköisestä arkistosta, jossa sijaitsee vain Skaalassa olemassa olevia pysyvästi säilytettäviä objekteja. SIP-hakemistot kasaataan käsin ja niiden sisältämät objektit valitaan tilanteen mukaan sopiva aineistokokonaisuus, tai sen osa, kerrallaan.

PAS-palvelun vastaanotolla on rajoituksia siirtopaketteihin liittyen. Tämän hetkisen (20211110) tiedon mukaan SIP:ejä ei oteta PAS-palvelun vastaanotossa käsittelyyn mikäli yhden SIP:n koko ylittää 1,7 Tt tai se sisältää yli 50 000 objektia.

Sähköisessä arkistossa voidaan säilyttää PAS-palveluun lähetettyjä ja sinne säilytykseen hyväksytyjä objekteja eri kansioissa kuin PAS-palveluun lähetettävistä odottavista objekteista. Joka tapauksessa SIP:ien tiedot niiden sisältämien objekteineen kirjataan Skaalaan, jonne lopulta kirjataan myös tiedot PAS-palveluun hyväksytyistä objekteista; jokaisen objektin status PAS-palvelussa on myös Musiikkiarkistossa aina tarkastettavissa.

## SIP:n paketointi ja lähettäminen PAS-palveluun

SIP:n paketointi tehdään CSC:n kehittämällä paketointikomponenteilla: *Pre-Ingest Tool for creating submission information packages*. (GitHub:ssa Digital-Preservation-Finland: *dpres-siptools*, *dpres-sip-compiler*).

Paketoinnin jälkeen SIP lähetetään PAS-palvelun vastaanottoon tar- tai zip-muodossa SFTP-rajapinnan kautta. Jokaiselle siirrolle annetaan ainutlaatuinen tunniste, joka on muotoa: <SIP>-<UUID>.

PAS-palvelun vastaanotossa on neljä hakemistoa sopimustunnistetta – hyödyntävän organisaation sopimuskohtaista käyttäjätunnusta – kohden: *accepted*, *rejected*, *transfer* ja *disseminated*. Viides *approved*-hakemisto on erikoistapaus, jota käsitellään tarvittaessa erikseen. SIP kopioidaan *transfer*-hakemistoon ja PAS-vastaanotto ottaa paketin käsittelyyn välittömästi onnistuneen tiedosto-siirron jälkeen, mikäli käsittelyn aloittamista ei erikseen estetä nimeämällä paketteja ".part"-tunnuksin.

Jokaisesta siirrosta luodaan tarkastusraportti *rejected*- tai *accepted*-hakemistojen alihakemistoihin, jotka on nimetty siirtopäivän, siirron (SIP) sekä siirtotunnuksen (UUID) mukaan: *Rejected*-hakemistoon PAS-vastaanoton hylkäämät paketit tarkastusraporteineen. *Accepted*-hakemistoon PAS-vastaanoton hyväksymien pakettien tarkastusraportit. Jälkimmäisessä tapauksessa aineiston säilytysvastuu on siirtynyt PAS-palvelulle.

## **PAS-palvelun hyväksymä SIP**

Säilytysvastuun siirryttyä PAS-palvelulle SIP:n siirtäminen PAS-palveluun kirjataan onnistuneeksi ja PAS-palvelun muodostaman AIP:n sekä siirron tunnisteet talletetaan Skaalaan. Siirtopakettikohtaisen kirjanpidon ansiosta saadaan PAS-palvelun säilytysvastuulla oleva aineisto listattua objekti-kohtaisesti. Lisäksi PAS-vastaanoton luomat tarkastusraportit otetaan talteen *accepted*-hakemiston siirtokohtaisesta alihakemistosta ja talletetaan Musiikkiarkiston järjestelmiin.

## **PAS-palvelun hylkäämä SIP**

Hylättyjen pakettien kaikki PREMIS-tapahtumat – SIP:n luominen, lähettäminen sekä PAS-vastaanoton raporttien kirjaaminen – voidaan tapauskohtaisesti tallettaa Skaalaan tai jättää huomiotta. Oleellista lopulta on vain se, että jokainen Musiikkiarkiston pysyvästi säilytettävä objekti saadaan jossain vaiheessa lähetettyä hyväksytysti PAS-palveluun. *Skaalassa ei tarvitse säilyttää täydellistä toimintologia jokaisesta epäonnistuneesta siirtoyrytyksestä, mutta kaikissa tilanteissa on oltava täysin selvää, mitkä Musiikkiarkiston pysyvästi säilytettävät objektit ovat PAS-palvelun säilytysvastuulla.*

Lähtökohtaisesti jokainen objekti liitetään vain yhteen SIP:iin ja lähetetään PAS-palveluun vain kerran. PAS-palvelun hylätessä SIP:ejä, tilanteet korjataan tapauskohtaisesti, esimerkiksi päivittämällä hylätyt objektit PAS-palvelun määritysten mukaisiksi objekti migroimalla tai normalisoimalla. Tapahtuma, korvaavan sekä alkuperäisen korvattavan objektin suhde kirjataan Skaalaan PREMIS-tapahtumana ja SIP:n sisältö päivitetään uutta validointia ja edelleen PAS-palveluun lähettämistä varten.

SIP-tunnus voidaan pitää samana, mutta sen sisältämät tiedostot päivitetään tarvittaessa; tiedostonimet eivät välttämättä muutu, mutta korvaavat objektit saavat aina uudet pysyvät UUID- ja UUID-SHORT-tunnisteet, sillä jokaisen objektin – ja niiden versioiden – on oltava kaikissa tilanteissa yksilöitävissä. Vaihtoehtoisesti SIP voidaan kokonaisuudessaan myös hylätä, jolloin objektit on vietävä PAS-palveluun myöhemmin luotavissa SIP:eissä.

Korvatulle tiedostolle annetaan uusi nimi, jolloin ajantasaisen tiedoston nimi voidaan pitää samana, sillä lähtökohtaisesti objektin semanttinen suhde sen sisältävässä arkistossa ei muutu. Korvattun tiedostonimen perään lisätään päätte "\_\_\_migrated---vvvkkpp[-osaobjektitunnus]". Vanhan korvattun tiedoston nimellä ei ole väliä, jos

- jokaisesta migraatiosta kirjataan PREMIS-tapahtuma Skaalaan, ja
- korvattua objektia säilytetään eri hakemistossa kuin ajantasaista objektia.

Tällöin migraatiotapahtuman objektit ovat aina tunnistettavissa ja tapahtuman päivämäärä on kirjattu Skaalaan. Mutta mikäli vanhentuneen objektin tiedostonimeä ei muuteta, on migroituja tiedostoja käsitellessä objektin tunnisteina aina käytettävä UUID:eitä, UUID-SHORT:ejä tai Skaalan tietokantataulujen pää- ja viiteavaimia.

*HUOM! Korvatut objektit ovat edelleen aktiivisesti säilytettäviä objekteja aina poistamiseensa asti.*

*HUOM! Normalisointi on migraatiotapahtuman tarkennus. Kirjatun PREMIS-tapahtuman ollessa migration tai normalization, tiedostonimen perään lisätään sama päätte: "\_\_\_migrated---vvvkkpp[-osaobjektitunnus]".*

## PAS-palvelun REST-rajapinnan toiminnot

REST-rajapinnan kautta voidaan tarkastaa esimerkiksi säilytyksessä olevan aineiston määrä ja hakea aineistoa eri ehtoja käyttäen. Rajapinnan toiminnot kohdistuvat säilytykseen hyväksytyihin aineistoihin, AIP:eihin.

PAS-paketoinnin yhteydessä luodusta METS-XML-tiedostosta voidaan hakea minkä tahansa kentän mukaan, teknisen metadatan kohdalla esimerkiksi

- tiiviste: <premis:messageDigest>
- alkuperäinen tiedostonimi: <premis:originalName>
- formaatti: <premis:formatName>
- UUID-tunniste: <premis:objectIdentifierValue>
- Skaalan sisäinen objektin ID: <premis:objectIdentifierValue>

Teknisen metadatan osalta on huomattava, että Skaalan on talletettu säilytettävien objektien tiivistet ja UUID-tunnisteet. Skaalan PAS-objektien luomisvaiheessa annettu UUID-tunniste pysyy muuttumattomana PAS-vastaanotossa. Myös tiivisteen on pysyttävänä muuttumattomana, jotta PAS-vastaanotto ylipäättensä hyväksyy aineiston säilytykseen, joten näiltä osin REST-rajapinnan hakuehdot voidaan poimia Skaalasta.

Kuvailevan metadatan osalta voidaan METS-tiedoston kautta hakea tietoa, joka on lähetetty SIP:ssä kuvailevan metadatan XML-tiedostossa. Esimerkiksi Dublin Core -muotoisen metadatatiedoston hakutermit voivat olla esimerkiksi

- nimeke: <dc:title>
- tunniste: <dc:identifier>

PAS-Palveluun indeksoidaan METS:n sisältö ilman nimiavaruuksia: *premis*, *dc*. Haettaessa aineistoa voidaan hakuavaimena käyttää METS:n hierarkiarakenteen muotoista polkua tai esimerkiksi vain osiota *messageDigest* tai *identifier*, ks. Aineiston hallinta Musiikkiarkistossa - Liite 5c - PAS-palvelun REST-rajapinnan käyttäminen.

*HUOM! SIP:lle voi antaa vain yhden contentId:n, jos ei käytetä sip-compilerin compile-komentoa. Aineiston tunnisteet on kirjattava kuvailevaan metadataan sip-compileria käytettäessä.*

## PAS-palvelun säilytysvastuulla olevien objektien päivittäminen tai poistaminen

PAS-palvelun säilytysvastuulla olevan aineiston päivittämisen prosessit luodaan ja dokumentoidaan, kun ne tulevat ajantasaisiksi. Mahdollisia tilanteita ovat esimerkiksi

- aineiston uudelleen järjestäminen
  - Tiedostonimi on korvattava vastaamaan ajantasaista sarjanmuodostusta
- aineiston poistaminen
  - Musiikkiarkiston säilytysvastuulla olevaa aineistoa poistetaan säilytyksestä
- aineiston korvaaminen, migrointi
  - Objektin päivittäminen teknisesti ajantasaiseen tiedostomuotoon

Jokaisesta aineistoon liittyvästä elinkaaritapahtumasta (PREMIS-tapahtuma) tehdään yksityiskohmainen kirjaus, joka säilytetään pysyvästi Skaalassa. Näin ollen esimerkiksi aineistomigraatioissa on aina tiedossa, mikä on ollut alkuperäinen säilytettävä objekti, ja mitä toimenpiteitä alkuperäiseen objektiin tai siitä johdettuihin objekteihin on kohdistettu. Skaalaan tallennetut PREMIS-tapahtumat tallennetaan paketoituvaiheessa SIP:ien CSV-tiedostoihin, joten objektien elinkaaritiedot tallentuvat pysyvästi myös PAS-palveluun.

# Liitteet

## Liite 1. PAS-dokumentit ja -työkalut sekä PAS-palvelun käyttäminen

Tässä dokumentissa ei ole tarkkoja asennusohjeita, sillä CSC:n luomat PAS-työkalut on asennettu CSC:n henkilökunnan avustuksella CentOS 7 -käyttöjärjestelmäympäristöön vuonna 2021. Myös paketoituympäristön konfiguroinnin tulee luonnollisesti olla kohdillaan. Ajantasainen ohjeistus työkalujen asennukseen on alla mainituissa CSC:n hallinnoimissa verkkosijainneissa. Lisäksi Musiikkiarkiston sisäisissä dokumenteissa on perustiedot PAS-palvelun vaatimista asennustoimenpiteistä ja perusteelliset työkulkukuvaukset PAS-palvelun käyttämisestä.

- <https://digitalpreservation.fi/specifications>
  - <https://digitalpreservation.fi/specifications/metadata>
  - <https://digitalpreservation.fi/specifications/fileformats>
  - <https://digitalpreservation.fi/specifications/interfaces>
  - <https://www.digitalpreservation.fi/documents/looginensailyttaminen>
- <https://github.com/Digital-Preservation-Finland>
  - <https://github.com/Digital-Preservation-Finland/dpres-siptools>
    - Pääasiallinen paketoitikonponentti
  - <https://github.com/Digital-Preservation-Finland/dpres-sip-compiler>
    - Musiikkiarkiston käyttämä automatisoitu paketoitityökalu, joka käyttää dpres-siptools-komponenttia
      - SIP-hakemistossa on oltava mukana ennalta määrätyt tiedot, jotta paketointi olisi mahdollista
      - Vaatimukset on sovittu PAS-käyttöönotton yhteydessä CSC:n kanssa, ja ne on esitelty tarkemmin Liitteessä 5
  - <https://github.com/Digital-Preservation-Finland/dpres-ipt>
    - Ei testattu: Objektien tai METS-dokumentin paikallinen validointi (sisältyy paketoitityönkulkuun)
  - <https://github.com/Digital-Preservation-Finland/access-rest-api-client>
    - Asennettu: Yhteys rajapintaan SFTP-yhteydellä tai REST:llä
      - AIP:ien hakemista ja DIP:ien lataamista varten
      - REST-API SIP:ien lataamiselle CSC:llä työn alla (20221215)

### PAS-paketoituympäristön asetukset

Tarvitaan SIP:n allekirjoittamisen tiedostot ja SSH-yhteyden sekä sip-tools-compilerin asetustiedostot.

#### SIP:n allekirjoittamisen vaatimat tiedostot

Polku tiedostot sisältävään kansioon: `"/home/<user>/dpres-config/".`

Kotikansiossa kansio "dpres-config", jossa sijaitsevat tiedostot

- sip\_sign\_pas.pem
  - Allekirjoittaa SIP:n, ks. konfigurointitiedosto alla
- sip\_sign\_pas\_musiikkiarkisto.pub
  - Julkinen avain, joka on lähetetty CSC:lle käyttöönotton yhteydessä. Ei tarvita paketoituvaiheessa
- sign.sh
  - Vanha allekirjoitustapa tai avainparien luomistapa. CSC:n ohjeistus ei ajan tasalla (20221215)
  - Voi poistaa: sip-compiler tai siptools hoitaa allekirjoittamisen
  - Ei tuotantoympäristössä (20221215)

## SSH-yhteyden tiedostot

Polku tiedostot sisältävään kansioon: "/home/<user>/.ssh/".

Kotikansiossa kansio .ssh, jossa sijaitsevat tiedostot

- id\_rsa\_pastesti\_musiikkiarkisto
  - Sisäinen yksityinen avain, jota ei saa luovuttaa eteen päin tai näyttää kenellekään!
- id\_rsa\_pastesti\_musiikkiarkisto.pub
  - Julkinen avain, joka on lähetetty CSC:lle käyttöönoton yhteydessä. Ei tarvita yhteyttä ottaessa(?)
- known\_hosts
  - käyttöjärjestelmän tarvitsema tiedosto, jotta se ymmärtää host-nimen. Luodaan automaattisesti

## Sip-compiler-konfigurointitiedosto

Polku tiedoston sisältävään kansioon: "/home/<user>/.config/dpres-sip-compiler/".

CSC:n GitHubista kansioista "dpres\_sip\_compiler/conf/" kopioitu tiedosto "config\_musicarchive\_template.conf", jolle annetaan organisaatiokohtaiset määrytykset sip-compilerin käyttöä varten.

Kotikansiossa kansio ".config", jossa kansio "dpres-sip-compiler", jossa tiedosto "config.conf"

- name=x
- contract=urn:uuid:xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx
  - Jokaisen SIP:n METS-tiedostossa fi:CONTRACTID
- sign\_key=/home/<user>/dpres-config/sip\_sign\_pas.pem
  - Allekirjoitustiedoston polku annettava
- adaptor=x
  - Määrittää myös CSC:n kanssa sovitut organisaatiokohtaiset määrytykset, joiden mukaan paketointi suoritetaan: meta\_ending, csv\_ending, used\_checksum, desc\_root\_remove (riippuu enemmän käytetystä kuvailevan metadatan standardista(?))
- meta\_ending=\_\_metadata.xml
- csv\_ending=\_\_metadata.csv
- used\_checksum=MD5
- desc\_root\_remove=True

## PAS-palvelun käyttämisen vaatimukset (Musiikkiarkiston sisäinen dokumentti)

PAS-palvelu hyväksyy yhteydet vain organisaatioiden ilmoittamista IP-osoitteista. PAS-palvelun käyttäjä- ja sopimustunnukset sekä yhteysasetukset, ks. Musiikkiarkiston salasanatiedosto.

## Liite 2. Skaalan PAS-osion luomisen SQL-komennot (Musiikkiarkiston sisäinen dokumentti)

Erillisessä tekstitiedostossa:

- Aineiston hallinta Musiikkiarkistossa - Liite 2 - Skaalan PAS-osion luomisen SQL-komennot.mysql

## Liite 3. PAS-objektien luominen Skaalaan (Musiikkiarkiston sisäinen dokumentti)

Erillisessä tekstitiedostossa:

- Aineiston hallinta Musiikkiarkistossa - Liite 3 - PAS-objektien luominen Skaalaan.mysql

## Liite 4. PAS-objektien ylläpito Skaalassa, PREMIS-

## **tapahtumat (Musiikkiarkiston sisäinen dokumentti)**

Erillisessä tekstitiedostossa:

- Aineiston hallinta Musiikkiarkistossa - Liite 4 - PAS-objektien ylläpito Skaalassa, PREMIS-tapahtumat.mysql
  - PREMIS-tapahtuma: deletion
  - PREMIS-tapahtuma: migration
  - PREMIS-tapahtuma: filename change

## **Liite 5. SIP-toiminnot (Musiikkiarkiston sisäiset dokumentit)**

Paketoinnissa hyödynnetään CSC:n Musiikkiarkistolle tekemää *sip-compiler*-lisäosaa, jonka käyttäminen nopeuttaa itse paketointivaihetta seuraavista syistä

- Objekteja ei kopioida SIP-hakemistoon
  - Objektit on kopioitu SIP-hakemistoon – esimerkiksi ulkoiselle kovalevylle –, josta paketointivaiheessa luodaan tar- tai zip-paketti
    - Tar- tai zip-pakettia luodessa kopioiminen toki tapahtuu, mutta ilman *sip-compiler*-lisäosaa paketointikomento kopioisi tiedostot lähdehakemistosta SIP-hakemistoon, jonka jälkeen edelleen data kopioidaan tar- tai zip-pakettiin
- MD5-tiivisteet lähetetään SIP:n mukana CSV-tiedostossa, joten niiden laskeminen paketointivaiheessa ohitetaan
  - LTFS-nauhavarmennuksesta luovutaan KP-PAS-Palvelun käyttöönoton jälkeen, joten tiivisteiden laskeminen on suoritettava osana PAS-objektien viemistä Skaalaan
- Objektien validointi paketointivaiheessa on mahdollista ohittaa

Edellä on huomattava, että MD5-tiivisteet on aikaisemmin laskettu osana LTFS-nauhavarmennusta, josta luovutaan KP-PAS-Palvelun käyttöönoton jälkeen. Jatkossa on luotava prosessi, jossa mahdollisimman pian aineiston vastaanoton tai digitoinnin jälkeen lasketaan objektien tiivisteet ja viedään objektit Skaalaan. Samassa vaiheessa kannattaa suorittaa myös objektien validointi, jotta SIP-paketointivaiheessa on jo etukäteen tiedossa objektien yhteensopivuus KP-PAS-Palvelun kanssa.

Ohjeet ovat erillisissä tekstitiedostoissa:

### **SIP:n luominen**

- Aineiston hallinta Musiikkiarkistossa - Liite 5a1 - SIPn luominen.txt
- Aineiston hallinta Musiikkiarkistossa - Liite 5a2 - SIPn luomisen kirjaaminen Skaalaan.mysql
- Aineiston hallinta Musiikkiarkistossa - Liite 5a3 - SIPn paketointi CentOS 7 -ympäristössä.txt

### **SIP:n lähettäminen**

- Aineiston hallinta Musiikkiarkistossa - Liite 5b1 - SIPn lähettäminen PAS-palveluun.txt
- Aineiston hallinta Musiikkiarkistossa - Liite 5b2 - SIPn lähettämisen kirjaaminen Skaalaan.mysql

### **SIP:n tai AIP:n hakeminen sekä hallinnointi**

- Aineiston hallinta Musiikkiarkistossa - Liite 5c - PAS-palvelun REST-rajapinnan käyttäminen.txt

## **Liite 6. SQL-kyselyt (Musiikkiarkiston sisäinen dokumentti)**

Erillisessä tekstitiedostossa:

- Aineiston hallinta Musiikkiarkistossa - Liite 6 - SQL-kyselyt.mysql