



## SÄHKE-hanke

### Tekninen mallintaminen

Versio ja pvm	Laatinut	Tarkpvm	Tarkastanut	Hyvpvm	Hyväksynyt
2.0 / 04.02.2005	Eija Sorakivi	15.02.2005	Markus Merenmies	18.02.2005	Ohjausryhmä



## Muutoshistoria

Versio ja pvm	Laatija	Muutoksen kuvaus
2.0 / 04.02.2005	Eija Sorakivi	TE:n toimittama versio
1.9T/25.01.2005	Anneli Rantanen	KA:n ehdottamat muutokset 21.12.2004
1.8T/21.12.2004	Anneli Rantanen	Versiomuutos ja pieniä korjauksia
1.5T/16.12.2004	Markus Merenmies	KA:n muutoksia. Poistettu säilyttämiseen liittyviä osia
1.4T/14.12.2004	Anneli Rantanen	Versiokorjaus
1.2T/30.11.2004	Anneli Rantanen	Ylä- ja alaotsikot korjattu
1.2T/29.11.2004	Lasse Akselin	Sovitut korjaukset tiff-talletusmuotoon
1.0/31.05.2004	Markus Merenmies	SÄHKE-pilotointi-projektin aloituskokous

## Sisällysluettelo

<b>1 JOHDANTO</b> .....	<b>4</b>
1.1 Sähke-hankkeesta .....	4
1.2 Teknisestä mallista .....	4
1.3 Liittymät eri mallien kesken .....	5
<b>2 TIEDOSTOMUODOT</b> .....	<b>5</b>
2.1 Tekstimuotoon tallennetut asiakirjat .....	5
2.1.1 Rakenteiset asiakirjat .....	6
2.2 PDF-muotoon tallennetut asiakirjat .....	6
2.3 Skannattu asiakirja .....	7
2.3.1 Asiakirjat, joissa ei ole merkityksellisiä värejä .....	7
2.3.2 Asiakirjat, joissa merkityksellisiä värejä .....	7
2.4 Muu sähköinen aineisto .....	8
<b>3 KONVERSIOT JA LAADUNVARMENNUS</b> .....	<b>8</b>
3.1 Konversiotavat .....	8
3.2 Laadunvarmennus .....	9
3.2.1 Metatietojen laadunvarmennus .....	9
3.2.2 Sähköisen aineiston laadunvarmennus .....	9
3.2.3 Siirtotiedoston laadunvarmennus .....	9
<b>4 SIIRTOTIEDOSTON RAKENNE</b> .....	<b>9</b>
4.1 Siirtotiedoston metatiedot .....	11
4.1.1 Siirrettävän aineiston tietovälinekohtaiset tiedot (siirto.xml) .....	12
4.1.2 Aineiston sisällön kuvaus (jarjkuv.xml) .....	12
4.2 Aineiston metatiedot (metatied.xml) .....	13
4.2.1 Viittaukset sähköisiin asiakirjoihin .....	13
4.3 Keskeneneräiset asiat .....	15
<b>5 SIIRTOTIEDOSTON VIIMEISTELY</b> .....	<b>15</b>
5.1 Sähköisen allekirjoituksen validointi .....	16
5.2 Suojaaminen .....	16
5.3 Siirtotiedoston tekninen validointi .....	16



---

<b>6 SIIRTOTIEDOSTON KOKOAMINEN.....</b>	<b>16</b>
<b>7 AINEISTON SIIRTO ARKISTOLAITOKSEEN .....</b>	<b>16</b>

### **Liitteet:**

- Liite 1: SÄHKE-metatietomäärittelyn XML Schema
  - a: aineiston\_metatiedot.xsd
- Liite 2: Siirtotiedoston metatietokuvaukset
- Liite 3: Siirtotiedoston metatietojen tietomalli (graafi)
- Liite 4: Siirtotiedoston metatietojen XML Schemat
  - a: siirto.xsd
  - b: jarjestelmakuvaukset.xsd
- Liite 5: Esimerkit siirtotiedoston metatiedoista
  - a: siirto.xml
  - b: jarjkuv.xml
  - c: metatied.xml

### **Lähteet:**

- [1] H Sivonen: Kansallisarkistolle tehty tekninen selvitys tiedostoformaateista sekä pakkaus- ja pake-  
tointitavoista, Kansallisarkisto, 2002
- [2] Tanskan arkistolaitos: Vejledning til cirkulaere om statslige myndigheders aflevering af elektroniske  
arkivsystemer til Statens Arkiver, <http://www.sa.dk/sa/statamtkom/arklov/nyafleirk.htm>



## Tekninen mallintaminen

### 1 Johdanto

#### 1.1 Sähke-hankkeesta

Kansallisarkiston SÄHKE-hanke on sähköisten tietoaaineistojen käsittelyn ja pitkäaikaissäilytyksen kehittämishanke, jossa tarkastellaan asiankäsittelyjärjestelmiä arkistotoimen ja pitkäaikaissäilytyksen näkökulmasta. Hankkeessa on kehitetty viranomaisten integroiduille asianhallintajärjestelmille asetettavia asiakirjahallinnon ja arkistotoimen vaatimuksia.

Nämä vaatimukset on kuvattu kolmessa eri mallissa: toiminnallinen, tekninen ja abstrakti. Toiminnallisessa mallissa on kuvattu asian ja asiakirjan elinkaaren hallinnan näkökulmasta asiankäsittelyjärjestelmälle asetettavia vaatimuksia. Abstrakti mallintaminen on keskittynyt erityisesti asiankäsitteilykokonaisuudessa käsiteltävän aineiston metatietojen ja niiden rakenteiden määrittelyyn. Työn tuloksena on muodostunut liitteenä olevat tietomallit, jotka sisältävät kaikki tunnistetut metatiedot. Teknisessä mallissa on kuvattu ne tekniset reunaehdot, jotka sähköisen aineiston tulee täyttää jotta sen arkistointi ja siirto arkistolaitokseen olisi mahdollista.

#### 1.2 Teknisestä mallista

SÄHKE-hankkeessa on tuotettu kuvaus arkistolaitokseen toimitettavan siirtotiedoston rakenteesta ja sisällöstä. Siirtotiedostolla tarkoitetaan asiankäsitteilyjärjestelmän tuottamaa arkistolaitokseen siirrettävää aineistoa, jossa ainakin sen rekisteriosa on tallennettu sähköisesti. Varsinainen aineisto voi sisältää sekä sähköisiä että fyysisiä asiakirjoja.

Siirtotiedoston avulla on määritelty yhtenäinen tapa tuottaa arkistolaitoksen palvelujärjestelmään<sup>1</sup> tallennettava aineisto. Sen avulla asiankäsitteilyjärjestelmien tuottama aineisto myös vastaanotetaan ja sitä käsitellään yhtenäisessä muodossa arkistolaitoksessa. Jatkossa aineisto on käytettävissä arkistolaitoksen palvelujärjestelmässä.

Teknisen mallintamisen tuloksena on määritelty tekniset reunaehdot siirtotiedostolle. Siirrettävän aineiston sisältämien asiakirjojen osalta on huomioitu ne tiedostomuodot, jotka arkistolaitos ottaa vastaan. Siirtotavoista on kuvattu aineiston siirtoprosessi fyysistä tietovälinettä käyttäen arkistolaitokseen nykyisten ohjeiden mukaisesti. Siirtotiedostossa kuvattu aineisto voi olla

<sup>1</sup> Palvelujärjestelmä on sähköisen säilytyksen operatiivinen järjestelmä, jossa säilytetään sähköisiä aineistoja arkistolaitoksessa.



myös osaksi tai kokonaan tallennettu fyysiselle tietovälineelle esim. mikrofilmille.

Luvussa kaksi on kuvattu tiedostomuodot, joita arkistolaitos ottaa vastaan. Luvussa kolme on kuvattu pääkohdat konversiotavoista ja konversion tuloksena muodostuvien tiedostojen laadunvarmennuksesta. Siirtotiedoston koostaminen on esitetty yleisellä tasolla luvussa neljä. Luvussa viisi on kuvattu siirtotiedoston sisältämät metatiedot, joiden tarkempi tietosisältö on kuvattu tämän asiakirjan liitteissä. Lopuksi luku kuusi sisältää kuvaukset siirtotavoista, joilla aineisto voidaan siirtää arkistolaitokseen.

Tässä tekninen mallintaminen –vaiheen lopputuloksena tuotetussa asiakirjassa kuvatut tekniset ohjeistukset ja ratkaisut perustuvat pääasiallisesti eri aihealueiden (esim. tiedostomuodot, pakkaustavat) osalta tehtyihin selvityksiin [1].

### 1.3 Liittymät eri mallien kesken

Sähke-määrittelyissä kuvatuilla kolmella mallilla on selkeät liittymäkohdat toisiinsa ja niitä voidaan tarkastella joko erikseen tai yhdessä. Toiminnallinen malli kuvaa sen, miten abstraktissa tietomallissa kuvatut metatiedot syntyvät asiankäsitteilyprosessissa. Tekninen malli vastaavasti kuvaa sen, miten metatiedot ja asiakirjat voidaan siirtää arkistolaitokseen.

Toiminnallinen ja abstrakti malli kuvaavat molemmat asiakirjallisen aineiston syntyyn ja käsittelyyn liittyviä asioita. Toiminnallisessa mallissa kuvataan vaatimukset prosessin näkökulmasta ja abstraktissa mallissa prosessissa syntyvän tiedon näkökulmasta. Näitä kahta mallia voidaan tarkastella yhtäaikaisesti, sillä toiminnallisessa mallissa viitataan toimintojen yhteydessä metatietoihin, joiden tarkempi kuvaus on abstraktissa mallissa.

## 2 Tiedostomuodot

Siirtotiedoston sisältämä sähköinen aineisto tallennetaan seuraavissa luvuissa esitettyjen tiedostomuotojen mukaisesti. Lisäksi aineisto ryhmitellään tiedostomuotokohtaisesti hakemistoihin \tiff ja \text.

### 2.1 Tekstimuotoon tallennetut asiakirjat

Tekstimuotoon tallennettu asiakirja on tuotettu tekstieditorilla (esim. Emacs, Notepad). Se ei sisällä merkkeihin liitettyjä tyylejä (esim. alleviivaus, lihavointi). Toisaalta myös tekstinkäsittelyohjelmalla tuotettu asiakirja voi olla tekstimuotoinen, jos se on tuotettu ja tallennettu tekstieditorilla tuotetun asiakirjan tavoin.



Tekstimuotoon tallennetut asiakirjat liitetään siirtotiedostoon tekstimuotoisina. Merkkikoodauksena käytetään Unicode UTF-8:aa.<sup>2</sup>

Käytettävät kontrollimerkit ovat:

Kontrollimerkki	Merkin järjestyksennumero
CR (Carriage Return)	13
LF (Line Feed)	10
TAB (Horizontal Tab)	09

Rivinvaihtomerkinä hyväksytään UNIX- (LF) ja Windows-muoto (CRLF).

### 2.1.1 Rakenteiset asiakirjat

Rakenteisten asiakirjojen erona pelkkiin tekstiasiakirjoihin on se, että tekstin joukkoon on tallennettu asiakirjan sisältöä ja rakennetta kuvaavia elementtejä esim. XML-rakenteen mukaisesti. Tällä hetkellä arkistolaitos ottaa vastaan vain pelkkiä tekstiasiakirjoja eli asiakirjoja, joita ei ole tallennettu rakenteiseen muotoon. Poikkeustapauksissa pysyvästi säilytettävät asiakirjat voidaan arkistolaitoksen luvalla säilyttää HTML-muodossa.

### 2.2 PDF-muotoon tallennetut asiakirjat

Arkistolaitos ei toistaiseksi vastaanota PDF-tiedostomuotoon tallennettuja asiakirjoja. PDF/A tiedostomuotoa voidaan käyttää asiankäsittelyjärjestelmässä aktiivinaikaisena tallennusmuotona jos voidaan varmistaa, että siitä voidaan siirtotiedostoa varten muodostaa arkistolaitoksen ohjeiden mukainen tiedosto. PDF/A-tiedostomuotoa käytettäessä tulee huolehtia samoista asiakirjalliseen tietoturvaan liittyvistä seikoista kuin alkuperäisen asiakirjan tapauksessa.

<sup>2</sup> UTF-8 on Unicoden vaihtelevanpituinen koodaustapa. Sen etuna on osittainen yhteensopivuus vanhempien järjestelmien kanssa, jotka käsittelevät merkkejä kahdeksanbittisinä tavuina. UTF-8 on rakennettu siten, että ASCII-merkistöön kuuluvat merkit (kirjaimet A-Z) säilyvät siinä samoina kuin ASCII:ssa, ja muut merkit koodataan vaihtelevan pituisena oktettijonona.

Koodipaikasta riippuen yksi merkki vie tallennettuna tilaa yhdestä neljään tavua. Tavujen määrä riippuu koodipaikasta.



## 2.3 Skannattu asiakirja

Paperiasiakirjat voidaan muuntaa sähköiseen muotoon skannaamalla, jolloin siirtotiedostoa muodostettaessa on huomioitava skannattujen asiakirjojen tiedostomuoto ja tallennustapa.

### 2.3.1 Asiakirjat, joissa ei ole merkityksellisiä värejä

Tähän ryhmään kuuluvat asiakirjat, jotka sisältävät pääsääntöisesti tekstiä tai niiden sisältämällä väreillä ei ole merkitystä sisällön tulkitsemiseksi (esim. asiakirja sisältää värillisen yritystunnuksen). Luovuttava viranomainen päättää, onko skannattavassa asiakirjassa esiintyvillä väreillä merkitystä. Päätöstä tehtäessä on kiinnitettävä huomiota asiakirjan tietosisällön säilymiseen

Skannatut asiakirjat tallennetaan CCITT/ITU-T Group 4 -pakattuun tif v. 6.0 -muotoon, joka mahdollistaa asiakirjan tallentamisen monisivuisena eli yhteen tiedostoon.

Asiakirja skannataan paperilta 200-400 ppi (pixels/inch) -tarkkuudella kaksiväriseksi (mustavalko) bittikartaksi. Tarkkuustaso on valittava siten, että asiakirjan teksti on tulkittavissa yksiselitteisesti. Tarkastaminen on tehtävä silmämääräisesti.

Asiakirjasta tallennetaan siirtotiedostoon sekä häviöttömästi pakattu mustavalkobittikartta että mahdollisesti siitä tuotettu pelkkää tekstiä sisältävä vedos tekstihakua varten. Tekstimuotoinen vedos tuotetaan OCR-ohjelmalla skannatusta kuvasta tai tallentamalla asiakirja tekstimuotoon (UTF-8) omaksi sivunumeroiduksi tiedostokseen.

### 2.3.2 Asiakirjat, joissa merkityksellisiä värejä

Tähän ryhmään kuuluvat asiakirjat, jotka sisältävät sisällön tulkitsemisen kannalta tärkeitä värejä tai harmaasävyjä. Asiakirja skannataan RGB-värillisenä (8 bittiä/kanava, yhteensä 24 bit eli True Color) ja tallennetaan TIF-tiedostomuotoon (tarkkuus 200-400 ppi) käyttäen LZW-pakkausta. Sähköisestä asiakirjasta on aina varmistettava silmämääräisesti, että tarkkuustaso on riittävä ja tarpeelliset värit ovat mukana.

Oletusarvoisesti asiakirjoista tehdään mustavalkoiset bittikartat. Asiankäsittelyprosessissa on määriteltävä ne työvaiheet, joissa tieto merkityksellisten värien säilyttämisestä tallennetaan järjestelmään asianomaisten asiakirjojen osalta. Mikäli asiakirja sisältää merkityksellisiä värejä tallennetaan tieto tästä asiakirjan formaatti metatiedon *merkityksellinen väri* elementtiin.

Asiakirjasta tuotetaan mahdollisesti myös tekstivedos tekstihakua varten.



Tiedostojen tallentaminen siirtotiedostoon ja niihin kohdistuvat viittaukset tehdään kuten kohdassa 4.2.1 ”Viittaukset sähköisiin asiakirjoihin”.

## 2.4 Muu sähköinen aineisto

Mikäli asiakirjallinen aineisto sisältää asiakirjoihin liittyviä tai erillisiä sähköisessä muodossa olevia valokuvia tulee näiden osalta toimia konversion yhteydessä siten, että valokuvan laatu säilyy riittävän hyvänä. Tallennusmuotona käytetään yhtenevää muotoa asiakirjojen kanssa siten, etteivät pakauksesta johtuvat häviöt ole havaittavissa häiritsevästi. Laadun varmistus on tehtävä silmämääräisesti.

Mikäli asiakirjallinen aineisto sisältää äänitiedostoja, on niiden pitkäaikais- ja pysyväistallentamisessa käytettävä pakkaamatonta WAV-formaattia (ns. CD-laatua, 24 bit, 48 kHz).

SÄHKE-määritys ei ota kantaa tässä yhteydessä rekistereihin ja muihin tietokantoihin liittyvän tiedon tallentamiseen tai siirtomuotoon.

## 3 Konversiot ja laadunvarmennus

Arkistolaitoksen vastaanottamat tiedostomuodot eivät todennäköisesti vastaa niitä tiedostomuotoja, joissa asiakirjoja tuotetaan ja hallitaan niiden aktiivivaiheena. Tällöin siirtotiedostoa varten asiakirja on konvertoitava. Asiakirjat konvertoidaan johonkin arkistolaitoksen hyväksymistä tiedostomuodoista.

Konversiomenetelmiä ja niihin liittyviä ohjelmistoja on olemassa useita, joten tässä kuvauksessa ei ole mahdollista esittää kattavasti tapoja, joilla siirrettävä aineisto konvertoidaan. Siirtotiedoston vastaanottajan ja sitä käsittelevän järjestelmän näkökulmasta on kuitenkin erittäin tärkeää, että konversion tulokset vastaavat sisällöltään alkuperäistä aineistoa ja niiden tallentaminen sekä käsittely arkistolaitoksen säilytys- ja palvelujärjestelmässä on mahdollista.

### 3.1 Konversiotavat

Siirrettävä aineisto on tallennettava tämän ohjeen mukaisesti teksti- ja/tai TIFF-muotoon.

Aineiston konvertoiminen TIFF-muotoon tapahtuu joko skannaamalla (paperiasiakirja) tai ohjelmallisesti tallentamalla (sähköinen asiakirja, esim. PDF-muodosta tuotetaan TIFF-muoto).

Kun asiakirja konvertoidaan tekstimuotoon, useita sen ulkoasuun ja tekstin aseteluun liittyviä määrittelyjä häviää. Esimerkiksi tekstinkäsittelyohjelman avulla tuotetut tyyli- ja taulukkoasetukset eivät tallennu tekstimuotoiseen



tiedostoon. Jos nämä konversion aiheuttamat hävikit vaikuttavat oleellisesti asiakirjan sisällön tulkitsemiseen, on asiakirja konvertoitava kuvaksi eli TIF-muotoon.

Asiakirjan konversiosta pitää tallentaa tieto asiakirjan metatietoihin (säilytyshistoria).

### 3.2 Laadunvarmennus

Siirtoaineiston laadunvarmistukseen liittyy kolme eri osa-aluetta: metatietorakenteiden oikeellisuus sekä sähköisessä muodossa olevan aineiston ja siirtotiedoston teknisen laadun varmennus.

#### 3.2.1 Metatietojen laadunvarmennus

Metatietojen laadunvarmennuksella tarkoitetaan varsinaisen aineiston ja siirtotiedostoon liittyvien metatietojen rakenteen tarkastamista (XML-tiedostot). Jokaista metatietomäärittystä varten on tehty oma XML Schema, joiden avulla varmistetaan, että aineisto on määritellyn rakenteen mukaista.

XML-rakenteiden tarkastamiseen tulee käyttää validointiohjelmaa, joka tukee XML Schema versio 1.0:n mukaisia rakenteita.

#### 3.2.2 Sähköisen aineiston laadunvarmennus

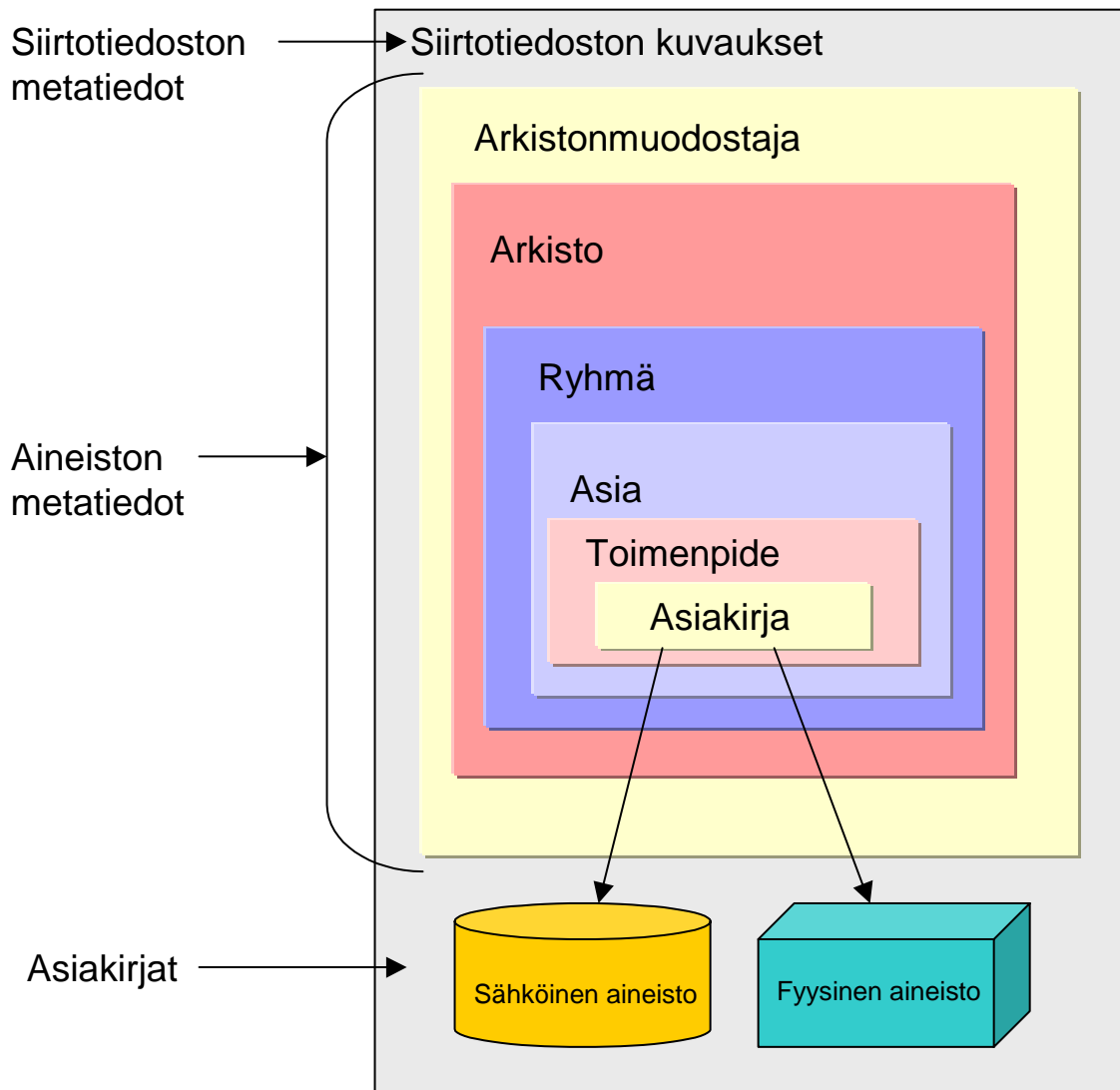
Konvertoidut sähköiset asiakirjat on tarkastettava ennen kuin ne voidaan liittää osaksi siirtotiedostoa. Tarkastaminen tehdään pistokokein, kunnes saavutetaan riittävä varmuus konversion onnistumisesta.

#### 3.2.3 Siirtotiedoston laadunvarmennus

Siirtotiedoston teknisellä laadunvarmennuksella varmistetaan, että siirrettävä aineistokokonaisuus on muodostettu teknisesti oikein. Siirrettävän aineiston osalta varmistetaan, että se on koottu ohjeiden mukaisesti ja sen sisältö on oikea. Tällöin varmistetaan myös se, että siirtotiedosto voidaan avata ja sen sisältö on käytettävissä arkistolaitoksessa.

## 4 Siirtotiedoston rakenne

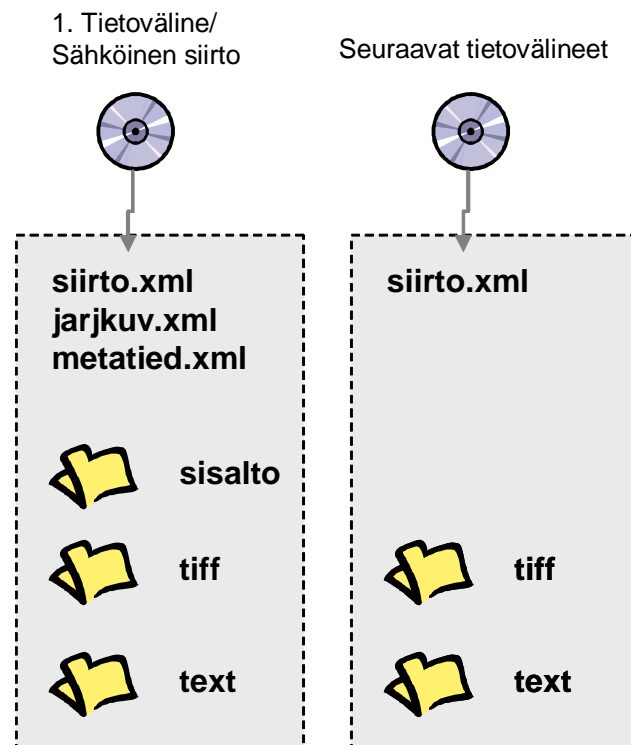
Siirtotiedoston rakenne loogisella tasolla on esitetty kuvassa 1 ”Siirtotiedoston looginen rakenne”. Siirtotiedosto muodostuu joukosta metatietoja ja aineistosta (asiakirjoista), johon viitataan aineiston metatiedoista. Koko siirrettävän aineistokokonaisuuden kuvailuun käytettäviä metatietoja kutsutaan puolestaan siirtotiedoston metatiedoiksi.



**Kuva 1.** Siirtotiedoston looginen rakenne.

Aineisto voidaan siirtää sähköisesti tietoliikenneyhteyden välityksellä, fyysisellä tietovälineellä tai molempien avulla. Kun aineisto tallennetaan ja siirretään fyysisellä tietovälineellä, siirtoaineisto jaetaan yhdelle tai useammalle tietovälineelle aineiston ja tietovälineen koon asettamien vaatimusten mukaisesti.

Siirrettävän aineiston tekninen rakenne on esitetty kuvassa 2 ”Siirtotiedoston tekninen rakenne”. Jos aineisto jaetaan ja tallennetaan useammalle tietovälineelle, seuraavat tietovälineet sisältävät aina aineistoa kuvaavan tiedoston (”siirto.xml”) ja hakemistorakenteen mukaisesti tallennetun osan siirrettävää aineistosta.



Kuva 2. Siirtotiedoston tekninen rakenne

Kun aineisto siirretään kokonaan yhdellä tietovälineellä tai kertasiirtona sähköisesti tietoliikenneyhteyden välityksellä, siirtotiedoston rakenne on sama kuin ensimmäisen tietovälineen rakenne. Jos tietoliikenteen välityksellä tapahtuva siirto muodostuu useammasta lähetettävästä kokonaisuudesta, muut kokonaisuudet kuvataan kuten kuvassa 2 esitetty ”seuraavat tietovälineet”.

Siirtotiedoston metatiedot tallennetaan XML-rakenteisina ennalta sovittujen tiedostonimien mukaisesti (siirto.xml ja jarjkuv.xml). Aineiston metatiedot on tallennettu erilliseen XML-rakenteiseen tiedostoon (metatied.xml) ja varsinainen asiakirja-aineisto tallennetaan tiedostomuotoiksi hakemistoihin.

Siirtotiedoston rakenne on kuvattu tarkemmin seuraavissa luvuissa ja siirtotiedoston metatiedot on kuvattu liitteissä 2 ”Siirtotiedoston metatietokuvaukset” ja 3 ”Siirtotiedoston metatietojen tietomalli”.

#### 4.1 Siirtotiedoston metatiedot

Siirtotiedoston tarkastamista ja käsittelyä varten tarvitaan siirrettävään kokonaisuuteen liittyviä kuvailutietoja. Lisäksi niitä tarvitaan siirtotiedoston sisällyksen tunnistamiseen, aineiston purkamiseen ja tallentamiseen palvelujärjestelmään.



#### 4.1.1 Siirrettävän aineiston tietovälinekohtaiset tiedot (siirto.xml)

Siirrettävän aineiston tietovälinekohtaiset tiedot tallennetaan ”siirto.xml” –nimiseen tiedostoon. Jos siirtotiedosto on tallennettu useammalle kuin yhdelle tietovälineelle, siirto.xml -tiedosto sisältyy niihin jokaiseen.

Siirto.xml –tiedostoon kuvataan seuraavaa:

- Siirrettävän aineiston tunniste ja mahdollinen viittaus edellisiin siirtotiedostoihin,
- siirtotiedoston tuottajan tiedot sekä tietojärjestelmän nimi, josta aineisto siirretään,
- aineiston ajallinen kattavuus,
- metatietojen XML Schema -määrittelyn nimi,
- siirtotiedoston metatietojen tekniset tiedot ja
- aineiston tai ko. tietovälineen sisältö (hakemistot ja tiedostot).

#### 4.1.2 Aineiston sisällön kuvaus (jarjkuv.xml)

Siirtotiedostoon tallennetaan myös kuvauksia asiankäsittelyjärjestelmästä (esim. systeemyödokumentaatio, näyttökuvia), jolla aineisto on tuotettu. Näiden tietojen avulla kuvataan aineiston alkuperäinen toimintaympäristö. Järjestelmäkuvaukset tallennetaan ”jarjkuv.xml” -nimiseen tiedostoon.

Metatietojen avulla kuvataan säilytettävää aineistoa tuottavasta järjestelmästä ainakin seuraavat tiedot:

- Järjestelmän yleiskuvaus ja käyttötarkoitus,
- otteita järjestelmän käyttöliittymästä ja tulosteista sekä
- systeemyöasiakirjat.

Systeemyöasiakirjoista siirtotiedostoon liitetään käyttöohjeet, ja mikäli katsotaan tarpeelliseksi myös määrittely-, suunnittelu- ja toteutusasiakirjat sekä mahdolliset testaus- ja käyttöönottoasiakirjat. Systeemyöasiakirjat ja järjestelmää kuvaavat näkymät kuvaillaan SÄHKE-metatietomäärittelyn mukaisen asiakirjan metatietojen avulla ja tarvittaessa aineiston julkisuus voidaan määrittellä metatietomäärittelyssä olevalla julkisuusrakenteella. Metatietojen ja asiakirjojen välisten viittausten muodostuminen kuvataan kohdassa 4.2.1 ”Viittaukset sähköisiin asiakirjoihin”.

Jos aineisto tallennetaan useammalle kuin yhdelle tietovälineelle tai lähetetään useammassa kuin yhdessä erässä, ”jarjkuv.xml” -tiedosto tallennetaan aina vain ensimmäiselle tietovälineelle/toimituserään.



## 4.2 Aineiston metatiedot (metatied.xml)

Aineiston metatiedot kuvataan SÄHKE-metatietomäärittelyn rakenteen mukaisesti. Metatiedot tallennetaan ”metatied.xml” -nimiseen tiedostoon.

Jos siirtotiedosto on tallennettu useammalle kuin yhdelle tietovälineelle, metatied.xml -tiedosto tallennetaan vain ensimmäiselle tietovälineelle.

### 4.2.1 Viittaukset sähköisiin asiakirjoihin

Viittaukset asiakirjojen metatiedoista asiakirjatiedostoihin muodostetaan metatietorakenteella, johon tallennetaan tiedostomuotokohtaiset hakemistotunnisteet.

TIFF-muotoon konvertoitu asiakirja ja vastaava tekstivedos voivat koostua joko yhdestä tai useammasta tiedostosta. Mikäli kyseessä on monisivuinen TIFF-formaatti, eli hakemistoon tallennetaan vain yksi tiedosto, viittaus asiakirjoihin tehdään seuraavasti:

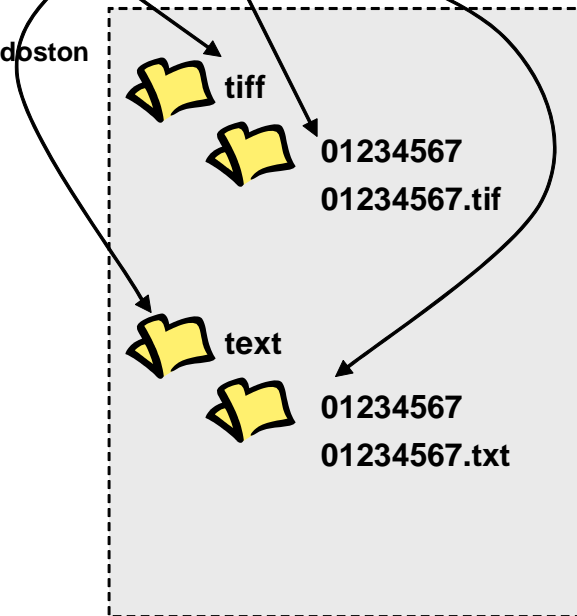
- Asiakirjan teksti- ja kuvatiedostot tallennetaan omaan rakenteeseen (hakemistoon), joka nimetään asiakirjan yksilöivän identifiointitunnuksen mukaisesti. Tiedostot sisältävä hakemisto tallennetaan siirtotiedostossa tiedostomuotokohtaiseen hakemistoon.
- Hakemistonimien pituudeksi suositellaan 8 merkkiä. Jos tämä ei ole mahdollista asiakirjan identifiointitunnuksesta muodostettujen hakemistojen kohdalla, on huolehdittava, ettei hakemistopolun kokonaispituus ylitä käyttöjärjestelmän tai tallevälineen asettamia rajoituksia.
- Tiedostot nimetään vakiodusti ”<asiakirjan identifiointitunnus>.tif”. Tiedostonimi on pituudeltaan 8+3 merkkiä.
- Asiakirjan metatietojen ”Sijaintipaikka”-nimiseen metatietorakenteeseen (storageLocation) tallennetaan tiedostomuoto (tiff tai txt). Jokaiselle tiedostoformaatile tehdään oma to, jonne vastaavat tiedostot tallennetaan.

Esimerkki yhtenä tiedostona tallennetun asiakirjan tiff- ja tekstivedostiedostojen viittaus- ja tallennustavasta on esitetty kuvassa 3 ”Esimerkki yhtenä tiedostona tallennetun asiakirjan viittaus- ja tallennustavasta”.

**Asiakirjan metatiedot:**

```
...  
<storageLocation type="text">text\01234567</storageLocation>  
<storageLocation type="tiff">tiff\01234567</storageLocation>  
...
```

**Tiedostot siirtotiedoston rakenteessa:**

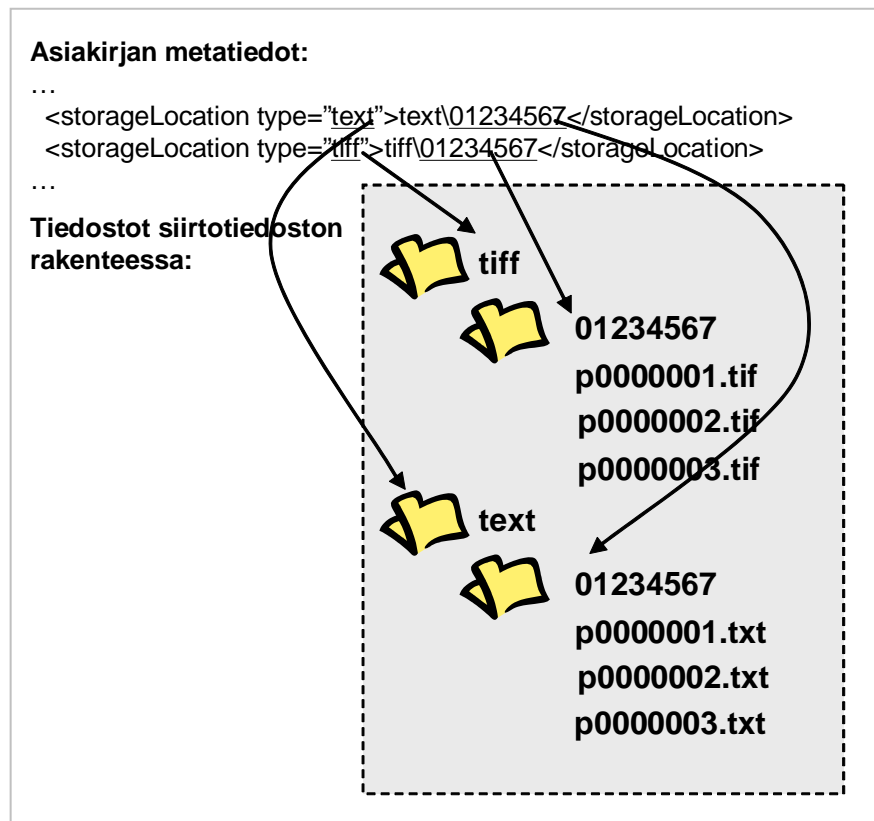


**Kuva 3.** Esimerkki yhtenä tiedostona tallennetun asiakirjan viittaus- ja tallennustavasta.

Jos kyseessä on yksisivuinen TIFF-formaatti, asiakirja tallennetaan tarvittaessa useina tiedostoina. Tällöin viittaus asiakirjoihin tehdään seuraavasti:

- Asiakirjan teksti- ja kuvatiedostot tallennetaan omaan rakenteeseen (hakemistoon), joka nimetään asiakirjan yksilöivän identifiointitunnuksen mukaisesti. Tiedostot sisältävä hakemisto tallennetaan siirtotiedostossa tiedostomuotokohtaiseen hakemistoon.
- Hakemistonimien pituudeksi suositellaan 8 merkkiä. Jos tämä ei ole mahdollista asiakirjan identifiointitunnuksesta muodostettujen hakemistojen kohdalla, on huolehdittava, ettei hakemistopolun kokonaispituus ylitä käyttöjärjestelmän tai tallevälineen asettamia rajoituksia.
- Sivukohtaiset nimetään siten, että nimi alkaa kirjaimella ”p” ja loput merkit muodostavat sivunumeron, esim. ”p0000001.tif”. Tiedostonimi on pituudeltaan 8+3 merkkiä.
- Asiakirjan metatietojen ”Sijaintipaikka”-nimiseen metatietorakenteeseen (storageLocation) tallennetaan viittaus hakemistoon, johon sivukohtaiset tiedostot on tallennettu. ”Sijaintipaikka” – metatietorakenteeseen tallennetaan myös tiedostomuoto (tiff tai text). Jokaiselle tiedostoformaatile tehdään oma hakemisto, jonne vastaavat tiedostot tallennetaan.

Esimerkki kolmesivuisen asiakirjan tiff- ja tekstivedostiedostojen viittaus- ja tallennustavasta on esitetty kuvassa 4 ”Esimerkki useana tiedostona tallennetun asiakirjan viittaus- ja tallennustavasta”.



**Kuva 4.** Esimerkki yhtenä tiedostona tallennetun asiakirjan viittaus- ja tallennustavasta.

#### 4.3 Keskeneräiset asiat

Aineiston siirtoajankohtana keskeneräisenä olevat asiat tallennetaan myös siirtotiedostoon. Asioille tehdään tekninen päätöstoimenpide, joka ei muuta asian tilaa (asian tila = ”kesken”). Siirtotiedostoon tallentuu asiasta sen hetkiset toimenpiteet ja toimenpiteisiin liitetyt asiakirjat.

#### 5 Siirtotiedoston viimeistely

Siirtotiedoston viimeistelyllä tarkoitetaan tehtäviä, joiden avulla varmistetaan, että aineisto on valmis siirrettäväksi. Samalla aineisto sinetöidään sähköisellä allekirjoituksella (mieluiten ns. roolivarmenne), mikäli aineisto toimitetaan tietoliikenneyhteyden välityksellä. Yleensä nämä tehtävät hoitaa asiakirjahallinnosta vastaava henkilö.



## 5.1 Sähköisen allekirjoituksen validointi

Viranomaisen tulee huolehtia siitä, että siirtotiedoston sähköiseen allekirjoittamiseen mahdollisesti käytetty menetelmä on arkistolaitoksen ohjeiden mukainen. Sähköisen allekirjoitusprosessin oikeellisuus on varmennettava ennen aineiston siirtämistä.

## 5.2 Suojaaminen

Siirtotiedoston muodostamis- ja siirtoprosessin aikana on tarkoin huolehdittava, etteivät tiedot kulkeudu sellaisiin tietojenkäsittely-ympäristöihin, joissa tietojen alkuperäisyys ja eheys tai mahdollinen salassapito voisi vaarantua.

Mikäli aineisto sisältää salassa pidettävää tietoa, aineisto on suojattava siirtämisen ajaksi. Fyysisten tietovälineiden osalta voidaan käyttää erilaisia perinteisiä sinetöintimenetelmiä. Tietoliikenneyhteyden yli siirrettävä aineisto on puolestaan salakirjoitettava sellaisin menetelmin, joita edellytetään aineistoon sisältyvien tietojen salaisimpien osien siirtoon. Salaamiseen voidaan käyttää HST-infrastruktuurin mukaisesti vastaanottajan julkista avainta.

Mikäli käytetään kaksisuuntaista salakirjoitusalgoritmia, siirtotiedoston luettavaksi saamisen edellyttämä salakirjoitusavain toimitetaan erikseen sovitun menettelyn mukaisesti siirtotiedostosta erillään.

## 5.3 Siirtotiedoston tekninen validointi

Siirtotiedoston teknisessä validoinnissa tarkastetaan muodostettu siirtotiedosto (eli tiedosto on avattavissa ja sisältö on oikea).

## 6 Siirtotiedoston kokoaminen

Ennen aineiston siirtoa, siirtotiedoston kokoamisen lisäksi, aineisto voidaan **pakata**, jolloin siirtotiedoston koko on mahdollisimman pieni. Kokoamiseen liittyvillä toimenpiteillä voidaan edistää aineiston sisällön muuttumattomuutta siirron aikana. Paketoinnissa täytyy huolehtia siitä, että myös aineiston hakemistorakenne tallennetaan.

Paketoinnissa voidaan käyttää arkistolaitoksen ohjeiden mukaista pakkaustapaa.

## 7 Aineiston siirto arkistolaitokseen

Siirtotiedosto voidaan toimittaa arkistolaitokseen toistaiseksi vain fyysisillä tietovälineillä.



Arkistolaitos hyväksyy toistaiseksi seuraavat tietovälineet sähköisten aineistojen siirtovälineiksi:

- 8 mm nauhakasetit
- 3,81 mm DDS (Digital Data Storage) nauhakasetit (DDS/DC tai DDS-2, DDS-3 tai DDS-4- formaatti)
- CD-R -levy

Lisäksi sopimuksen mukaan voidaan tapauskohtaisesti erikseen hyväksyä mm. seuraavat siirtovälineet

- DVD – levy
- muut nauhatyypit

Lisätietoja:

AIIM International and NPES Partner to Standardize use of PDF for Document Archive and Preservation:

[http://www.aiim.org/article\\_edoc.asp?ID=25517](http://www.aiim.org/article_edoc.asp?ID=25517) ja

[http://www.aiim.org/article\\_pr.asp?ID=24680](http://www.aiim.org/article_pr.asp?ID=24680)

XML Schema: <http://www.w3.org/XML/Schema>