



KANSALLISARKISTO

Digitoidut kuva-aineistot – siirtopaketin muodostaminen

Sisällys

1.	OHJEEN TARKOITUS.....	1
2.	SIIRRETTÄVÄ AINEISTO JA SIIRTOERÄT: SIIRTOERIEN MUODOSTAMINEN	2
3.	SIIRTOPAKETIN MUODOSTAMINEN	3
	HAKEMISTORAKENTEET	3
	SIIRTOPAKETTI.....	5
4.	TIETOAINEISTON KUVAILU.....	6



1. Ohjeen tarkoitus

Ohjeessa kuvataan Kansallisarkistoon siirrettävän siirtopaketin muodostaminen viranomaisen itse digitoimista aineistoista. Ohje koskee Kansallisarkiston digitointimäärittysten 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 taulukossa 1 mainittujen dokumenttien kanssa.

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

Taulukko 1. Viittaukset muihin ohjeisiin ja dokumentteihin

Dokumentti	Sisältö
Vastaanotettavia tiedostoja koskevat ohjeet -ohje ja sen 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 mitäkin tiedostomuotoa voi siirtää. Ohje on Kansallisarkiston verkkosivuilla.
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 kuvailutiedot. Siirtäjä voi hyödyntää metatietolomaketta siirrettävän aineistokokonaisuuden rakenteen, arkistollisen kontekstin kuvailutietojen ja käyttörajoitusten kokonaisuuden hahmottamisessa. Lomake on Kansallisarkiston verkkosivuilla.

2. Siirrettävä aineisto ja siirtoerät: Siirtoerien muodostaminen

Siirrettävä tietoaineisto 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 siirtoeriä on useampia, Siirtäjän tulee huolehtia, 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 tietoaineisto 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 tiedostoihin.
- Lue ohjeet tarkkaan, niin välttyt turhalta työltä.
- Ohjeen tulkintaan liittyvissä kysymyksissä ole yhteydessä Sähköisen arkistoinnin palveluun.

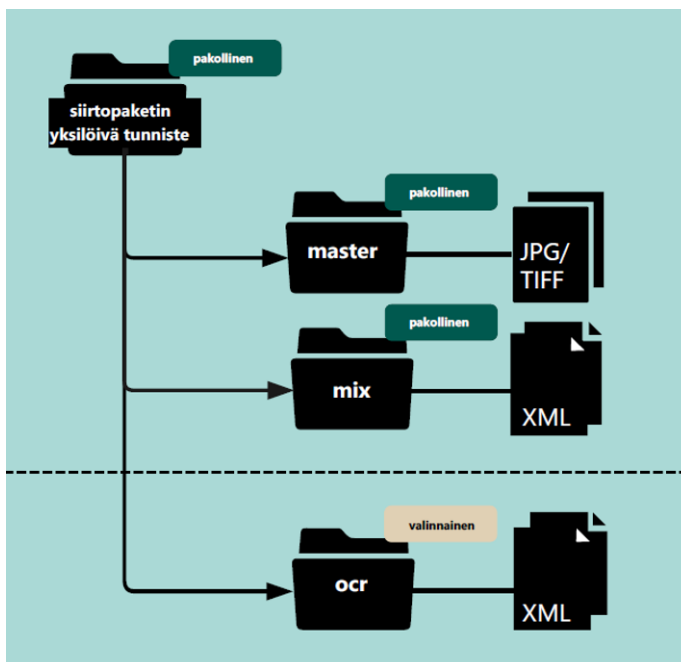
3. Siirtopaketin muodostaminen

Hakemistorakenteet

Digitointiprosessissa tuotetut digitointimäärityksen mukaiset tiedostot pitää tallentaa kuvan 2 mukaiseen hakemistorakenteeseen. Kansallisarkiston määritysten mukaisesti digitoitujen aineistojen yksityiskohtaiset metatietovaatimukset ja muut vaatimukset on lueteltu Kansallisarkiston digitointivaatimuksissa (katso taulukko 1).

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.

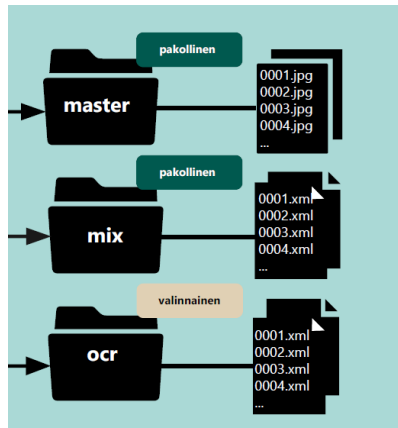
Siirrettävä tietoaineisto (master), kuvatiedostojen tekniset metatiedot (mix) ja mahdolliset tekstintunnistuksessa luodut tiedostot (ocr) tallennetaan omiin **alihakemistoihinsa**. Siirtopaketti saa sisältää vain sovelletussa digitointimäärityksessä kuvatun hakemistorakenteen mukaisia hakemistoja. Näin ollen ocr-alihakemisto voi sisältyä vain vuoden 2019 digitointimääritysten (KA/3357/07.01.02.04.02/2019) mukaan muodostettuihin siirtopaketteihin.



Kuva 2 Hakemistorakenne

ocr-hakemisto voi sisältyä vain digitointimäärityksen KA/3357/07.01.02.04.02/2019 mukaan muodostettuihin siirtopaketteihin

Tietoaineisto yhdistetään siihen liittyviin teknisiin metatietoihin ja mahdollisiin tekstintunnistuksessa luotuihin tietoihin käyttämällä samaa lukujonoon perustuvaa tiedostonimeä (katso kuva 3). Vain tiedoston pääte (esim. .jpg tai .xml) erottaa ne toisistaan. Toisin sanoen mix- ja ocr-hakemistojen sisältämien tiedostojen tiedostonimien tulee vastata master-hakemiston kuvatiedostojen nimiä. Kuvatiedostoa vastaava lukujono liittyy digitoinnissa luodun kuvatiedoston ja siihen liittyvät mix-metatiedot sekä tekstintunnistetun sisällön yhteen.



Kuva 3 Esimerkki nimeämisestä

- 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**).
Tekstintunnistus mahdollistaa myöhemmin sisältöhakujen tekemisen siirtopaketin sisältämiin aineistoihin.

Hakemistojen nimet ja niiden sisällöt on kuvattu taulukoissa 2 ja 3. Hakemistojen nimet ovat merkkikoriippuvaisia. Alihakemistojen nimet kirjoitetaan pienellä alkukirjaimella.

Taulukko 2. Juurihakemiston nimi ja sisältö

Hakemisto	Selite/sisältö	Koskee digitointimäärittystä
juurihakemisto	<p>Juurihakemisto pitää nimetä siirtopaketin tunnisteella, esim. <i>paketti1</i> tai <i>vuodet9195</i>. Siirtäjä määrittelee paketin tunnisteeseen ja käyttää sitä myös siirtokäyttöliittymässä siirtopaketin kontekstimetatietojen luonnin yhteydessä. Olennaista on, että tunniste yksilöi siirtopaketin muista siirtoerän paketeista.</p> <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. 	<p>KA/3357/07.01.02.04.02/2019</p> <p>KA/15780/07.01.02.04.02/2020</p>

Seuraavien alihakemistojen (master, mix ja ocr) kohdalla tiedostojen nimeämisen periaate on aina sama. Hakemistoissa olevat tiedostot nimetään lukujonon avulla. Lähtökohtaisesti lukujonon tulee koostua neljästä numerosta. Numerointi on juokseva niin, että ensimmäinen tiedosto saa nimen 0001, toinen 0002 jne. Alihakemistot itsessään nimetään sisältönsä mukaisesti master, mix tai ocr.

Taulukko 3. Alihakemistojen nimet ja sisältö

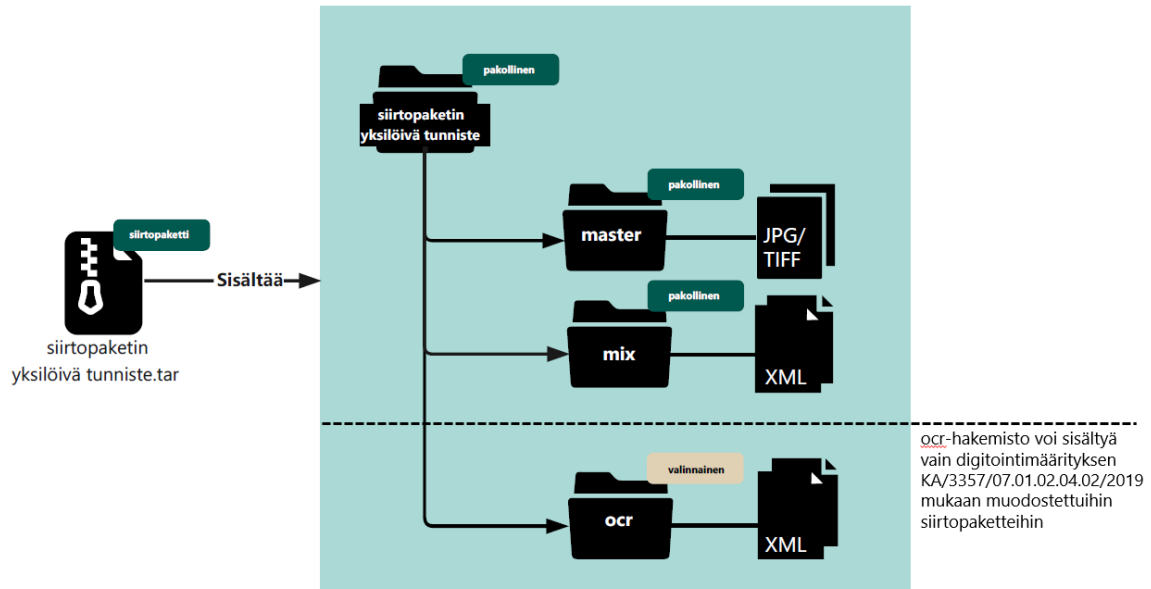
Alihakemisto	Selite/sisältö	Koskee digitointimäärittystä
master	Alihakemisto sisältää kuvatiedostot yksittäisinä tiedostoina. <ul style="list-style-type: none">Tiedostot nimetään lukujonon (0001, 0002, 0003, 0004) avulla.mix- ja ocr-hakemistojen tiedostonimien tulee vastata master-hakemiston kuvatiedostojen nimiä.	KA/3357/07.01.02.04.02/2019 KA/15780/07.01.02.04.02/2020
mix	Alihakemisto sisältää pakolliset tekniset metatiedot kaikista master-hakemiston sisällä olevista kuvatiedostoista. <ul style="list-style-type: none">Tiedostot nimetään lukujonon (0001, 0002, 0003, 0004) avulla.tiedostonimien tulee vastata master-hakemiston kuvatiedostojen nimiä.	KA/3357/07.01.02.04.02/2019 KA/15780/07.01.02.04.02/2020
ocr	Alihakemisto sisältää kuvatiedostoista mahdollisesti tunnistetun tekstin (OCR tai HTR) AltoXML-tiedostot. <ul style="list-style-type: none">Tiedostot nimetään lukujonon (0001, 0002, 0003, 0004) avulla.tiedostonimien tulee vastata master-hakemiston kuvatiedostojen nimiä.	KA/3357/07.01.02.04.02/2019

Siirtopaketti

Juurihakemisto nimetään siirtopaketin tunnisteella ja **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 4 Siirtopaketti on TAR-tiedosto, joka sisältää hakemistorakenteen mukaiset hakemistot tiedostoineen.



Huom! Siirtopaketit eli TAR-tiedostot kannattaa ennen siirtoa tallentaa niin, että yhteen sarjaan (tai muuhun laajempaan kokonaisuuteen) kuuluvat siirtopaketit ovat samassa hakemistossa. Tätä monipolisempia hakemistorakenteita saattavat merkittävästi hidastaa siirtopakettien lähettämistä, kun käyttäjä yrittää valita useamman paketin siirrettäväksi samalla kertaa.

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.