



# KANSALLISARKISTO

## Muut digitaaliset tietoineistot – siirtopaketin muodostaminen

### Sisällys

1	OHJEEN TARKOITUS.....	1
2	SIIRRETTÄVÄ AINEISTO JA SIIRTOERÄT.....	1
3	SIIRTOPAKETIN MUODOSTAMINEN.....	2
	HAKEMISTORAKENTEET .....	2
	SIIRTOPAKETTI .....	4
4	TIETOAINEISTON KUVAILU.....	5

## 1 Ohjeen tarkoitus

Ohjeessa kuvataan, miten muodostetaan Kansallisarkistoon siirrettävä siirtopaketti sellaisesta digitaalisesta tietoaineistosta, joka ei sovi siirrettäväksi muissa Sähköisen arkistoinnin palvelun määrittelemissä siirtopakettityypeissä. Siirrettävä aineisto voi olla esimerkiksi verkkolevyillä säilytettyä monimuotoista aineistoa mukaan lukien AV-aineistot. Ohjetta tulee hyödyntää yhdessä taulukon 1 dokumenttien kanssa.

Tässä ohjeessa Siirtäjällä tarkoitetaan organisaatiota, joka on Kansallisarkistoon siirrettävän tietoaineiston rekisterinpitäjä/hallinnoija.

### Taulukko 1 Viittaukset muihin ohjeisiin ja dokumentteihin

Dokumentti	Tarkoitus
Vastaanotettavia tiedostoja koskevat ohjeet -ohje ja sen LIITE 1 Luettelo Kansallisarkistoon vastaanotettavista tiedostomuodoista sekä LIITE 2 Tekniset vaatimukset AV-tiedostoille	Ohjeessa kuvataan Kansallisarkistoon vastaanotettavia tiedostoja koskevat yleiset vaatimukset, tekstitiedostoissa sallitut merkistöt ja siirtokelpoiset tiedostoformaatit. Ohjeeseen on eritelty, missä siirtorakenteessa mitäkin tiedostomuotoa voi siirtää. Ohje on Kansallisarkiston verkkosivuilla.
Metatietolomake	Siirtäjä ilmoittaa metatietolomakkeella Sähköisen arkistoinnin palvelulle siirtoerää koskevat metatiedot. Siirtäjä voi hyödyntää metatietolomaketta siirrettävän aineistokokonaisuuden rakenteen, arkistollisen kontekstin kuvailutietojen ja käyttörajoitusten kokonaisuuden hahmottamisessa. Lomake on Kansallisarkiston verkkosivuilla.

## 2 Siirrettävä aineisto ja siirtoerät

**Muilla digitaalisilla tietoaineistoilla** tarkoitetaan tässä ohjeessa audiovisuaalista tietoaineistoa, kuva-aineistoa tai muuta digitaalista tietoaineistoa, joka ei sovellu siirrettäväksi muissa siirtorakenteissa. Siirrettävien tiedostojen on oltava Kansallisarkiston hyväksymässä muodossa.

**Siirrettävä tietoaineisto** yksilöidään Kansallisarkiston ja Siirtäjän välisessä **siirtosuunnitelmassa**, jonka Siirtäjä on toimittanut Kansallisarkistoon. Siirrosta sopimisen ja siirtosuunnitelman perusteella siirrettävästä tietoaineistosta koostetaan yksi tai useampi **siirtoerä**. Yksittäinen siirtoerä on looginen, ehyt ja

valmis kokonaisuus, jonka aineistoihin tai tietoihin ei kohdistu enää muokkaustarpeita.

Kun siirtoeriä on useampia, Siirtäjän tulee huolehtia, että jokaisen siirtoerän osalta kaikki arkistoitavaksi määrätty aineisto tulee siirrettyksi.

Siirtoerä jaetaan yhteen tai useampaan siirrettävään kokonaisuuteen, joka tallennetaan ohjeiden mukaiseen hakemistorakenteeseen ja paketoidaan siirtoa varten tiedostoksi (**siirtopaketti**), katso luku 3.

Kansallisarkisto ei tarjoa työkaluja siirtopakettien muodostamiseen, vaan Siirtäjän on muodostettava siirtopaketti käytössään olevilla työkaluilla tai luotava tarvittava ratkaisu sovellustoimittajansa/palveluntarjoajansa tai muun teknisen kumppanin kanssa.

**Kuva 1** Siirrettävä aineisto muodostaa yhden tai useamman loogisen siirtoerän, josta muodostetaan yksi tai useampi siirtopaketti.



### Huom!

- Onnistunut tietoaineistojen vastaanotto edellyttää siirtopaketin muodostamista ohjeiden mukaisesti.
- Pakettiin ei saa lisätä mitään muuta kuin ohjeenmukaiset hakemistot tiedostoineen.
- Lue ohjeet tarkkaan, niin vältyt turhalta työltä.
- Ohjeen tulkintaan liittyvissä kysymyksissä ole yhteydessä Sähköisen arkistoinnin palveluun.

## 3 Siirtopaketin muodostaminen

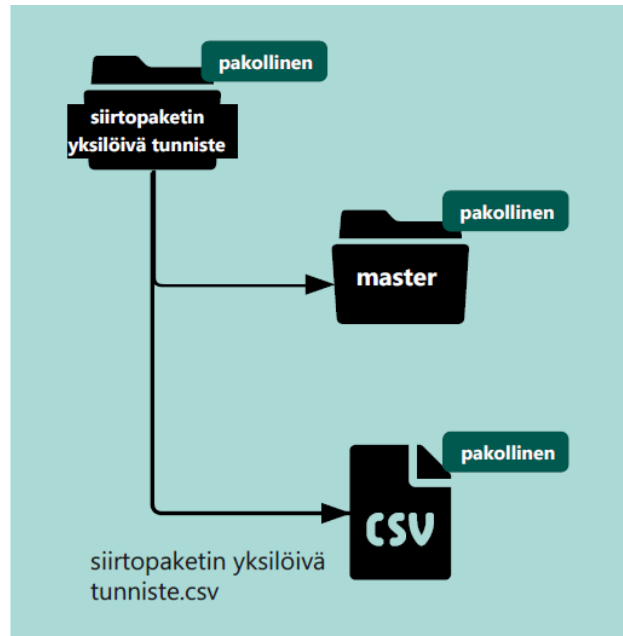
### Hakemistorakenteet

**Siirrettävä tietoaineisto** pitää tallentaa kuvan 2 mukaiseen hakemistorakenteeseen (kansiorakenne).

Juurihakemisto nimetään Siirtäjän määrittämällä siirtopaketin yksilöllisellä tunnisteella. Tunnistetta käytetään myös siirtokäyttöliittymässä siirtopaketin kontekstimetatietojen luonnin yhteydessä. Tunnisteen avulla siirtopaketti yhdistetään siihen liittyviin metatietoihin. Mikäli siirtopaketille annettu yksilöllinen tunniste ja siirtokäyttöliittymässä määritetty tunniste eivät ole identtiset, ei tietoaineiston siirto käyttöliittymässä onnistu.

Siirrettävä tietoaaineisto tallennetaan omaan **alihakemistoonsa** (master).

Jokaiselle alihakemistossa siirrettävälle tiedostolle on laskettava tarkistesumma, jotta Kansallisarkisto voi varmistua, että säilytykseen otetaan eheä tiedosto. Tarkistesummat tallennetaan csv-tiedostoon, joka tallennetaan juurihakemistoon.



**Kuva 2 Siirtopaketin hakemistorakenne**

Hakemistojen ja tiedostojen nimet ja sisältö on kuvattu taulukoissa 2 ja 3. Hakemistojen nimet ovat merkkikoririippuvaisia. Alihakemiston nimi kirjoitetaan pienellä alkukirjaimella.

**Taulukko 2** Juurihakemiston sisältö

Hakemisto/ Tiedosto	Selite/sisältö
juurihakemisto	<p>Juurihakemisto (pääkansio) pitää nimetä siirtopaketin tunnisteella, esimerkiksi <i>Paketti1</i> tai <i>vuodet9195</i>. Siirtäjä määrittelee tämän tunnisteeseen siirtokäyttöliittymässä siirtopaketin kontekstimetatietojen luonnin yhteydessä. Olennaista on, että tunniste yksilöi siirtopaketin muista saman siirtoerän paketeista.</p> <ul style="list-style-type: none"> <li>Tunniste saa sisältää seuraavia merkkejä: a–z, A–Z ja 0–9.</li> <li>Siirtopaketin tunnisteena ei saa käyttää Siirtäjälle kontekstimetatietojen luomista varten toimitettua metatietotunnistetta.</li> </ul>

Alihakemistoissa olevat tiedostot nimetään lukujonon avulla. Lähtökohtaisesti lukujonon tulee koostua neljästä numerosta. Numerointi on juokseva niin, että ensimmäinen tiedosto saa nimen 0001, toinen 0002 jne. Alihakemisto itsessään nimetään pienellä alkukirjaimella ”master”.

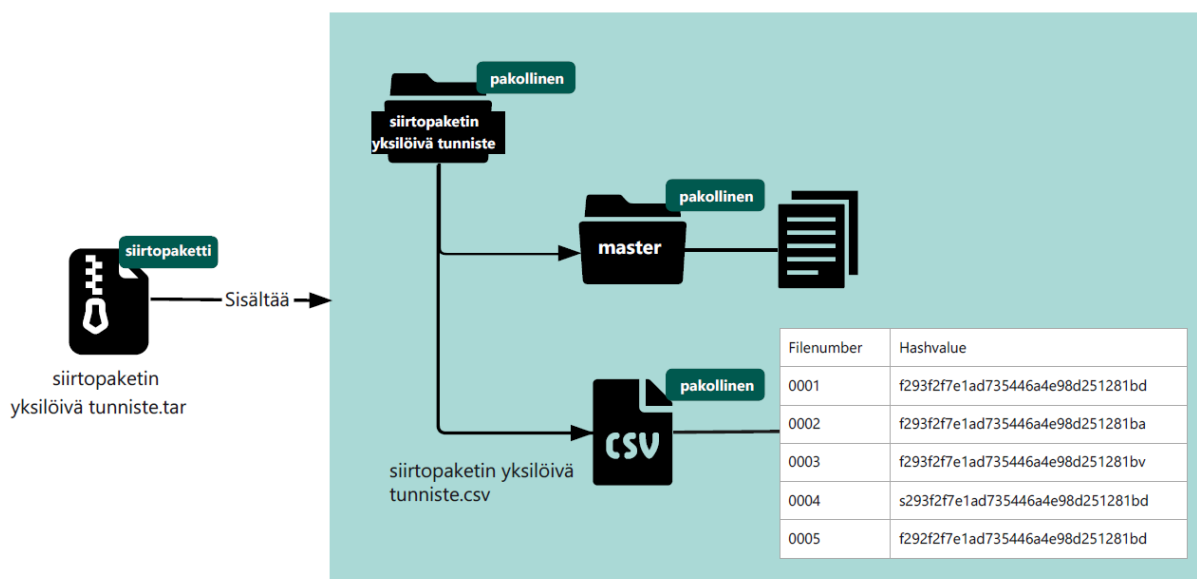
### Taulukko 3 Alihakemistojen sisältö

master	Alihakemisto sisältää siirrettävän tietoaineiston. Hakemisto on pakollinen, ja sen on sisällettävä tiedostoja. <ul style="list-style-type: none"><li>Hakemisto saa sisältää kaikkia muita SAPA:n vastaanottamia tiedostomuotoja paitsi pääsääntöisesti Rakenteinen tietoaineisto -vastaanottorakenteessa siirrettävät tiedostoformaattit (xml, json, csv).</li><li>Tiedostot nimetään lukujonon avulla.</li></ul>
siirtopaketin tunniste.csv	csv-tiedosto sisältää master-hakemiston sisältämien tiedostojen tiedostonimet (sarake Filenumber) ja tiedostoille lasketut tarkistesummat (sarake Hashvalue). Filenumber-sarakkeeseen ei sisällytetä tiedostopäätettä. Tarkistesumma annetaan muodossa MD5. <ul style="list-style-type: none"><li>Tiedosto tulee nimetä siirtopaketin tunnisteella, eli samalla merkkijonolla kuin juurihakemisto, esimerkiksi <i>Paketti1.csv</i> tai <i>vuodet9195.csv</i>.</li><li>Tiedoston merkistön pitää olla UTF-8.</li><li>Kenttien ympärillä <b>ei saa olla</b> lainausmerkkejä.</li></ul>

### Siirtopaketti

Juurihakemisto nimetään siirtopaketin tunnisteella. Hakemistot (kansiot) tiedostoineen paketoidaan siirtopaketiksi eli yhdeksi TAR-tiedostoksi. TAR-tiedoston saa lisäksi pakata häviöttömään GZIP- (.tar.gz) tai BZIP2- (.tar.bz2) muotoon.

Kun siirtopaketti ladataan siirtokäyttöliittymään, sen nimenä käytetty tunniste ohjaa aineiston osaksi oikeaa kuvailukokonaisuutta Kansallisarkiston metatietovarannossa.



**Kuva 2. Hakemistorakenne ja siirtopaketti. Tiedostot tallennetaan siirtopakettia varten kuvan mukaiseen hakemistorakenteeseen. Juurihakemisto nimetään siirtopaketin tunnisteella (kuvassa Identifier) ja alihakemistot tiedostoineen paketoidaan yhdeksi TAR-tiedostoksi.**

#### 4 Tietoaineiston kuvailu

Siirrettävän tietoaineiston arkistointi vaatii aina erillistä kuvailua. Kuvailun avulla varmistetaan, että tietoaineisto säilyy saavutettavana ja ymmärrettävänä, vaikka käyttäjä ei tuntisi aineistoa tai sen alkuperäistä käyttötarkoitusta ja historiaa.

Siirtopaketit eli arkistoyksiköt kuvaillaan siirtokäyttöliittymässä ennen niiden vastaanottoa. Siirtäjä toimittaa Sähköisen arkistoinnin palvelulle myös siirrettävän tietoaineistokokonaisuuden ja siihen kuuluvien siirtoerien arkistollista kontekstia, käyttörajoituksia ja aineistoon liittyviä toimijoita kuvaavat metatiedot erillisellä metatietolomakkeella.

Siirtäjä voi hyödyntää metatietolomaketta myös aineistokokonaisuuden rakenteen hahmottamisessa ja siirtopakettien kontekstimetatietojen suunnittelussa ennen pakettien metatietojen määrittelyä siirtokäyttöliittymässä.