



Kansallisarkisto

Muut digitaaliset tietoaineistot – siirtopaketin muodostaminen



Sähköisen arkistoinnin palvelu



Sisällys

1	OHJEEN TARKOITUS	1
1.1	VIITTAUKSET MUIHIN OHJEISIIN JA DOKUMENTTEIHIN.....	1
2	SIIRRETTÄVÄ AINEISTO JA SIIRTOERÄT	1
2.1	SIIRRETTÄVÄ KOKONAISUUS: MUUT DIGITAALISET TIETOAINESOT	1
2.2	SIIRTOERIEN MUODOSTAMINEN.....	2
3	SIIRTOPAKETIN MUODOSTAMINEN	2
3.1	HAKEMISTORAKENTEET.....	2
3.2	SIIRTOPAKETTI	3
4	AINESTON KUVAILU	4

1 Ohjeen tarkoitus

Ohjeessa kuvataan, miten muodostetaan Kansallisarkistoon siirrettävä siirtopaketti sellaisesta digitaalisesta tietoaaineistosta, 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 kohdan 1.1. dokumenttien kanssa.

1.1 Viittaukset muihin ohjeisiin ja dokumentteihin

Taulukko 1. Viittaukset muihin ohjeisiin ja dokumentteihin

Dokumentti	Tarkoitus
Vastaanotettavia tiedostoja koskevat ohjeet ja 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 tiedostoformaattit. Ohjeeseen on eritelty, missä siirtorakenteessa mitään tiedostomuotoa voi siirtää. Ohje on Sähköisen arkistoinnin palvelun kotisivuilla.
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 Sähköisen arkistoinnin palvelun kotisivuilla.

2 Siirrettävä aineisto ja siirtoerät

2.1 Siirrettävä kokonaisuus: Muut digitaaliset tietoaaineistot

Muilla digitaalisilla tietoaaineistoilla tarkoitetaan tässä ohjeessa audiovisuaalista tietoaaineistoa, kuva-aineistoa tai muuta digitaalista tietoaaineistoa, joka ei sovellu siirrettäväksi rakenteinen tietoaaineisto, digitoitu kuva-aineisto¹ tai SÄHKE2-aineistot -rakenteissa.

Siirrettävien tiedostojen on oltava Kansallisarkiston hyväksymässä muodossa.

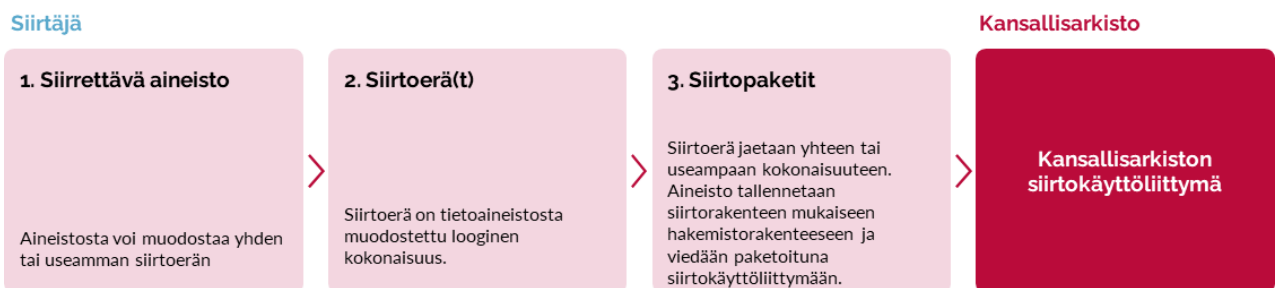
¹ Digitoidulla kuva-aineistolla tarkoitetaan Kansallisarkiston määritysten mukaisesti digitoitua tietoaaineistoa: [Kansallisarkiston vaatimukset hävittämiseen tähtäävään digitointiin \(2019\)](#) ja [Kansallisarkiston määritykset arkistoitavien erikoisaineistojen digitointiin \(2021\)](#). Määritysten noudattaminen mahdollistaa digitoidun tietoaaineiston säilyttämisen yksinomaan digitaalisena ja analogisen tietoaaineiston hävittämisen.

2.2 Siirtoerien muodostaminen

Siirrettävä aineisto yksilöidään Kansallisarkiston ja siirtäjän välisessä **siirtosuunnitelmassa**, jonka siirtäjä on toimittanut Kansallisarkistoon. Siirrosta sopimisen ja siirtosuunnitelman perusteella siirrettävästä aineistosta koostetaan yksi tai useampi **siirtoerä**. Yksittäinen siirtoerä on looginen, ehyt ja valmis kokonaisuus, jonka aineistoihin tai tietoihin ei kohdistu enää muokkaustarpeita.

Kun aineistoa siirretään vaiheittain (erissä), siirtäjän tulee huolehtia siitä, että jokaisen siirtoerän osalta kaikki arkistoitavaksi määrätty aineisto tulee siirretyksi.

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



Kuva 1. Siirrettävä aineisto siirretään yhdessä tai useammassa siirtoerässä. Jokaisesta siirtoerästä muodostetaan yksi tai useampi siirtokäyttöliittymään vietävä siirtopaketti.

3 Siirtopaketin muodostaminen

3.1 Hakemistorakenteet

Siirrettävä aineisto pitää tallentaa kuvan 2 mukaiseen hakemistorakenteeseen (kansiorakenne).

Juurihakemisto (pääkansio) pitää nimetä siirtopaketin yksilöllisellä tunnisteella. Siirtäjä määrittelee tämän tunnisteeseen siirtokäyttöliittymässä siirtopaketin kontekstimetatietojen luonnin yhteydessä. Jotta siirtopaketti siirtyy hyväksytysti säilytettäväksi, tulee tunnisteeseen vastata täysin siirtopaketille siirtokäyttöliittymässä määritellyä nimeä. Siirrettävä tietoaineisto tallennetaan omaan **alihakemistoonsa** (alihakemisto).

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

Hakemistojen ja tiedostojen nimet ja sisältö on kuvattu taulukossa 2. Hakemistojen nimet ovat merkkikokoriippuvaisia. Alihakemiston nimi kirjoitetaan pienellä alkukirjaimella.

Taulukko 2. Hakemistojen/tiedostojen nimet ja sisältö

Hakemisto/ Tiedosto	Selite/sisältö
juurihakemisto	Juurihakemisto (pääkansio) pitää nimetä siirtopaketin tunnisteella, esimerkiksi <i>Paketti1</i> tai <i>vuodet9195</i> . Siirtäjä määrittelee tämän tunnisteiden siirtokäyttöliittymässä siirtopaketin kontekstimetatietojen luonnin yhteydessä. Olennaista on, että tunniste yksilöi siirtopaketin muista saman siirtoerän paketeista. <ul style="list-style-type: none">Tunniste saa sisältää seuraavia merkkejä: a-z, A-Z ja 0–9.Siirtopaketin tunnisteena ei saa käyttää siirtäjälle kontekstimetatietojen luomista varten toimitettua metatietotunnistetta.
master	Alihakemisto sisältää siirrettävän tietoaineiston. Hakemisto on pakollinen, ja sen on sisällettävä tiedostoja. <ul style="list-style-type: none">Hakemisto saa sisältää kaikkia muita SAPA:n vastaanottamia tiedostomuotoja paitsi pääsääntöisesti Rakenteinen tietoaineisto -vastaanottorakenteessa siirrettävät tiedostoformaatit (xml, json, csv).Tiedostot nimetään lukujonon (0001, 0002, 0003, 0004) avulla. Lähtökohtaisesti lukujonon tulee koostua neljästä numerosta. Numerointi on juokseva niin, että ensimmäinen tiedosto saa nimen 0001, toinen 0002 jne.
siirtopaketin tunniste.csv	CVS-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">Tiedosto tulee nimetä siirtopaketin tunnisteella, eli samalla merkkijonolla kuin juurihakemisto, esimerkiksi <i>Paketti1.csv</i> tai <i>vuodet9195.csv</i>.Tiedoston merkistön pitää olla UTF-8.Kenttien ympärillä ei saa olla lainausmerkkejä.

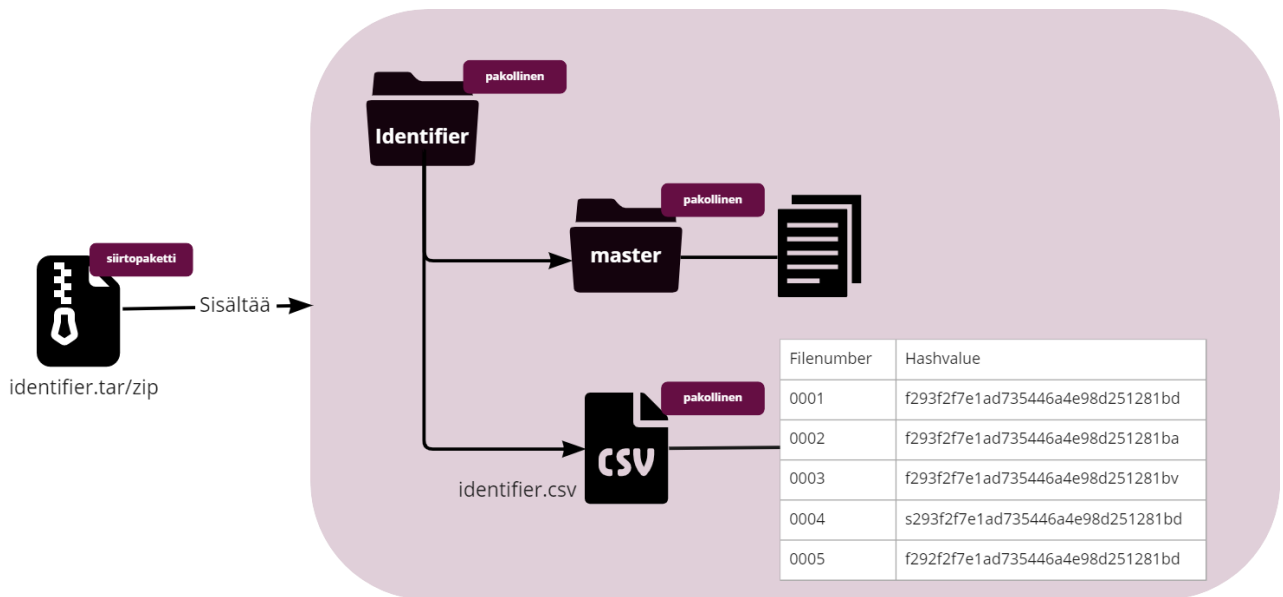
VINKKI!

Laskentataulukko-ohjelmistot muuntavat numeroita sisältävät merkkijonot usein automaattisesti luvuiksi (esim. 0001 -> 1). Solun muotoilu kannattaa muuttaa tekstiksi, jolloin merkkijono tallentuu syötetyssä muodossa.

3.2 Siirtopaketti

Juurihakemisto nimetään siirtopaketin tunnisteella. Hakemistot (kansiot) tiedostoineen paketoidaan siirtopaketiksi eli yhdeksi TAR-tiedostoksi. TAR-tiedoston saa lisäksi pakata häviöttömään GZIP- tai BZIP2-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 (pääkansio) nimetään siirtopaketin tunnistella (kuvassa Identifier) ja hakemistot (kansiot) tiedostoineen paketoidaan yhdeksi TAR-tiedostoksi.

4 Aineiston kuvailu

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

Siirtäjä toimittaa sähköisen arkistoinnin palvelulle siirtoerän 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.