



# Kansallisarkisto

Digitoitu kuva-aineisto – siirtopaketin muodostaminen



Sähköisen arkistoinnin palvelu



## Sisällys

<b>1.</b>	<b>OHJEEN TARKOITUS.....</b>	<b>1</b>
1.1.	VIITTAUKSET MUIHIN OHJEISIIN JA DOKUMENTTEIHIN .....	1
<b>2.</b>	<b>SIIRRETTÄVÄ AINEISTO JA SIIRTOERÄT: SIIRTOERIEN MUODOSTAMINEN .....</b>	<b>1</b>
<b>3.</b>	<b>SIIRTOPAKETIN MUODOSTAMINEN.....</b>	<b>2</b>
3.1.	HAKEMISTORAKENTEET .....	2
3.2.	SIIRTOPAKETTI.....	4
<b>4.</b>	<b>AINEISTON KUVAILU.....</b>	<b>6</b>

## 1. Ohjeen tarkoitus

Ohjeessa kuvataan Kansallisarkistoon siirrettävän siirtopaketin muodostaminen viranomaisen itse digitoimista aineistoista.

Ohje koskee Kansallisarkiston digitointimääritysten KA/3357/07.01.02.04.02/2019 ja KA/15780/07.01.02.04.02/2020 mukaisesti digitoituja aineistoja (katso taulukko 1).

Ohjetta tulee hyödyntää yhdessä kohdan 1.1. dokumenttien kanssa.

### 1.1. Viittaukset muihin ohjeisiin ja dokumentteihin

Taulukko 1. Viittaukset muihin ohjeisiin ja dokumentteihin

Dokumentti	Sisältö
Vastaanotettavia tiedostoja koskevat ohjeet ja LIITE 1 Luettelo Kansallisarkistoon vastaanotettavista tiedostomuodoista	Ohjeessa kuvataan Kansallisarkistoon vastaanotettavia tiedostoja koskevat yleiset vaatimukset, tekstimuotoisissa tiedostoissa sallitut merkistöt ja siirtokelpoiset tiedostoformaattit. Ohjeeseen on eritelty, missä siirtorakenteessa mitään tiedostomuotoa voi siirtää. Ohje on Sähköisen arkistoinnin palvelun kotisivuilla.
Kansallisarkiston vaatimukset hävittämiseen tähtäävään digitointiin (2019) KA/3357/07.01.02.04.02/2019	Kansallisarkiston vaatimukset digitoinnille, joka mahdollistaa analogisen ilmentymän hävittämisen.
Kansallisarkiston määräykset arkistoitavien erikoisaineistojen digitointiin (2021) KA/15780/07.01.02.04.02/2020	Kansallisarkiston määräykset arkistoitavien erikoisaineistojen digitointiin.
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: Siirtoerien muodostaminen

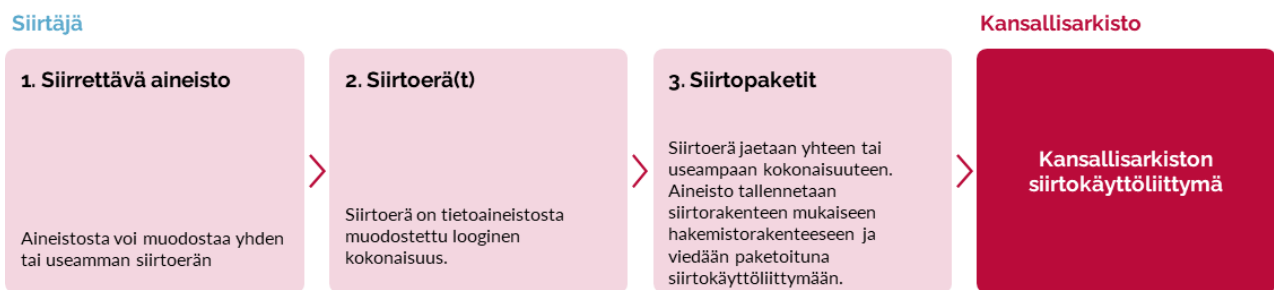
**Siirrettävä aineisto** yksilöidään Kansallisarkiston ja siirtäjän välisessä **siirtosuunnitelmassa**, jonka siirtäjä on toimittanut Kansallisarkistoon. Siirrostä sopimisen ja siirtosuunnitelman perusteella siirrettävästä aineistosta koostetaan yksi tai useampi

**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 muotoon ja 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 siirretään yhdessä tai useammassa siirtoerässä. Jokaisesta siirtoerästä muodostetaan yksi tai useampi siirtokäyttöliittymään vietävä siirtopaketti.

### 3. Siirtopaketin muodostaminen

#### 3.1. Hakemistorakenteet

Digitointiprosessissa tuotetut tiedostot pitää tallentaa noudatetun digitointimäärityksen mukaan joko kuvan 2 tai kuvan 3 mukaiseen hakemistorakenteeseen (katso luku 3.2.). Kansallisarkiston määritysten mukaisesti digitoitujen aineistojen yksityiskohtaiset metatietovaatimukset ja muut vaatimukset on lueteltu Kansallisarkiston digitointivaatimuksissa. Katso taulukko 1.

**Juurihakemisto** (pääkansio) pitää nimetä siirtopaketin yksilöllisellä tunnisteella. Siirtäjä määrittelee tämän tunnisteeseen siirtokäyttöliittymässä siirtopaketin kontekstimetatietojen luonnin yhteydessä. Jotta siirtopaketti siirtyy hyväksytysti säilytettäväksi, tulee siirtopaketin tunnisteeseen vastata täysin siirtokäyttöliittymässä määriteltyä nimeä. Siirrettävä tietoaineisto, kuvatiedostojen tekniset metatiedot ja mahdolliset tekstintunnistuksessa luodut tiedostot tallennetaan omiin **alihakemistoihinsa** (aliansio).

mix- ja ocr-hakemistojen sisältämien tiedostojen tiedostonimien tulee vastata master-hakemiston kuvatiedostojen nimiä.

**Esimerkki:**

- master-hakemistossa on kuvatiedosto **0001.jpg**.

- mix-hakemistossa **pitää** olla kyseisen kuvan tekniset metatiedot (xml-tiedosto **0001.xml**).
- ocr-hakemistossa **saa** olla kuvatiedoston tekstintunnistettu sisältö (xml-tiedosto **0001.xml**). Kuvatiedostoa vastaava numerointi liittää digitoinnissa luodun kuvatiedoston ja siihen liittyvän tekstitunnistetun sisällön yhteen. Tekstitunnistus mahdollistaa myöhemmin sisältöhakujen tekemisen siirtopaketin sisältämiin aineistoihin.

ocr-hakemisto voi sisältyä vain vuoden 2019 digitointimääritysten (KA/3357/07.01.02.04.02/2019) mukaan muodostettuihin siirtopaketteihin. Siirtopaketti saa sisältää vain sovelletussa määrityksessä kuvatun hakemistorakenteen mukaisia hakemistoja.

Hakemistojen nimet ja niiden sisällöt on kuvattu taulukossa 3. Hakemistojen nimet ovat merkkikokoriippuvaisia. Alihakemistojen nimet kirjoitetaan pienellä alkukirjaimella.

**Taulukko 3.** Hakemistojen nimet ja sisältö

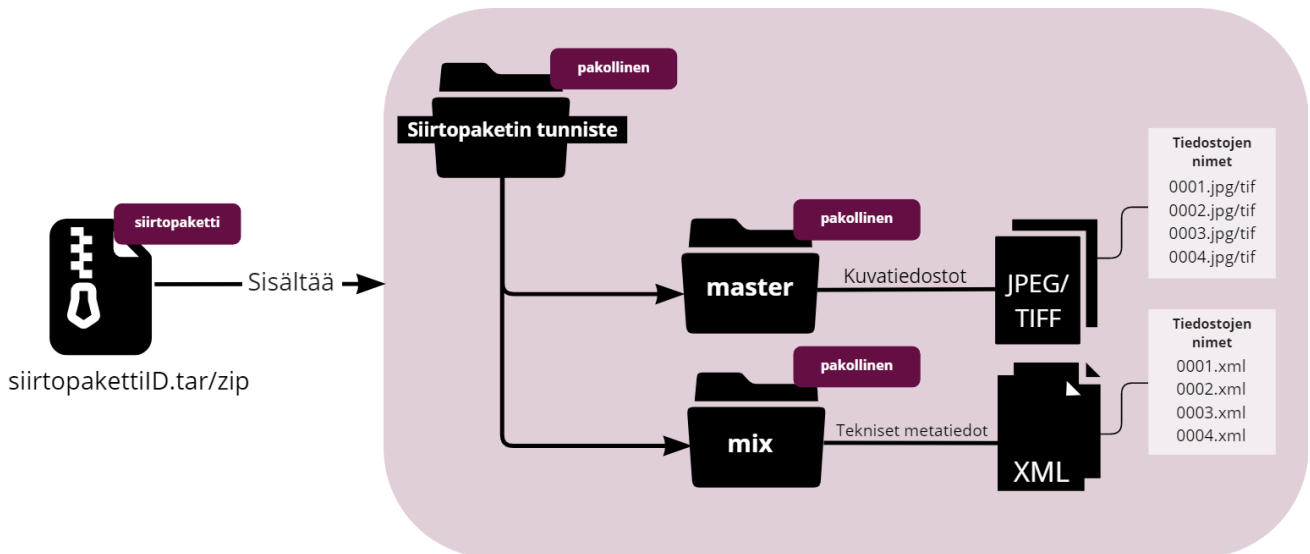
Hakemisto	Selite/sisältö	Koskee digitointimääritystä
juurihakemisto	Juurihakemisto (pääkansio) pitää nimetä siirtopaketin tunnisteella, esim. <i>paketti1</i> tai <i>vuodet9195</i> . Siirtäjä määrittelee paketin tunnisteeseen siirtokäyttöliittymässä siirtopaketin kontekstimetatietojen luonnin yhteydessä. Olennaista on, että tunniste yksilöi siirtopaketin muista siirtoerän paketeista. <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>	KA/3357/07.01.02.04.02/2019 KA/15780/07.01.02.04.02/2020
master	Alihakemisto sisältää kuvatiedostot yksittäisinä tiedostoina. <ul style="list-style-type: none"><li>• 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.</li></ul>	KA/3357/07.01.02.04.02/2019 KA/15780/07.01.02.04.02/2020

Hakemisto	Selite/sisältö	Koskee digitointimäärittystä
	<ul style="list-style-type: none"><li>• mix- ja ocr-hakemistojen tiedostonimien tulee vastata master-hakemiston kuvatiedostojen nimiä.</li></ul>	
mix	<p>Alihakemisto sisältää pakolliset tekniset metatiedot kaikista master-hakemiston sisällä olevista kuvatiedostoista.</p> <ul style="list-style-type: none"><li>• Tiedostot nimetään lukujonon (0001, 0002, 0003, 0004) avulla. Lähtökohtaisesti lukujonon tulee koostua neljästä numerosta. Numerointi on juokseva niin, että ensimmäinen saa nimen 0001, toinen 0002 jne.</li><li>• tiedostonimien tulee vastata master-hakemiston kuvatiedostojen nimiä.</li></ul>	KA/3357/07.01.02.04.02/2019 KA/15780/07.01.02.04.02/2020
ocr	<p>Alihakemisto sisältää kuvatiedostoista mahdollisesti tunnistetun tekstin (OCR tai HTR) AltoXML-tiedostot.</p> <ul style="list-style-type: none"><li>• Tiedostot nimetään lukujonon (0001, 0002, 0003, 0004) avulla. Lähtökohtaisesti lukujonon tulee koostua neljästä numerosta. Numerointi on juokseva niin, että ensimmäinen saa nimen 0001, toinen 0002 jne.</li><li>• tiedostonimien tulee vastata master-hakemiston kuvatiedostojen nimiä.</li></ul>	KA/3357/07.01.02.04.02/2019

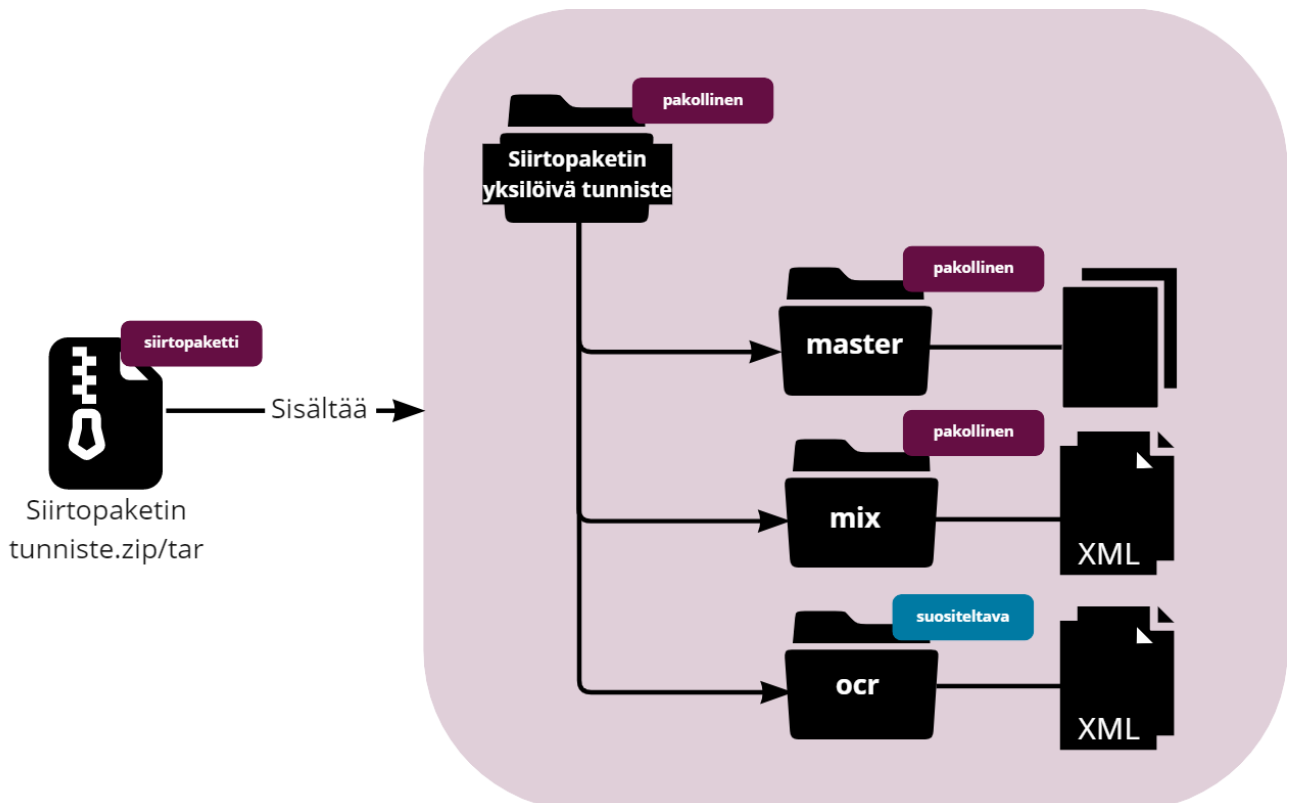
### 3.2. Siirtopaketti

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. KA/15780/07.01.02.04.02/2020** Digitoitujen erikoisaineistojen hakemistorakenne ja siirtopaketti. Tiedostot tallennetaan siirtopakettia varten kuvan mukaiseen hakemistorakenteeseen. Juurihakemisto nimetään siirtopaketin tunnisteella. Hakemistot (kansiot) tiedostoineen paketoidaan yhdeksi TAR-tiedostoksi.



**Kuva 3. KA/3357/07.01.02.04.02/2019** takautuvasti digitoitujen asiankäsittelyn tai muun toiminnan tuloksena muodostuneiden aineistokokonaisuuksien tai arkistoitujen aineistojen hakemistorakenne ja siirtopaketti. Tiedostot tallennetaan siirtopakettia varten kuvan mukaiseen hakemistorakenteeseen. Juurihakemisto nimetään siirtopaketin tunnisteella. 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.