



# Kansallisarkisto

Rakenteiset tietoaaineistot – siirtopaketin muodostaminen



Sähköisen arkistoinnin palvelu

## Sisällys

<b>1. OHJEEN TARKOITUS</b> .....	<b>1</b>
1.1. VIITTAUKSET MUIHIN OHJEISIIN JA DOKUMENTTEIHIN.....	1
<b>2. SIIRRETTÄVÄ AINEISTO JA SIIRTOERÄT</b> .....	<b>1</b>
2.1. SIIRRETTÄVÄ KOKONAISUUS: RAKENTEINEN DATA .....	1
2.2. AINEISTON POIMINTA .....	2
2.3. SIIRTOERIEN MUODOSTAMINEN .....	2
<b>3. FORMAATTIKOHTAISET VAATIMUKSET</b> .....	<b>3</b>
3.1. XML (EXTENSIBLE MARKUP LANGUAGE).....	3
3.2. CSV (COMMA SEPARATED VALUES) .....	4
3.3. JSON (JAVASCRIPT OBJECT NOTATION).....	4
3.4. SIARD 2.1 (SOFTWARE INDEPENDENT ARCHIVING OF RELATIONAL DATABASES) .....	5
<b>4. AINEISTON KUVAILU JA DOKUMENTOINTI</b> .....	<b>5</b>
4.1. OHEISDOKUMENTAATIO .....	6
4.2. KUVAILEVAT METATIEDOT JA KÄYTTÖRAJOITUKSET .....	8
<b>5. TIEDOSTOJA JA NIMEÄMISTÄ KOSKEVAT VAATIMUKSET</b> .....	<b>8</b>
<b>6. SIIRTOPAKETIN RAKENNE</b> .....	<b>9</b>
6.1. HAKEMISTORAKENNE.....	9
6.2. SIIRTOPAKETTI.....	10
<b>LIITTEET</b> .....	<b>12</b>
LIITE 1 DIAARIAINEISTOT ESIMERKKINÄ RAKENTEISESTA TIETOAINIESTOSTA JA DIAARIAINEISTOJEN XML-RAKENNE .....	12
DIAARIAINEISTOJEN XML-RAKENNE.....	12

## 1. Ohjeen tarkoitus

Ohjeessa kuvataan, miten rakenteisesta datasta muodostetaan Kansallisarkistoon siirrettävä siirtopaketti.

Ohje on yleiskäyttöinen, sillä ratkaisut ja tilanteet, joissa rakenteista dataa koostetaan, poikkeavat toisistaan.

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	Ohjeessa kuvataan Kansallisarkistoon vastaanotettavia tiedostoja koskevat yleiset vaatimukset, tekstitiedostoissa sallitut merkistöt ja siirtokelpoiset tiedostoformaatit. Ohje on Sähköisen arkistoinnin palvelun kotisivuilla. Ohjeeseen on eritelty, missä siirtorakenteessa mitäkin tiedostomuotoa voi siirtää.
Diaariaineistojen XML-rakenne (LIITE 1)	Alkuperäisestä diaari-sovelluksesta saatavat datakoosteet tulee muodostaa ensisijaisesti tässä ohjeessa esitetyn rakenteen mukaisesti.
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: rakenteinen data

**Rakenteisella datalla** tarkoitetaan tässä ohjeessa koneellisesti käsiteltävään muotoon rakenteistettua tietoaineistoa. Kansallisarkiston siirtokäyttöliittymän kautta voidaan siirtää **XML**-, **CSV**- ja **JSON**-muotoista sekä **SIARD**-formaateissa olevaa dataa.

Siirrettävä kokonaisuus koostuu varsinaisesta rakenteisessa muodossa olevasta **data-aineistosta** ja siihen liittyvästä **ohjeidokumenttaatiosta** (katso kohta 4.1). Varsinainen data-aineisto koostuu yleensä yhdestä tai useammasta tiedostosta (**datakooste tai SIARD-tiedosto**).

Siirtäjä vastaa siitä, että koostettu data on ehyttä ja hyödynnettävissä koneluettavassa muodossa.

Datakoosteissa ja oheisdokumentaatioissa käytettyjen tiedostojen on oltava Kansallisarkiston hyväksymässä muodossa.

## 2.2. Aineiston poiminta

Siirrettävä kokonaisuus poimitaan tavallisesti yhdestä tai useammasta tietolähteestä (esimerkiksi tietojärjestelmä). Aineistopoiminnan suunnittelussa ja toteutuksessa on tärkeää tunnistaa seulontapäätöksessä arkistoitavaksi määrätyt tietoyksiköt. Tavoite on, että arkistoinnin kannalta merkityksellinen tieto siirretään eheässä ja käyttökelpoisessa muodossa.

Poiminta voi perustua määritykseen, jota viranomaisen hyödyntää jo jossain säännöllisessä tai lakisäätöisessä tiedonsiirrossa tai -luovutuksessa. Tällainen käyttötarkoitus voi olla esimerkiksi tilastointi, tietojen siirtäminen tietojärjestelmien välillä tai välitettävä raportti. Poiminnan sisältöä tulee kuitenkin täydentää seulontapäätöksessä arkistoitavaksi määrätyillä tiedoilla, jos ne eivät ennestään sisälly määritykseen.

Tietoaaineisto tai sen tiedot eivät saa koostettaessa ja siirrettäessä vahingossa muuttua, ja mahdolliset muutokset on pystyttävä jäljittämään. Esimerkiksi erilaisten näyteaineistojen avulla voidaan arvioida, miten varmistetaan

- aineiston alkuperäiseen käyttötarkoitukseen liittyvän loogisen rakenteen dokumentointi ja
- datakoosteiksi puretun aineiston käytettävyys.

Tiedon kattavuutta ja eheyttä tulisi testata useissa eri vaiheissa. Siirtäjän tulee varmistaa, että siirrettävä aineisto sisältää kaiken siirrettäväksi sovitun ja arkistoitavaksi määrätyn. Testaamiseen tulisi ulottua teknisen eheyden lisäksi myös itse tietosisältöön. Tämä on erityisen tärkeää silloin, kun alkuperäisestä sovelluksesta tai järjestelmästä ollaan luopumassa. Tällöin tiedon eheyttä ei voida jälkikäteen varmentaa alkuperäisessä ympäristössä eikä poimintaa tarpeen vaatiessa toistaa.

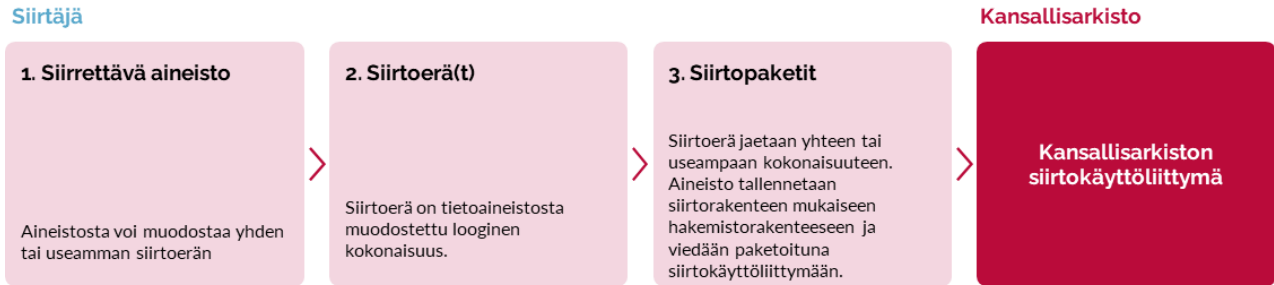
Kansallisarkisto ei tarjoa työkaluja aineistopoiminnan tai siirtopakettien muodostamiseen, vaan siirtäjän on tehtävä aineistopoiminta käytössään olevilla työkaluilla tai luotava tarvittava ratkaisu sovellustoimittajansa/palveluntarjoajansa tai muun teknisen kumppanin kanssa.

## 2.3. 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.

Siirtoerä jaetaan yhteen tai useampaan siirrettävään kokonaisuuteen, joka tallennetaan ohjeiden mukaiseen hakemistorakenteeseen ja paketoidaan siirtoa varten tiedostoksi

(siirtopaketti), katso luku 6. Esimerkiksi diaariaineiston vuosittainen siirtoerä voidaan jakaa siirtopaketteihin tehtäväryhmittäin.



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

### 3. Formaattikohtaiset vaatimukset

#### 3.1. XML (Extensible Markup Language)

XML-rakenne on hyvä keino säilyttää alkuperäisen datan hyödynnettävyys erityisesti silloin, kun alkuperäinen tietorakenne on hierarkkinen tai muodostuu valmiiksi määritellyistä määrämuotoisista tietorakenteista.

XML-rakenteessa osa tietoelementtien kuvailusta voidaan tallentaa osaksi itse data-aineistoa. Formaattiin sisältyvien validointimenetelmien ja informaation määrämuotoisuuden avulla voidaan tukea tiedon koneellista käsittelyä ja varmistaa, että informaatio on rakenteellisesti oikeanlaista.

Jos XML on muodostettu skeeman avulla, tulee datakoosteessa viitata käytettyyn skeemaan suoralla viittauksella siirtopaketin hakemistorakenteeseen, esim. `xi:schemaLocation="../schemas ../schemas/esimerkki.xsd"`

Skeemat, joihin datakoosteissa viitataan, tulee aina tallentaa siirtopaketin **schemas**-hakemistoon. Skeemat nimetään xml:ssä mainittujen skeemasijaintien tiedostonimien mukaan (ilman polkua), esim. **esimerkki.xsd**. Katso myös luku 6.1. XML:n rakenne validoidaan vastaanoton yhteydessä skeeman avulla.

XML:n rakenne suositellaan aina muodostamaan skeeman avulla. Näin varmistetaan siirrettävän tietoaineiston rakenteen säilyminen pitkäaikaissäilytyksessä. Tämän ohjeen liitteessä 1 on Kansallisarkiston diaariaineistolle määrittelämä XML-rakenne.

**Taulukko 2.** XML-rakennetta koskevat vaatimukset

Hyväksyttävä versio	PRONOM: fmt/101
---------------------	-----------------

Huom! Tarkista ennen siirtoa, että XML:n merkistö on siirtokelpoisessa muodossa (ISO 8859-15, UTF-8, UTF-16, UTF-32) ja ilmoitettu tiedostossa oikein. Usein oletuksena on Windows-1252-merkistö, jolloin XML-deklaraatiossa on attribuutti `encoding=Windows-`

1252". Tiedoston merkistö tulee muuntaa siirtokelpoiseksi ja sen attribuutti käytetyn merkistön mukaiseksi.

### 3.2. CSV (Comma Separated Values)

CSV-rakenne ei yleensä sovellu monimutkaisten, sisäisesti hierarkkisten ja ristiviittauksia sisältävien tietorakenteiden tallentamiseen. Rakennetta voidaan kuitenkin hyödyntää, kun siirrettävä tietoaaineisto muodostaa yksinkertaisen taulukkomaisen kokonaisuuden.

CSV-muodon käyttö edellyttää aina tietorakenteen dokumentointia. Ilman sitä datakoosteen sisältämää informaatiota ei ole mahdollista tulkita.

CSV-tiedoston kukin rivi kuvaa yksittäistä tietuetta. Rakenteen alussa tulee olla otsikkorivi, joka sisältää varsinaisten tietokenttien (sarakkeiden) otsikot. Muut CSV-rakennetta koskevat vaatimukset on koottu taulukkoon 3.

**Taulukko 3.** CSV-rakennetta koskevat vaatimukset

Rivien erotinmerkit	CR tai CR-LF
Kenttien erotinmerkit	Erottimena voi toimia esimerkiksi pilkku (,), puolipilkku (;), pystyviiva ( ) tai tabulaattori
Tietokentissä olevat tekstielementit erotinmerkkien välissä pitää yhdistää kokonaisuuksiksi käyttäen ' tai " merkkejä.	Esimerkki: 120155–1234; 'Testi Nieminen';040-323212;'Osoitteenkatu 3'
Hyväksyttävä versio	PRONOM: x-fmt/18

Huom! Siirtokäyttöliittymä voi tulkita tietyt CSV-tiedoston kentissä käytetyt merkit erotinmerkeiksi ja hylätä tiedoston validoinnin yhteydessä. Jos kentät sisältävät erotinmerkeiksi tulkittavia merkkejä, tulee niiden ympärille lisätä lainausmerkit, jolloin järjestelmä tulkitsee lainausmerkkien sisään jäävän osuuden tekstiksi.

### 3.3. JSON (JavaScript Object Notation)

JSON on erityisesti tiedonvälittämisessä käytetty rakenteinen tekstitiedostomuoto. Se yhdistää tietosisällön sekä tietorakenteen samalla tavalla kuin XML, mutta on rakenteena kevyempi.

Tietosisällön esittäminen perustuu JSON-muodossa tieto-objektien ja niihin liittyvien arvojen listoihin. JSON-tiedostoon tallennettavat kentät ja arvot on aina määriteltävä aineistokohtaisesti.

**Taulukko 4.** JSON-rakennetta koskevat vaatimukset

Hyväksyttävä versio	PRONOM: x-fmt/817
---------------------	-------------------

**3.4. SIARD (Software Independent Archiving of Relational Databases)**

SIARD on relaatiotietokantojen arkistointiin kehitetty avoimen lähdekoodin säilytysformaatti, joka perustuu avoimiin standardeihin (Unicode, XML ja SQL:2008) sekä ZIP64-tiedostoformaattiin.

Rakenteeltaan SIARD on määrämuotoinen, pakkaamaton ZIP-tiedosto, jonka sisältö on jaettu kahteen kansioon. Toinen kansio sisältää tietokannan taulujen sisällöistä muodostetut XML-tiedostot ja toinen XML-tiedostoon tallennettuna tietokannan rakenteen ja kuvauksen. SIARD ei tallenna käyttöliittymän ominaisuuksia eikä siten ole sellaisenaan ole käytettävissä oleva, operatiivinen tietokanta, mutta sen sisältö voidaan palauttaa käytettäväksi tietokannanhallintajärjestelmään.

**Taulukko 5.** SIARD-tiedostoa koskevat vaatimukset

ZIP-versio	32- tai 64-bittinen, myöhempi kuin versio 4.5
Pakkaus	Tiedoston häviötön pakkaus on sallittu.
Tiedoston päätte	.siard
Tiedostojen suojaus	Salaus tai salanasuojaus ei ole sallittu.
Hyväksyttävä versio	PRONOM: x-fmt/1196

**4. Aineiston kuvailu ja dokumentointi**

Tietoaineiston todistusvoimaisuuden ja ymmärrettävyyden säilyttämiseksi data-aineisto vaatii tuekseen erillistä kuvailua ja dokumentaatiota, sillä alkuperäinen järjestelmä ei enää siirron jälkeen tue aineiston käyttöä.

Kuvailun ja dokumentaation avulla varmistetaan, että data-aineiston rakenne ja tietosisältö voidaan ymmärtää, vaikka käyttäjä ei tuntisi aineistoa tai sen alkuperäistä käyttötarkoitusta ja historiaa.

Pysyvästi säilytettäväksi/arkistoitavaksi määrätyn aineiston lukemiseen, käyttöön ja ymmärtämiseen samoin kuin todistusvoimaisuuden varmistamiseen tarvittavat asiakirjat säilytetään myös pysyvästi. Tällaisia ovat esimerkiksi tietojärjestelmän käyttäjille suunnatut käsikirjat/manuaalit/käyttöohjeet, systeemidokumentit, tiedostojen kuvaukset, koodiluettelot ja tietojärjestelmäselosteet.<sup>1</sup>

<sup>1</sup>Valtionhallinnon asiakirjojen seulonta ja hävittäminen 2010. Määräys ja ohje 3.8.2010 (AL/19273/07.01.01.00/2008), luku 9, sivut 37, 40–41). <https://kansallisarkisto.fi/viranomaisille/Julkishallinnon-asiakirjahallinnon-ja-arkistotoimen-ohjaus/maeaeaykset/valtionhallinnon-asiakirjojen-seulonta>.

#### 4.1. Oheisdokumentaatio

Oheisdokumentaatio on data-aineiston käytettävyyttä ja tulkintaa tukevaa aineistoa, joka tallennetaan siirtopaketissa omaan **documentation**-hakemistoonsa, katso luku 6 Siirtopaketin rakenne.

Oheisdokumentaatiolla annetaan yleiskuvaus siirrettävästä aineistosta ja toiminnasta, jonka tuloksena aineisto on muodostunut sekä kuvataan data-aineiston tietorakenteet. Jokaisen datakoosteen sisäinen tietorakenne pitää dokumentoida kattavasti. Sarakkeet ja kentät tulee kuvata, jotta koosteen sisältö säilyy ymmärrettävänä. Kun datakoosteita on useita, pitää erityistä huomiota kiinnittää niiden välisten suhteiden ja rakenteiden dokumentointiin.

Dokumentaatiossa voidaan hyödyntää kuvauksia ja ohjeita, jotka tietojärjestelmästä ja aineistoista on laadittu alkuperäisen käyttötarkoituksen aikana. Dokumentaatiota on kuitenkin yleensä laajennettava ja tarkennettava vastaamaan nimenomaan arkistosiirron ja arkistoinnin vaatimuksia. Tarvittava oheisdokumentaatio arvioidaan siirrosta sopimisen yhteydessä.

**Taulukko 6.** Oheisdokumentaation avulla kuvattavia kokonaisuuksia

Kuvaus	Selite
Yleiskuvaus siirrettävästä aineistosta	<ul style="list-style-type: none"><li>• vapaa kuvaus</li><li>• sekä esimerkiksi asiakirjajulkisuuskuvaus ja muut valmiit kuvaukset</li><li>• muut tarvittavat dokumentit</li></ul>
Data-aineiston syntykonteksti	<p>Kuvaukset toiminnasta, jonka tuloksena aineisto on muodostunut</p> <ul style="list-style-type: none"><li>• vapaa kuvaus</li><li>• sekä esimerkiksi liittyvät lait, asetukset, normit</li><li>• toimintaohjeet, toimintakertomukset, raportit tms.</li><li>• muut tarvittavat dokumentit</li></ul>
Kuvaus data-aineiston muodostumisprosessista ja tietojen alkuperästä	<ul style="list-style-type: none"><li>• mistä lähteistä tietoja on saatu tai saadaan</li><li>• miten tiedot on kerätty eri lähteistä</li><li>• mahdolliset ohjeet ja säännöt, jotka määrittelevät, miten</li></ul>



	<p>tiedot on kirjattu/tallennettu alkuperäisessä ympäristössä</p> <ul style="list-style-type: none"><li>• miten tietoja on käsitelty</li></ul>
Data-aineiston tietorakenteiden kuvaus	<ul style="list-style-type: none"><li>• aineiston sisäinen tietorakenne</li><li>• yhteys Kansallisarkiston seulontapäätöksessä yksilöityihin, säilytettäväksi määrättyihin tietokokonaisuuksiin</li></ul> <p>Kuvaustapana voidaan käyttää graafisia tietomallikuvauksia (esim. ER-malli) ja teksti- tai taulukkokodokumentteja tietojen rakenteista ja suhteista.</p>
Data-aineiston sisältämien koodistojen ja niiden arvojoukkojen kuvaukset	<ul style="list-style-type: none"><li>• data-aineistossa hyödynnetyt yleiset koodistot ja niiden arvojoukot</li><li>• data-aineistossa hyödynnetyt organisaation sisäiset koodistot ja niiden arvojoukot</li></ul>
Data-aineistossa hyödynnetyt standardit ja suositukset	<ul style="list-style-type: none"><li>• hyödynnetyt yleiset standardit ja suositukset</li><li>• hyödynnetyt alakohtaiset standardit ja suositukset</li></ul>
Kattavat esimerkit data-aineiston käyttötavoista, käytöstä ja hyödyntämisestä	<ul style="list-style-type: none"><li>• vapaa kuvaus</li><li>• ohjeet</li><li>• dokumentoivat kuvakaappaukset</li></ul>
Siirrettävään kokonaisuuteen kuuluvat datakoosteet	<ul style="list-style-type: none"><li>• seloste (esimerkiksi taulukko) datakoosteista</li><li>• viittaus tiedoston nimen ja sisällön välillä</li></ul>
Datakoosteiden rakenne	<ul style="list-style-type: none"><li>• datakoosteiden sisäinen tietorakenne</li><li>• datakoosteiden väliset suhteet ja rakenteet</li></ul>

Pääperiaate on, että oheisdokumentaatio on itsenäisesti ymmärrettävissä, eikä se ole riippuvaista ulkopuolisista tietovarannoista tai palveluista. Myös avoimesti verkossa saatavilla olevien koodistojen tietosisältö tulee arkistoida oheisdokumentaationa.

**Esimerkki:**

Osa julkishallinnon yhteisistä koodistoista on saatavilla esimerkiksi Digi- ja väestötietoviraston ylläpitämässä koodistopalvelussa (<https://koodistot.suomi.fi/>). Siirrettävän tietoaineiston alkuperäisessä sovellusympäristössä on hyödynnetty Tilastokeskuksen tässä koodistopalvelussa ylläpitämää ammattiluokitusta (Ammattiluokitukset 2010). Oheisdokumentaatioon on tällöin dokumentoitava tieto käytetystä koodistosta (vuosi ja aineistossa käytetty koodiston versio). Koodiston tietosisältö pitää liittää osaksi oheisdokumentaatiota.

Oheisdokumentaatioon tulee liittää myös seloste siirtopaketin datakoosteista sekä viittaukset tiedoston nimen ja sisällön välillä. Seloste voi olla esimerkiksi taulukkotiedosto, jossa on yhdessä sarakkeessa esitetty tiedoston sisältö ja toisessa sarakkeessa tiedoston nimi.

**Taulukko 7.** Esimerkki viittauksesta tiedoston ja sen sisällön välillä

Tiedoston sisältö	Tiedostonimi
Virasto X:n toiminnan kuukausitilasto lokakuu 2017	0001.xml
Virasto X:n toiminnan kuukausitilasto marraskuu 2017	0002.xml

#### 4.2. Kuvailevat metatiedot ja käyttörajoitukset

Siirtäjä toimittaa sähköisen arkistoinnin palvelulle siirtoerän arkistollista kontekstia, käyttörajoituksia ja aineistoon liittyviä toimijoita kuvaavat metatiedot erillisellä **metatietolomakkeella**.

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

#### 5. Tiedostoja ja nimeämistä koskevat vaatimukset

Tiedostoja koskevat yleiset vaatimukset, luettelo vastaanotettavista tiedostomuodoista ja tiedostoissa sallituista merkistöistä löytyvät Kansallisarkistoon vastaanotettavia tiedostoja koskevasta ohjeesta.

Oheisdokumentaation ei voi tällä hetkellä sisältyä XML-, CSV- ja JSON-formaatissa olevia tiedostoja tai kuvatiedostoja (TIFF, JPEG).

Datakoosteet ja oheisdokumentaation kuuluvat 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. Katso myös taulukko 7.

Skeemat nimetään XML:ssä mainittujen skeemasijaintien tiedostonimien mukaan (ilman polkua), katso luku 3.1 ja taulukko 7.

## 6. Siirtopaketin rakenne

### 6.1. Hakemistorakenne

Datakoosteet, oheisdokumentaatio ja mahdolliset skeemat, joita datakoosteissa hyödynnetään, 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 määriteltyä nimeä.

Datakoosteet, oheisdokumentaatio ja mahdolliset skeemat tallennetaan omiin **alihakemistoihinsa** (aliansiot).

Jokaiselle datakoosteelle on laskettava tarkistesumma (MD5), jotta Kansallisarkisto voi varmistua, että säilytykseen otetaan eheä tiedosto. Tiedostojen tarkistesummat tallennetaan erilliseen CSV-tiedostoon, joka tallennetaan juurihakemistoon.

Hakemistojen nimet ja sisältö on kuvattu taulukossa 7. Hakemistojen nimet ovat merkkikokoriippuvaisia, ja alihakemistojen nimet kirjoitetaan pienellä alkukirjaimella.

**Taulukko 8.** Hakemistojen nimet ja sisältö

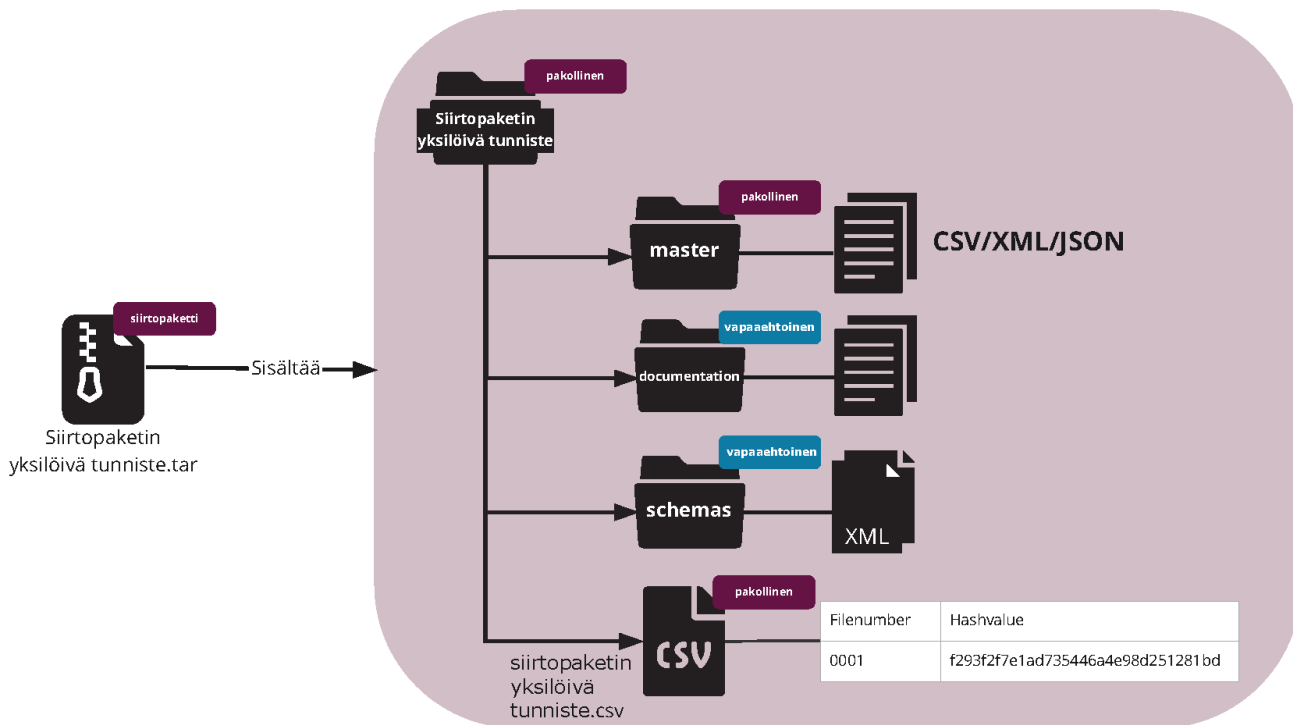
Hakemisto	Selite/sisältö
juurihakemisto	Juurihakemisto (pääkansio) pitää nimetä siirtopaketin tunnisteella, esim. <i>Paketti1</i> tai <i>vuodet9195</i> . Siirtäjä määrittelee paketin tunnisteeseen siirtokäyttöliittymässä siirtopaketin kontekstimetatietojen luonnin yhteydessä. Olennaista on, että tunniste yksilöi siirtopaketin muista saman 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>
master	Alihakemisto sisältää siirrettävän tietoaineiston, datakoosteet. Hakemisto on pakollinen, ja sen on sisällettävä tiedostoja. <ul style="list-style-type: none"><li>• Hakemistoon saa tallentaa CSV-, XML-, JSON- ja SIARD-muotoisia tiedostoja.</li><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. SIARD-tiedosto nimetään aina 0001.siard.</li></ul>

	<ul style="list-style-type: none"><li>• Jos master-hakemiston datakoosteissa viitataan skeemoihin, pitää skeemat tallentaa schemas-hakemistoon.</li></ul>
documentation	<p>Alihakemisto sisältää mahdollisen oheisdokumentaation.</p> <ul style="list-style-type: none"><li>• Hakemistoon ei saa tallentaa XML-, CSV- ja JSON-formaatissa olevia tiedostoja tai kuvatiedostoja (TIFF, JPEG).</li><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>
schemas	<p>Alihakemisto sisältää mahdolliset skeemat. Skeemat ovat pakollisia silloin, kun xml-tiedosto on muodostettu skeeman mukaisesti ja skeemaan viitataan master-hakemiston datakoosteissa.</p> <ul style="list-style-type: none"><li>• Skeemat nimetään xml:ssä mainittujen skeemasijaintien tiedostonimien mukaan (ilman polkua), esim. esimerkki.xsd, jos skeemaviittaus on seuraava xsi:schemaLocation=<a href="https://virasto.fi/skeemat">https://virasto.fi/skeemat</a> <a href="https://virasto.fi/skeemat/esimerkki.xsd">https://virasto.fi/skeemat/esimerkki.xsd</a></li></ul>
siirtopaketin tunniste.csv	<p>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.</p> <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 ei saa olla lainausmerkkejä.</li></ul>

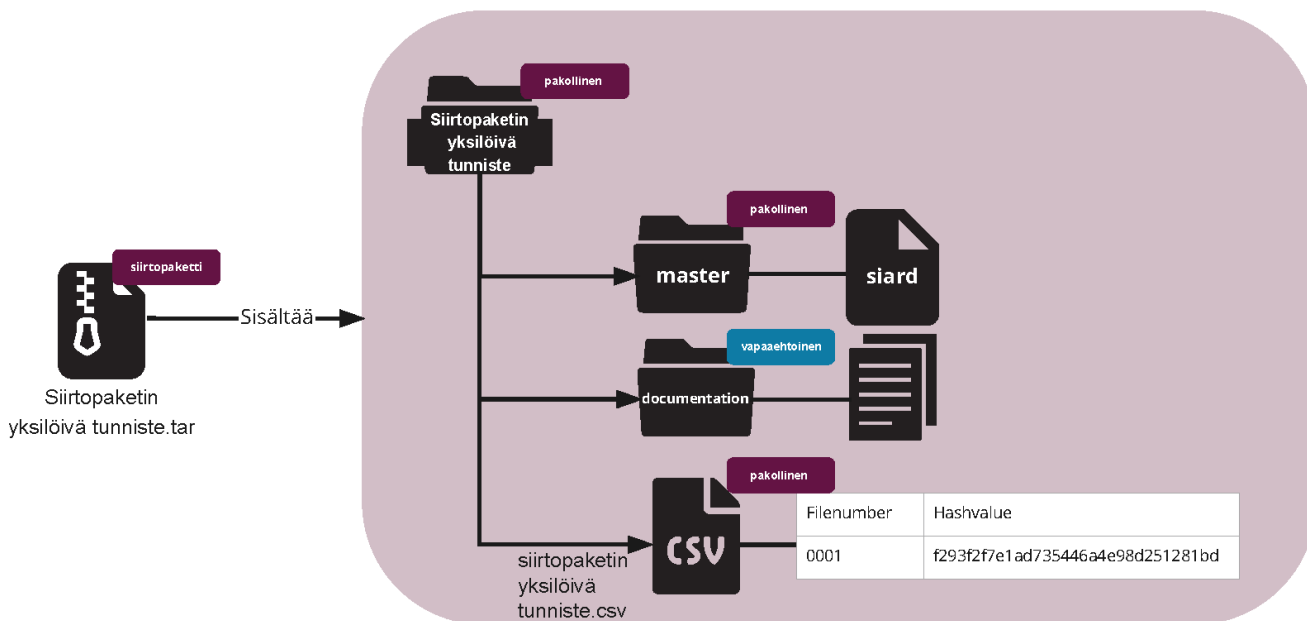
## 6.2. Siirtopaketti

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 nimetään siirtopaketin tunnisteella ja paketoidaan yhdeksi TAR-tiedostoksi.



**Kuva 3.** SIARD-tiedoston sisältävä siirtopaketti vastaa rakenteeltaan datakoosteita sisältävää siirtopakettia, mutta master-kansio voi sisältää vain yhden tiedoston, eikä schemas-kansiota tarvita lainkaan.

## Liitteet

### LIITE 1 Diaariaineistot esimerkkinä rakenteisesta tietoaaineistosta ja diaariaineistojen XML-rakenne

Diaariaineistoilla<sup>2</sup> tarkoitetaan tässä ohjeessa digitaalisessa muodossa olevia asian ja asiakirjojen rekisteröintitietoja. Ohjetta sovelletaan esimerkiksi silloin, kun viranomaisen vanhasta diaarisovelluksesta ollaan luopumassa.

Siirrettävien diaaritietojen kokonaisuus muodostuu varsinaisesta data-aineistosta eli diaaritiedosta sekä rakenteeseen liittyvistä diaarikaavoista ja asia- tai tehtävryhmyksestä, jotka toimitetaan oheisdokumentaationa.

Alkuperäisestä diaarisovelluksesta saatavat datakoosteet tulee muodostaa ensisijaisesti tässä ohjeessa esitetyn rakenteen mukaisesti.

#### Diaariaineistojen XML-rakenne

Diaariaineistoista muodostettaviin datakoosteisiin voidaan tuottaa asia-, toimenpide- ja asiakirjatason tietoelementtejä tässä ohjeessa esitetyn rakenteen mukaisesti.

Rakenteessa on esitetty pakolliset minim tiedot asiatasolla sekä valinnaisia tietoelementtejä asia-, toimenpide- ja asiakirjatasoilla. Laajimmillaan diaariaineistot voivat sisältää tietoelementtejä kaikilta kolmelta tasolta.

Lähtökohtana on, että datakoosteet muodostetaan diaarisovelluksessa tai asiankäsittelyjärjestelmässä olevien tietojen pohjalta. Rekisteröintitietojen tulee olla samalta ajalta kuin käytettyyn rakenteeseen liittyvä dokumentaatio. Tehtäväluokituksen tai asiaryhmyksen selitykset ja mahdolliset muutokset on esitettävä aineiston oheisdokumentaatioissa.

Rakenteessa asiakirjan voi liittää asialle joko toimenpiteen kautta tai ilman toimenpidettä suoraan asialle. Samalla asialla ei kuitenkaan voi olla sekä toimenpiteen kautta että suoraan asialle liitettyjä asiakirjoja. Ainoastaan toinen vaihtoehtoista on mahdollinen. Huom! XML, CSV ja JSON-tiedostoja voidaan siirtää rakenteiset tietoaaineistot - hakemistorakenteen master-kansiossa

Taulukoissa 1, 2 ja 3 esitetään XML-rakenne asiatasolla (taulukko 1), toimenpidetasolla (taulukko 2) ja asiakirjatasolla (taulukko 3). Taulukoiden rivit kuvaavat rakenteeseen liittyvän XML-skeeman elementtejä.

**Taulukko 1.** Asian tiedot

Metatiedon nimi ja tunniste	Selitys	Tietotyyppi	Pakollisuus ja toistettavuus
-----------------------------	---------	-------------	------------------------------

<sup>2</sup> Diaariaineistojen pysyvä säilyttäminen digitaalisessa muodossa ja vastaanottaminen sähköisen arkistoinnin palveluun perustuu Kansallisarkiston seurantapäätökseen [AL/16465/07.01.01.03.02/2016](#) (12.9.2016).

Numero <nr>	Asian rekisteröinnin yhteydessä annettu numero, joka esiintyy osana diaarinumeroa	Vapaa teksti	Pakollinen
Tehtävä <function>	Tehtävän tai asiaryhmän tunniste, joka yleensä esiintyy osana diaarinumeroa	Vapaa teksti	Pakollinen
Vuosi <year>	Diaarinumerossa esiintyvä vuosiluku, joka yleensä merkitsee vuotta, jolloin asia on avattu	Vapaa teksti	Pakollinen
Diaarinumero <caseid>	Asiainumero, diaarinumero tai rekisterinumero, esim. 13/43/1998  Vapaa teksti joka yhdistää kolmen yllä kuvatun elementin (<nr>, <function> ja >year>) tiedot	Vapaa teksti	Pakollinen
Tehtävuokka <functionclass>	Tehtävuokitus on mahdollista esittää monitasoisena hierarkiana	Vapaa teksti	Valinnainen
Asian lähettäjä/vastaanottaja <actor>	Voi olla myös asian vireillepanija, avaaja jne.  Rooli esitetään attribuutissa	Vapaa teksti	Pakollinen
Asian otsikko <title>	Asian otsikko	Vapaa teksti	Pakollinen
Asian kuvaus <description>	Asian otsikkoa tarkentava asian sisällön kuvaus. Voidaan käyttää myös vapaamuotoisen lisätiedon esittämiseen.	Vapaa teksti	Valinnainen
Päivämäärä <casedate>	Asian avaamisen tai päättymisen päivämäärä  Rooli esitetään attribuuttina avaus tai päätös	Päivämäärä vvvv-kk-pv  Datatyyppi määrittää	Pakollinen
Asian suunta <direction>	Kuvaa, onko asia saapuva vai lähtevä	Saapuva, Lähtevä  Arvojoukko	Valinnainen
Julkisuus <publicity>	Tieto siitä, onko asia julkinen vai salassa pidettävä	Julkinen, Osittain salassa pidettävä, Salassa pidettävä	Pakollinen

		Arvojoukko	
Henkilötietoluonne <personaldata>	Tieto siitä, sisältyykö asian tietoihin henkilötietoja vai ei	Sisältää henkilötietoja, Ei sisällä henkilötietoja	Valinnainen, ei toistettava
		Arvojoukko	
Käyttörajoituksen peruste <restriction>	Lainkohta, johon käyttörajoitus perustuu (julkisuuslaki tai erityislaki)	Vapaa teksti	Pakollinen, jos julkisuus on salassa pidettävä tai osittain salassa pidettävä
			Toistettava

**Taulukko 2. Toimenpiteen tiedot**

Metatiedon nimi ja tunniste	Selitys	Tietotyyppi	Pakollisuus ja toistettavuus
Toimenpiteen otsikko <title>	Toimenpiteen otsikko	Vapaa teksti	Valinnainen
Toimenpiteen kuvaus <description>	Toimenpiteen otsikkoa tarkentava kuvaus. Voidaan käyttää myös vapaamuotoisen lisätiedon esittämiseen.	Vapaa teksti	Valinnainen
Toimenpiteen laatija <actor>	Voi olla lähettäjä, vastaanottaja, hyväksyjä jne.  Rooli esitetään attribuutissa Rooli on vapaasti määriteltävissä	Vapaa teksti	Valinnainen, toistettava
Toimenpiteen tyyppi <actiontype>	Toimenpiteet voivat olla esimerkiksi asian välitoimenpiteitä tai lopputoimenpiteitä.	Vapaa teksti	Valinnainen
Toimenpiteen päivämäärä <date>	Saapumispäivämäärä tai laatimispäivämäärä  Rooli attribuuttina Rooli on vapaasti määriteltävissä.	Päivämäärä vvvv-kk-pv Datatyyppi määrittää	Valinnainen, toistettava
Toimenpiteen suunta <direction>	Tieto siitä, onko kyseessä saapuva vai lähtevä toimenpide.	Saapuva, Lähtevä  Arvojoukko	Valinnainen



**Taulukko 3.** Asiakirjan tiedot

<b>Metatiedon nimi ja tunniste</b>	<b>Selitys</b>	<b>Tietotyyppi</b>	<b>Pakollisuus ja toistettavuus</b>
Asiakirjan otsikko <title>	Asiakirjan otsikko	Vapaa teksti	Valinnainen
Asiakirjan kuvaus <description>	Asiakirjan otsikkoa tarkentava sisällön kuvaus. Voidaan käyttää myös vapaamuotoisen lisätiedon esittämiseen.	Vapaa teksti	Valinnainen
Asiakirjan lähettäjä- tai vastaanottaja <actor>	Voi olla myös asiakirjan laatija, hyväksyjä tai allekirjoittaja  Rooli esitetään attribuuttina	Vapaa teksti	Valinnainen, toistettava
Asiakirjatyypin <doctype>	Asiakirjatyypin esim. päätös, hakemus, ilmoitus	Vapaa teksti	Valinnainen
Asiakirjan päivämäärä <date>	Asiakirjan laatimisen, lähettämisen, hyväksymisen tai allekirjoittamisen päivämäärä  Rooli esitetään attribuuttina Rooli on vapaasti määriteltävissä	Päivämäärä vvvv-kk-pv Datatyyppi määrittää	Valinnainen, toistettava
Julkisuus <publicity>	Tieto siitä, onko asiakirja julkinen vai salassa pidettävä	Julkinen, Osittain salassa pidettävä, Salassa pidettävä  Arvojoukko	Valinnainen
Henkilötietoluonne <personaldata>	Tieto siitä, sisältyykö asiakirjan tietoihin henkilötietoja vai ei	Sisältää henkilötieto ja, Ei sisällä henkilötieto ja  Arvojoukko	Valinnainen, ei toistettava
Käyttörajoituksen peruste <restriction>	Lainkohta, johon käyttörajoitus perustuu (julkisuuslaki tai erityislaki)	Vapaa teksti	Pakollinen, jos asiakirja on salassa pidettävä tai osittain salassa pidettävä