

**SÄHKE2**  
Siirto-ohje

v. 1.0.2  
2.10.2009

1 (23)

**SÄHKE2-Siirto-ohje**

**Sovellusohje SÄHKE2-määräyksen mukaisen siirtotiedoston muodostamiseksi**

**SÄHKE2**  
Siirto-ohje

v. 1.0.2  
2.10.2009

2 (23)

<b>1 JOHDANTO</b> .....	<b>3</b>
1.1 Termistö.....	3
1.2 Liittymät muihin määräyksiin ja ohjeisiin.....	3
<b>2 YLEISTÄ SIIRTOPROSESSISTA</b> .....	<b>4</b>
2.1 Muodostuva aineisto.....	4
2.2 Prosessin vaiheet .....	4
<b>3 LUOVUTUSPAKETIN MUODOSTAMINEN</b> .....	<b>5</b>
3.1 Aineiston valmistelu .....	5
3.1.1 Konversio pitkäaikaissäilytettävään muotoon .....	5
3.1.2 Immateriaalioikeuksiin liittyvät rajoitukset .....	6
3.1.3 Validointi ja laadunvarmistus .....	7
3.1.4 Asiakirjojen yksilöinti OID/URN -tunnuksella .....	8
3.1.5 Salatun aineiston käsittely .....	9
3.2 Luovutuspaketin tiedostorakenne .....	9
3.3 Metatiedot ja niiden esittäminen XML-siirtotiedostona .....	10
3.3.1 Siirtotiedoston rakenne .....	11
3.3.2 SÄHKE1 -määräyksen mukainen siirtotiedosto .....	11
3.3.3 Siirron metatiedot.....	12
3.3.4 Siirrettävän kokonaisuuden metatiedot.....	13
3.3.5 Käyttöoikeuksiin liittyvät metatiedot .....	18
3.3.6 Tekninen validointi .....	18
3.3.7 Semanttinen validointi.....	19
3.4 Aineiston lohkominen eri luovutuspaketteihin.....	19
3.4.1 Tekniset rajoitteet pakettien koolle .....	20
3.4.2 Lohkominen sisällön ja luokittelun perusteella.....	20
<b>4 LUOVUTUSPAKETIN VALMISTELU</b> .....	<b>20</b>
4.1 Aineiston sähköinen allekirjoittaminen.....	20
4.1.1 Validointi ja laadunvarmistus .....	22
4.1.2 Poikkeustilanteet.....	22
4.1.3 Aineiston salaaminen ja pakkaaminen siirtoa varten.....	22
<b>5 LUOVUTUSPAKETIN SIIRTÄMINEN</b> .....	<b>23</b>
5.1 Online-siirto.....	23
5.2 Offline-siirto.....	23

# SÄHKE2

## Siirto-ohje

v. 1.0.2  
2.10.2009

3 (23)

## 1 JOHDANTO

Tämä sovellusohje on tarkoitettu täydentämään arkistolaitoksen antamaa SÄHKE2 - määräystä. Se on suunnattu sähköisesti arkistoitavaa aineistoa tuottaville organisaatioille, ja ohjeistaa teknisesti arkistoitavan aineiston lähetyspaketin (Submission Information Package, SIP) muodostamista, valmistelua ja lähetystä arkistolaitoksen VAPA-järjestelmään. Pääpaino on SÄHKE2-määräykseen liittyvän XML-siirtotiedoston skeeman soveltamisessa arkistoivan aineiston metatietojen esittämiseen. Ohjeessa käsitellään myös asiakirjojen yksilöintiä OID/URN -tunnisteella sekä asiakirjojen konversiota natiivimuodosta pitkäaikaissäilytettävään muotoon. Lisäksi käsitellään tilanne, jossa lähetettävä järjestelmä seuraa SÄHKE1-määräystä, myös sellaisessa tapauksessa, että siirtoon sisältyy SÄHKE1-määräyksestä puuttuvia metatietoja.

### 1.1 Termistö

**Luovutuspaketilla** (SIP-paketti, Submission Information Package) tarkoitetaan arkistolaitoksen järjestelmään siirrettäväksi tarkoitettua tietokokonaisuutta, joka koostuu siirrettävästä aineistosta, niihin liittyvistä metatiedoista sekä niihin liittyvistä sähköisistä allekirjoituksista.

Luovutuspaketin **muodostamisella** tarkoitetaan pitkäaikaissäilytettävän aineiston kokoamista sarjasta tiedostoja koostuvaksi arkistoitavaksi kokonaisuudeksi ja niihin liittyvien kuvailutietojen esittämistä XML-muotoisena siirtotiedostona.

Luovutuspaketin **valmistelulla** tarkoitetaan kyseisen kokonaisuuden allekirjoittamista ja pakkaamista lähetystä varten.

Luovutuspaketin **siirtämisellä** tarkoitetaan luovutuspaketin siirtämistä arkistolaitoksen järjestelmiin jollakin SÄHKE2 -määräyksessä hyväksytyllä menetelmällä.

**Siirtotiedostolla** tarkoitetaan luovutuspakettiin sisältyvää XML-tiedostoa, joka seuraa SÄHKE1- tai SÄHKE2-schemaa, ja joka sisältää siirrettävän aineiston metatiedot.

**SÄHKE-skeemalla** tarkoitetaan joko SÄHKE1- tai SÄHKE2-määräyksen yhteydessä julkaistua siirrettävän aineiston metatietojen kuvaustavan määrittävää XML-skeemaa.

### 1.2 Liittymät muihin määräyksiin ja ohjeisiin

Sovellusohje liittyy arkistolaitoksen SÄHKE2 -määräykseen. Ennen siihen perehtymistä on syytä tutustua määräyksen [normitekstiin](#), liitteisiin "[Metatietojen tuottaminen](#)" ja "[Metatietomalli](#)" sekä [siirtotiedoston XML-skeemaan](#).

SÄHKE1-määräykseen liittyvä aineisto on saatavissa [arkistolaitoksen WWW-sivuilta](#).

Asiakirjan yksilöintiä OID-tunnuksilla ohjeistaa [JHS-159 -suositus](#) "ISO OID-yksilöintitunnuksen soveltaminen julkishallinnossa."

# SÄHKE2

## Siirto-ohje

v. 1.0.2  
2.10.2009

4 (23)

Siirrettävä aineisto tallennetaan arkistolaitoksen VAPA-järjestelmään, josta saa tietoja [arkistolaitoksen WWW-sivuilta](#).

## 2 YLEISTÄ SIIRTOPROSESSISTA

Arkistolaitoksen VAPA-järjestelmään siirrettävä aineisto muodostuu arkiston käyttäjäorganisaation tietojärjestelmissä. Nämä ovat usein **asiankäsittelyjärjestelmiä**. Aineiston muodostamiseen liittyviä vaatimuksia sekä asiakirjallisen tiedon käsittelyprosessien aikana kertyviä metatietoja käsitellään SÄHKE-määräyksen normitekstissä sekä liitteessä 1, "Asiakirjallisten tietojen metatietojen tuottamisen periaatteet."

Siirtoprosessi arkistolaitoksen VAPA-järjestelmään käynnistetään kun on muodostunut arkistoitava kokonaisuus, johon ei enää liity operatiivisia käsittelyvaatimuksia. Tällöin aineisto on valmis pitkäaikaissäilytettäväksi. Siirron jälkeen alkuperäisen asiakirjan katsotaan sijaitsevan arkistolaitoksen järjestelmissä; käyttäjäorganisaation järjestelmiin mahdollisesti jäävät kappaleet katsotaan kopioiksi.

Arkistolaitoksen järjestelmän käyttö edellyttää sopimusta käyttäjäorganisaation ja arkistolaitoksen välillä. Tässä sopimuksessa täytyy sopia mm:

- Aineiston muodostavista järjestelmistä
- Aineiston siirtotavoista
- Siirtoaikataulusta
- SÄHKE-malliin kuulumattoman metatiedon käytöstä
  
- Siirrettävän aineiston luonteesta (fysiset asiakirjat, skannatut paperiasiakirjat, sähköiset asiakirjat, tiedostomuodot, ääni- ja videokokonaisuudet jne.)

### 2.1 Muodostuva aineisto

Siirtoprosessissa syntyy yksi tai useampi määrämuotoinen kokonaisuus, joka koostuu siirrettävistä asiakirjoista, niitä kuvailevista XML-tiedostoon kootuista metatiedoista, sekä aineiston eheyden ja kiistämättömyyden varmistavista allekirjoituksista. Tätä kokonaisuutta kutsutaan luovutuspaketiksi tai SIP-paketiksi (Submission Information Package).

### 2.2 Prosessin vaiheet

Siirtoprosessi alkaa luovutuspaketin tai pakettien **muodostamisella**. Tässä vaiheessa siirrettävä aineisto **konvertoidaan** pitkäaikaissäilytettävään muotoon ja **järjestetään** tiedostojärjestelmään määrämuotoiseksi paketiksi. Ellei OID-tunnuksia jo ole, jokainen asiakirja **yksilöidään** OID/URN -tunnuksella. Aineistoon liittyvät metatiedot kootaan aineiston muodostaneesta järjestelmästä, siirtoon liittyvät metatiedot luodaan siirtoa varten, ja metatiedot koostetaan SÄHKE2 -skeeman mukaiseksi XML-siirtotiedostoksi.

Kun luovutuspaketti on muodostettu, se **valmistellaan** siirtoa varten. Tässä vaiheessa pakettiin sisällytetyt tiedot **allekirjoitetaan** ja syntyneet allekirjoitukset liitetään pakettiin. Sen jälkeen paketti **salataan** ja **pakataan** siirtoa varten, mikäli tähän on tarve.

## SÄHKE2

### Siirto-ohje

v. 1.0.2  
2.10.2009

5 (23)

Lopuksi valmis luovutuspaketti **siirretään** arkistolaitoksen VAPA-järjestelmään. Siirto tapahtuu joko **offline**-siirtona fyysisellä medialla (CD- tai DVD-ROM, USB-kiintolevy) tai **online**-siirtona VAPA-järjestelmän extranet-rajapintaan, arkistolaitoksen kanssa tehdyn sopimuksen mukaisesti.

### 3 LUOVUTUSPAKETIN MUODOSTAMINEN

Arkistolaitokselle lähetettävä luovutuspaketti muodostetaan kokoamalla siirrettävä aineisto yhdeksi tai useammaksi asiakirjallisen tiedon käsittelyprosessin perusteella ryhmitellyksi kokonaisuudeksi, valmistelemalla aineisto siirtoa varten, järjestämällä se tiedostojärjestelmään määrämuotoiseksi rakenteeksi, sekä kokoamalla siihen liittyvät metatiedot SÄHKE-scheman mukaiseksi siirtotiedostoksi.

#### 3.1 Aineiston valmistelu

Ennen kuin siirrettävä aineisto voidaan koota siirtopaketiksi, se täytyy valmistella pitkäaikaissäilytystä varten. Tällöin se konvertoidaan pitkäaikaissäilytettävään muotoon, sen laatu varmistetaan, ja jokainen kokonaisuuteen kuuluva asiakirja ja tiedosto yksilöidään OID/URN-tunnuksella.

##### 3.1.1 Konversio pitkäaikaissäilytettävään muotoon

Ennen kuin aineisto siirretään arkistolaitokselle, se on konvertoitava johonkin SÄHKE2-määräyksessä luetelluista hyväksytyistä pitkäaikaissäilytysformaateista. Tieto formaatista on sisällytettävä siirtotiedostoon, kohtaan Document.Format. Tuetut formaatit ja niiden siirtotiedostossa käytettävät nimet ovat:

Formaatti	Nimi siirtotiedostossa	Versioiden nimi siirtotiedostossa	Pakkaus	Kuvaus
Unicode UTF-8 -tekstitiedosto	UTF-8	N/A	N/A	Vapaa, formatoimaton teksti, merkistö kattaa kaikki maailman kielet
TIFF Rev 5, Rev 6	TIFF	5,6	CCIT Group 3 (CCIT3), CCIT Group 4 (CCIT4)	Bittikarttakuva
PDF/A, ISO 19005-1:2005, IDT tai uudempi	PDF/A	1a, 1b, tulevaisuudessa myös muita	N/A	Muotoiltu asiakirja, voi sisältää myös kuvia
PCM WAV, 16 bit, 44.1 kHz	WAV	16 bit 44.1 kHz	N/A	Ääni
MPEG-1, Audio Layer 3, 128 kbit/s	MPEG1	AL3 128 kbit/s	N/A	Video

## SÄHKE2

### Siirto-ohje

v. 1.0.2  
2.10.2009

6 (23)

MPEG-2, 50 Mbit/s, 4:2:2	MPEG2	50 Mbit/s 4:2:2	N/A	Korkealaatuinen video
--------------------------	-------	-----------------	-----	-----------------------

Konversio pitkäaikaissäilytettävään muotoon tulee tehdä siten, että konvertoidusta asiakirjasta ei katoa merkityksellistä tietoa. Tallennusformaatti ja sen ominaisuudet tulee kuitenkin valita siten, että tiedostoista ei tule tarpeettoman suuria. Konversiokäytännöistä tulee sopia arkistolaitoksen kanssa, joka myös antaa tarvittaessa siihen liittyvää lisäopastusta.

Asiakirjan luonne	Suosittelut formaatti	Muut mahdolliset formaattit
Teksti, johon ei sisälly kuvia tai muotoilua, esimerkiksi sähköpostit tai yksinkertaiset muistiot	UTF-8	PDF/A
Semanttiset XML-formaatit, esim. CDA	UTF-8	PDF/A
Muotoiltu teksti, johon voi sisältyä kuvia, esimerkiksi useimmat toimistosovelluksilla (Microsoft Office, OpenOffice yms) tuotetut asiakirjat	PDF/A	TIFF
Skannatut paperiasiakirjat	TIFF	
Kuvat	TIFF	
Ääni	WAV	
Video	MPEG1	MPEG2

Semanttisissa XML-formaateissa olevat asiakirjat, jotka ovat sellaisenaan luettavissa ilman lisäkonversioita tai katselusovelluksia, voidaan siirtää arkistolaitokseen ilman konversiota. Tästä on kuitenkin sovittava arkistolaitoksen kanssa formaatti- ja tapauskohtaisesti, samalla kun solmitaan sopimus siirtojen tekemisestä.

Konvertoidun asiakirjan yhteydessä voidaan arkistoida myös natiivimuodossa oleva asiakirja. Tätä voidaan tarvita siinä tapauksessa, että konversiossa tapahtuu laadunvarmistuksesta huolimatta virheitä, joiden seurauksena syntyy epäily merkityksellisen tiedon katoamisesta; tällöin on usein mahdollista selvittää kysymykset tutkimalla natiivimuotoista tiedostoa.

### 3.1.2 Immateriaalioikeuksiin liittyvät rajoitukset

Arkistolaitokselle lähetettävän aineiston immateriaalioikeuksien täytyy kuulua lähettäjälle. Lähettäjä vastaa siitä, että tekijänoikeuksista ja muista immateriaalioikeuksista johtuvat rajoitukset eivät estä aineiston pitkäaikaissäilytystä. Jos aineistoon sisältyy osia, joiden immateriaalioikeudet kuuluvat kolmansille osapuolille, täytyy oikeudet ja niistä seuraavat rajoitukset kuvata metatietorakenteissa, kohdassa Record.Rights. Tällaisia osia voivat olla esimerkiksi videoleikkeiden taustamusiikki, asiakirjoihin sisältyvät kuvat ja PDF-tiedostoihin sisällytetyt kaupallisella lisenssillä käytettävät kirjasimet.

## SÄHKE2

### Siirto-ohje

v. 1.0.2  
2.10.2009

7 (23)

#### 3.1.3 Validointi ja laadunvarmistus

Lähtäjän tulee **validoida konvertoidut tiedostot teknisesti** aina konversion yhteydessä. Validointiin käytetään tähän tarkoitukseen tehtyä sovellusta. Sen lisäksi konversion tuloksen laatu täytyy varmistaa menetelmillä, joilla saadaan riittävä varmuus siitä, että merkityksellistä tietoa ei katoa.

Konvertoitujen tiedostojen tekninen validointi tehdään tätä varten tarkoitetuilla työkaluilla, esimerkiksi seuraavilla:

- DROID on Britannian kansallisarkiston kehittämä avoimeen lähdekoodiin perustuva tiedostomuodon tunnistus- ja validointityökalu. Siitä löytyy lisätietoja osoitteessa <http://droid.sourceforge.net/wiki/index.php/Introduction>.
- JHOVE on avoimeen lähdekoodiin perustuva Java-ohjelmointikielellä toteutettu tiedostomuodon validointityökalu. Siitä löytyy lisätietoja osoitteesta <http://hul.harvard.edu/jhove/>.

Formaatti	Mahdolliset validointityökalut
UTF-8	JHOVE, DROID
TIFF	JHOVE, DROID
PDF/A	JHOVE, DROID
WAV	JHOVE, DROID
MPEG-1	DROID
MPEG-2	DROID

Laadunvarmistus tehdään pistokokein.

**Huom:** Luovutuspaketteja ei saa ryhtyä siirtämään arkistolaitoksen järjestelmiin niin kauan kuin pistokokein löytyy virheellisesti konvertoituja asiakirjoja.

Formaatti	Tyypillisiä virhetilanteita
UTF-8	<ul style="list-style-type: none"><li>• Virheellinen merkistömuunnos, joka tekee tiedostosta lukukelvottoman tai rikkoo osan merkeistä</li><li>• rivin- tai kappaleenvaihtojen katoaminen tai lisääntyminen</li></ul>
TIFF	<ul style="list-style-type: none"><li>• Puutteellinen resoluutio: teksti puuroutuu niin pahasti, että asiakirja ei ole luettavissa.</li><li>• Väreihin liittyvät virheet: värejä katoaa tai ne vääristyvät niin paljon, että luettavuus kärsii.</li><li>• Asemointiin liittyvät virheet: elementti peittää toisen; elementti ylittää sivun reunan ja osa jää kuvasta pois.</li></ul>

## SÄHKE2

### Siirto-ohje

v. 1.0.2  
2.10.2009

8 (23)

	<ul style="list-style-type: none"><li>• Tarpeettoman korkea resoluutio tai värisyvyys, mikä kasvattaa tiedostot tarpeettoman suuriksi.</li></ul>
PDF/A	<ul style="list-style-type: none"><li>• Asemointiin liittyvät virheet: elementti peittää toisen; elementti ylittää sivun reunan ja osa jää kuvasta pois.</li></ul>
WAV	<ul style="list-style-type: none"><li>• Tiedoston kokoon liittyvät virheet (kirjoitus katkeaa ennen aikojaan).</li><li>• Koodaukseen liittyvät virheet (kuva tai ääni sekoaa).</li></ul>
MPEG-1	<ul style="list-style-type: none"><li>• Tiedoston kokoon liittyvät virheet (kirjoitus katkeaa ennen aikojaan).</li><li>• Koodaukseen liittyvät virheet (kuva tai ääni sekoaa).</li></ul>
MPEG-2	<ul style="list-style-type: none"><li>• Tiedoston kokoon liittyvät virheet (kirjoitus katkeaa ennen aikojaan).</li><li>• Koodaukseen liittyvät virheet (kuva tai ääni sekoaa).</li></ul>

**Huom:** Lisätietoja paperiasiakirjojen digitointiin ja sen yhteydessä suositeltaviin TIFF-tiedoston ominaisuuksiin saa arkistolaitoksen ohjeesta ["Digitoinnin laatuksiteerit"](#).

#### 3.1.4 Asiakirjojen yksilöinti OID/URN -tunnuksella

Asiakirjat tulee yksilöidä siten, että ne on yksiselitteisesti ja globaalisti tunnistettavissa ja erotettavissa muista asiakirjoista. Yksilöintiin suositellaan käytettävän ISO OID/URN -menetelmää.

ISO OID (Object Identifier) on piste-erotelluista numerosarjoista koostuva yksilöintimenetelmä, joka mahdollistaa tieto-olioiden kansainvälisen yksilöinnin.

OID perustuu hierarkiaan, jossa jokaista pisteellä eroteltua tasoa (solmuluokkaa) hallinnoi oma tahonsa. Esimerkiksi Suomen OID-tunnus on 1.2.246, ja sitä hallinnoi SFS. Organisaatio voi muodostaa omia OID-tunnuksiaan Y-tunnuksen perusteella. Tällöin tunnus alkaa muodolla 1.2.246.10.yyyyyyy, missä yyyyyyy on Y-tunnus. Yhteisöt voivat edelleen yksilöidä tietojärjestelmänsä ja asiakirjansa, esimerkiksi muodostamalla OID-tunnus ylläkuvatusta alkuosasta, tietojärjestelmälle annetusta tunnuksesta ja tietojärjestelmän asiakirjoille jakelemasta (sisäisestä tai ulkoisesta) yksilöintitunnuksesta.

Yksityiskohtaisempia tietoja ISO OID -tunnuksien muodostamisesta ja käytöstä saa JUHTA-työryhmän ohjeesta [JHS 159 ISO OID-yksilöintitunnuksen soveltaminen julkishallinnossa](#).

Luovutuspaketteihin sisältyvät asiakirjat on yksilöitävä ja yksilöintitiedot on sisällytettävä metatietoihin. Tällöin asiakirjan ISO OID -tunnus on esitettävä URN-muodossa. Tämä tapahtuu yksinkertaisesti liittämällä OID-tunnuksen eteen merkkijono urn:oid:.

## SÄHKE2

### Siirto-ohje

v. 1.0.2  
2.10.2009

9 (23)

- Esimerkki asiakirjan ISO OID tunnuksesta:  
1.2.246.10.2048198.10.1.11.2002.228
- Esimerkki samasta tunnuksesta esitettyä URN-muodossa:  
urn:oid:1.2.246.10.2048198.10.1.11.2002.228

#### 3.1.5 Salatun aineiston käsittely

Salatun aineiston siirtämistä säilytettäväksi arkistolaitoksen järjestelmiin ei suositella. Jos näin halutaan kuitenkin tehdä, siitä tulee sopia erikseen arkistolaitoksen kanssa tehtävässä sopimuksessa, jotta aineiston käytettävyys säilyy sen koko säilytysajan. Siinä täytyy määritellä:

- Käytettävä salausmenetelmä (algoritmi, työkalut)
- Avaimet
- Avainten säilytys
- Menettelyt salauksen purkamiseksi

Tiedot salausmenetelmästä täytyy tällöin sisällyttää aineiston metatietoihin.

**Huom:** Koko siirrettävä luovutuspaketti voidaan tarvittaessa salata siirron ajaksi (ks. "Aineiston salaaminen ja pakkaaminen siirtoa varten" alla). Tällöin salauksesta ei tule mainita metatiedoissa.

#### 3.2 Luovutuspaketin tiedostorakenne

Koottu aineisto järjestetään luovutuspaketiksi siten, että jokaiselle siirrettävälle asiakirjallisen tiedon käsittelyprosessille luodaan oma hakemisto, ja siirrettävät **asiakirjatiedostot** kootaan asiahakemistoon siten, että jokaista tiedostomuotoa varten luodaan oma hakemisto. Prosessihakemistojen nimet voidaan luoda esimerkiksi asiakirjallisen tiedon käsittelyprosessin OID-tunnuksen loppuosan perusteella. Hakemiston nimen pituus saa olla enintään 8 merkkiä, ja siinä saa käyttää merkkejä A-Z, a-z, 0-9 ja \_.

XML-siirtotiedosto sijoitetaan luovutuspaketin juureen, ja sille annetaan nimeksi **sahke1.xml**. Jos paketti sisältää useampia siirtotiedostoja, ne numeroidaan juoksevasti.

Allekirjoitustiedosto sijoitetaan luovutuspaketin juureen, ja sille annetaan nimeksi varm1.sig. Jos allekirjoitustiedostoja on useampia, ne numeroidaan juoksevasti.

#### Esimerkki

Luovutuspaketeissa siirretään asiakirjallisen tiedon käsittelyprosessi, jonka tunnus on **urn:oid:1.2.246.10.2048198.12.117.2008.244**.

Tällöin asiakirjallisen tiedon käsittelyprosessin hakemistonimi muodostetaan yhdistämällä OID-tunnuksen kaksi viimeistä solmuluokkaa, jolloin hakemiston nimeksi tulee **2008244**.

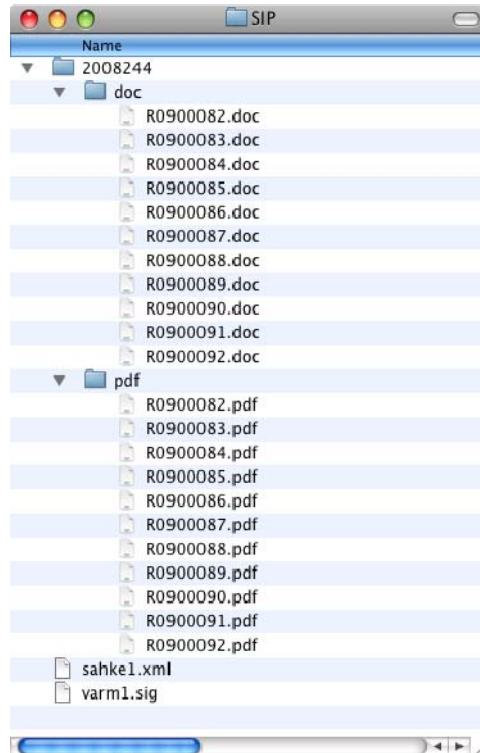
# SÄHKE2

## Siirto-ohje

v. 1.0.2  
2.10.2009

10 (23)

**Huom!** Tämä nimeämiskäytäntö ei ole sitova: mitä tahansa kunkin asiakirjallisen tiedon käsittelyprosessin yksilöivää nimeämiskäytäntöä saa käyttää.



### 3.3 Metatiedot ja niiden esittäminen XML-siirtotiedostona

Asiakirjojen metatiedot tulee muodostaa jo asiakirjan operatiivisen käytön yhteydessä. Tämä prosessi on kuvattu SÄHKE2-määräyksen liitteessä [Asiakirjallisten tietojen metatietojen tuottamisen periaatteet](#). Tässä ohjeessa oletetaan, että vähintään pakolliset metatiedot on muodostettu ja noudettavissa operatiivisista järjestelmistä.

Luettelo pakollisista metatiedoista sekä lyhyet kuvaukset metatietojen merkityksestä ja niille sallituista arvoista on SÄHKE2-määräyksen liitteessä [SÄHKE2 -metatietomalli](#).

Metatiedot esitetään XML-muotoisessa siirtotiedostossa, joka seuraa SÄHKE2-scheman mukaista rakennetta.

Siirrettävään aineistoon liittyvien metatietojen lisäksi myös itse siirrolla on metatietoja. Ne perustuvat palvelun käyttäjän ja arkistolaitoksen väliseen sopimukseen. Luovutuspaketin kokoavan integrointikomponentin tulee sisällyttää ne siirtotiedostoon.

Tämän ohjeen liitteinä on esimerkkejä erityyppisiin siirtoihin käytettävistä siirtotiedostoista.

## SÄHKE2

### Siirto-ohje

v. 1.0.2  
2.10.2009

11 (23)

#### 3.3.1 Siirtotiedoston rakenne

SÄHKE2-siirtotiedoston juurielementti on *Metadata*. Sen alle tulevat pakolliset *TransferInformation*, *ContactInformation*, sekä yksi (tai poikkeustapauksissa useampi) *CaseFile* -solmu. Ensinmainitut kaksi solmua sisältävät siirtoon liittyviä välttämättömiä metatietoja. *CaseFile*-solmut sisältävät siirrettävän aineiston metatiedot.

#### 3.3.2 SÄHKE1 -määräyksen mukainen siirtotiedosto

Tässä ohjeessa käytetyt esimerkit seuraavat SÄHKE2-määräyksen mukaista XML-siirtotiedostoa. Jos arkistolaitoksen kanssa tehty sopimus niin sallii, siirtoon voidaan käyttää myös SÄHKE1-määräyksen mukaisia siirtotiedostoja. Tämän ohjeen liitteenä on myös SÄHKE1-määräyksen mukaiset esimerkit siirtotiedostoista. Metatietojen merkityksen ja esitystavan osalta tätä ohjetta tulee soveltaa myös SÄHKE1-siirtotiedostoihin. Siinä tapauksessa, että SÄHKE1-mallin mukaisissa tiedostoissa siirretään metatietoja, jotka kuuluvat SÄHKE2-määräykseen, mutta joita ei ole määritelty SÄHKE1-määräyksessä, metatiedot kuvataan SÄHKE1-määräyksen vapaamuotoisia metatietoja sisältävissä solmuissa, käyttäen SÄHKE2-määräyksen rakenteita.

SÄHKE1-määräykseen sisältyy eräitä pakollisia metatietoja, joita ei sisälly SÄHKE2-määräykseen. Ne on sisällytetty tämän ohjeen liitteenä oleviin esimerkkeihin. Käytettäessä SÄHKE1-määräyksen mukaisia tiedostoja, tulee metatietojen arvoina käyttää SÄHKE2-määräyksen mukaisia ilmaisutapoja. SÄHKE2-määräykseen kuuluvat pakolliset metatiedot, joita ei ole SÄHKE1-määräyksessä, tulee kuvata esimerkin mukaisina *customMetadata*-solmuina. Rakenteet ja muodot ovat SÄHKE2-määräyksen mukaisia, mutta solmut pidetään SÄHKE1-nimiavaruudessa.

Tiedostorakenteeltaan luovutuspaketti on samanlainen kuin SÄHKE2-tapauksessa, paitsi, että *varm1.sig* -tiedostoa ei käytetä ja *sahke1.xml*-tiedoston lisäksi kokonaisuuteen sisältyy siirron metatiedot sisältävä tiedosto *siirto.xml*, joka seuraa SÄHKE1-määräystä. Järjestelmän kuvausta (*jarjkuv.xml*) ei tarvitse sisällyttää luovutuspakettiin.

- *FunctionalClassification* -solmu sisällytetään *group.customMetadata* -solmuun SÄHKE2-määräyksen mukaisella ilmaisutavalla.
- Tiedostomuoto ja -versio ilmaistaan kaksoispisteellä eroteltuna *document.preservation.format.softwareSupport* -solmussa, esim. `<softwareSupport>PDF/A:1b</softwareSupport>`
- Koska SÄHKE1-määräys ei edellytä luovutuspaketin kryptografista varmentamista, siihen liittyvät *HashAlgorithm* ja *HashValue* -solmut voidaan jättää pois. Jos ne halutaan sisällyttää siirtotiedostoon, ne voidaan ilmaista *document.customMetadata* -solmun alla SÄHKE2-määräyksen mukaisella tavalla.

**HUOM:** Tämän ohjeen yhteydessä on julkaistu uusi versio SÄHKE1-metatietoschemasta. Siihen on korjattu eräitä vanhassa versiossa olleita virheitä ja sitä on tietyiltä osin väljennetty sallimaan SÄHKE2-määräyksen mukaisia ilmaisutapoja. Alkuperäisen SÄHKE1-scheman mukaiset siirtotiedostot läpäisevät myös uuden scheman.

## SÄHKE2

### Siirto-ohje

v. 1.0.2  
2.10.2009

12 (23)

Esimerkkejä SÄHKE1-määräyksen mukaisista siirtotiedoista on tämän ohjeen liitteissä.

#### 3.3.3 Siirron metatiedot

Arkistolaitokseen siirrettävä aineisto on aina jonkin sopimuksen piirissä. Tähän sopimukseen sekä siirron teknisiin järjestelyihin liittyvät metatiedot sisällytetään siirtotiedoston *TransferInformation*-solmuun.

```
<s2:TransferInformation>
  <s2:NativeId>urn:oid:1.2.246.10.2048198.10.1.2008.3311.2</s2:NativeId>
  <s2:Title>Hallinnollisten asiain viraston asiankäsittelyjärjestelmä
HORKKA</s2:Title>

<s2:TransferContractId>urn:oid:1.2.246.10.2048198.10.1.2008.3311</s2:TransferContr
actId>

<s2:MetadataSchema>http://www.narc.fi/Arkistolaitos/sahkoinen/Sahke2_2009_06.xsd</
s2:MetadataSchema>
</s2:TransferInformation>
```

- Siirrot kattava sopimus saa sopimusta solmittaessa ISO-OID -tunnuksen. Tämä tunnus syötetään URN/OID -muodossa *TransferInformation.TransferContractId* -solmuun.
- Jokainen siirto saa oman yksilöivän ISO-OID -tunnuksen. Se muodostetaan liittämällä *TransferInformation.TransferContractId*:n perään omaan solmuluokkaansa yksilöivä luku, esimerkiksi juokseva numero. Tämä tunnus tallennetaan solmuun *TransferInformation.NativeId*.
- *TransferInformation.MetadataSchema* -solmu saa arvoksi käytetyn siirtotiedoston XML-scheman URL-osoitteen arkistolaitoksen WWW-sivuilta.

*ContactInformation*-solmuun sisällytetään siirrosta vastaavan teknisen yhteyshenkilön tiedot. Arkistolaitos ottaa yhteyttä tähän henkilöön jos siirrossa tapahtuu ongelmia.

```
<s2:ContactInformation>
  <s2:Organisation>
    <s2:Name>Hallinnollisten asiain virasto</s2:Name>
  </s2:Organisation>
  <s2:ContactPerson>
    <s2:Name>Hilkka Hakkeroinen</s2:Name>
    <s2:Address>Jokikatu 3</s2:Address>
    <s2:PhoneNumber>044-123123</s2:PhoneNumber>
    <s2:Email>Hilkka.Hakkeroinen@haveri.fi</s2:Email>
  </s2:ContactPerson>
</s2:ContactInformation>
```

## SÄHKE2

### Siirto-ohje

v. 1.0.2  
2.10.2009

13 (23)

#### 3.3.4 Siirrettävän kokonaisuuden metatiedot

Siirrettävän kokonaisuuden metatiedot koostuvat kunkin siirrettävän **asiakirjallisen tiedon käsittelyprosessin** metatiedoista, kuhunkin käsittelyprosessiin liittyvän **toimenpiteen** metatiedoista, jokaiseen toimenpiteeseen liittyvän **asiakirjan** metatiedoista, sekä jokaiseen asiakirjaan liittyvän **tiedoston** metatiedoista. Nämä rakenteet on kuvattu SÄHKE2 -määräyksen liitteessä 2 sekä tämän ohjeen liitteinä olevissa esimerkeissä.

##### 3.3.4.1 Asiakirjallisen tiedon käsittelyprosessin metatiedot

Jokaisen luovutuspakettiin liittyvän asiakirjallisen tiedon käsittelyprosessin metatiedot kuvataan solmussa *CaseFile*.

```
<s2:CaseFile>
  <s2:Created>2008-08-14</s2:Created>
  <s2:NativeId>urn:oid:1.2.246.10.2048198.12.117.244</s2:NativeId>
  <s2:Language>fi</s2:Language>
  <s2:Description>Parkkipaikkojen ja puistojen ylitarkastajan virantäytön valmistelu</s2:Description>
  <s2:Restriction>
    <s2:PublicityClass>Julkinen</s2:PublicityClass>
    <s2:PersonalData>Ei sisällä henkilötietoja</s2:PersonalData>
  </s2:Restriction>
  <s2>Title>Virantäytön valmistelu</s2>Title>
  <s2:RetentionPeriod>20</s2:RetentionPeriod>
  <s2:RetentionReason>Kalastuslaki, §5 3 mom.</s2:RetentionReason>
  <s2>Status>valmis</s2>Status>
  <s2:ClassificationScheme>
    <s2>MainFunction>
      <s2>Title>HENKILÖSTÖASIAT</s2>Title>
      <s2:FunctionCode>01</s2:FunctionCode>
      <s2:FunctionClassification>
        <s2>Title>Palvelussuhdeasiat</s2>Title>
        <s2:FunctionCode>01.01</s2:FunctionCode>
        <s2:SubFunction>
          <s2>Title>Rekrytointi ja palvelussuhteeseen valitseminen</s2>Title>
          <s2:FunctionCode>01.01.01</s2:FunctionCode>
        </s2:SubFunction>
      </s2:FunctionClassification>
    </s2>MainFunction>
  </s2:ClassificationScheme>
  <s2:Function>01.01.01</s2:Function>
  <s2:Agent>
    <s2:Role>Henkilöstöpäällikkö</s2:Role>
    <s2:Name>Murju, Rauno</s2:Name>
  </s2:Agent>
  <s2:Finished>2009-01-01</s2:Finished>
</s2:CaseFile>
```

## SÄHKE2

### Siirto-ohje

v. 1.0.2  
2.10.2009

14 (23)

- Jokainen asiakirjallisen tiedon käsittelyprosessi yksilöidään ISO-OID - tunnuksella, joka muodostetaan esimerkiksi liittämällä käytetyn tiedostojärjestelmän OID-tunnuksen perään solmuluokka, jossa on esimerkiksi juokseva numero. Tämä tunnus sisällytetään kunkin asiakirjallisen tiedon käsittelyprosessin metatietoihin, solmuun *CaseFile.Nativeld*.
- Tehtäväluokituksesta siirretään vain se osa, johon siirrettävät asiakirjat kuuluvat.

#### 3.3.4.2 Toimenpiteen metatiedot

Toimenpiteen metatiedot kuvataan solmussa *CaseFile.Action*.

```
<s2:Action>
  <s2:Title>Virantäytön valmistelu</s2:Title>
  <s2:Agent>
    <s2:Role>Henkilöstöpäällikkö</s2:Role>
    <s2:Name>Murju, Rauno</s2:Name>
  </s2:Agent>
  <s2:Type>valmistelu</s2:Type>
</s2:Action>
```

*Action.Type* -solmuun tulee toimenpiteen tyyppi. Käytetyt tyypit tulee määritellä työkulkuja kuvattaessa. Yleisimmät toimenpidetyypit sekä niihin liittyvät asiakirjatyypit (ks. alla) ovat:

vireilletulo	valmistelu	päätöksenteko	toimeenpano	muutoksenhaku	seuranta
aloite	esitys	lausunto	ilmoitus	kantelu	raportti
ehdotus	EU-tosite	liite	julkaisu	oikaisuvaatimus	
esitys	kartta	muistio	kirje	päätös	
hakemus	kertomus	päätös	kuulutus	valitus	
ilmoitus	kutsu	pöytäkirja	lasku		
kantelu	kuva	sopimus	määräys		
kirje	lista	suositus	ohje		
kutsu	lausunto	tilinpäätös	ote		
kuulutus	lausuntopyyntö	vaalilippu	päätös		
lasku	liite		sopimus		
lausuntopyyntö	luettelo				
liite	muistio				
lähete	ohje				
muistio	ohjelma				
määräys	ote				
oikaisuvaatimus	piirustus				
valitus	päätös				
	raportti				
	selvitys				
	sopimus				
	suunnitelma				
	sääntö				
	tarjous				
	tosite				

## SÄHKE2 Siirto-ohje

v. 1.0.2  
2.10.2009

15 (23)

**Huom:** Lista ei ole tyhjentävä, ja organisaatiot voivat täydentää sitä tarpeittensa mukaan. Siitä ei kuitenkaan ole syytä poiketa ilman syytä.

### 3.3.4.3 Asiakirjan metatiedot

Asiakirjan metatiedot kuvataan *CaseFile.Action.Record* -solmussa.

```
<s2:Record>
  <s2:Created>2009-01-05</s2:Created>
  <s2:NativeId>urn:oid:1.2.246.10.2048198.10.1.11.2009.84</s2:NativeId>
  <s2:Language>fi</s2:Language>
  <s2:Description>Hakemus parkkipaikkojen ja puistojen ylitarkastajan
virkaan</s2:Description>
  <s2:Restriction>
    <s2:PublicityClass>Julkinen</s2:PublicityClass>
    <s2:PersonalData>Sisältää henkilötietoja</s2:PersonalData>
  </s2:Restriction>
  <s2>Title>Hillevi Hiirulaisen hakemus</s2>Title>
  <s2:RecordRelation>
    <s2:References>urn:oid:1.2.246.10.2048198.10.1.11.2009.83</s2:References>
  </s2:RecordRelation>
  <s2:RecordRelation>
    <s2:IsReferencedBy>urn:oid:1.2.246.10.2048198.10.1.11.2009.87</s2:IsReferencedBy>
  </s2:RecordRelation>
  <s2:RecordRelation>
    <s2:IsReferencedBy>urn:oid:1.2.246.10.2048198.10.1.11.2009.90</s2:IsReferencedBy>
  </s2:RecordRelation>
  <s2:RetentionPeriod>20</s2:RetentionPeriod>
  <s2:RetentionReason>Kalastuslaki, §5 3 mom.</s2:RetentionReason>
  <s2>Status>valmis</s2>Status>
  <s2:Function>01 01 01</s2:Function>
  <s2>Type>hakemus</s2>Type>
  <s2:Accepted>2009-01-01</s2:Accepted>
  <s2:Authenticity>
    <s2:Checker>Murju, Rauno</s2:Checker>
    <s2>Date>2009-01-05</s2>Date>
    <s2>Description>Tarkistettu, allekirjoitettu käsin.</s2>Description>
  </s2:Authenticity>
</s2:Record>
```

- Jokainen asiakirja yksilöidään URN/OID -tunnuksella, JHS159 -suosituksen mukaisesti.
- *Record.Status* -solmu saa yleensä arvon "valmis", koska arkistolaitoksen järjestelmiin siirretään asiakirjoja, joiden operatiivinen käyttö on päättynyt.
- *Record.Function* -solmu viittaa *CaseFile.ClassificationScheme* -kokonaisuuden alla esiintyvään *FunctionCode*-solmuun.

## SÄHKE2

### Siirto-ohje

v. 1.0.2  
2.10.2009

16 (23)

- *Record.Type* -solmu sisältää asiakirjatyypin. Lista suositelluista asiakirjatyypeistä toimenpiteittäin yllä kohdassa "Toimenpiteen metatiedot."

#### 3.3.4.4 Viittaukset asiakirjojen välillä

Viittaukset asiakirjojen välillä kirjataan *Record.RecordRelation* -solmuihin. Viittaus tehdään URN/OID -tunnuksella. Kaksisuuntaisten viittausten semanttinen eheys tulee varmistaa. Toisin sanoen, jos asiakirjassa A mainitaan, että asiakirja B viittaa siihen, niin asiakirjasta A täytyy löytyä merkintä tästä viittauksesta:

##### Asiakirja A:

```
<s2:Record>
...
<s2:NativeId>urn:oid:1.2.246.10.2048198.10.1.11.2009.83</s2:NativeId>
...
<s2:RecordRelation>
<s2:IsReferencedBy>urn:oid:1.2.246.10.2048198.10.1.11.2009.84</s2:IsReferencedBy>
</s2:RecordRelation>
...
</s2:Record>
```

##### Asiakirja B:

```
<s2:Record>
...
<s2:NativeId>urn:oid:1.2.246.10.2048198.10.1.11.2009.84</s2:NativeId>
...
<s2:RecordRelation>
  <s2:References>urn:oid:1.2.246.10.2048198.10.1.11.2009.83</s2:References>
</s2:RecordRelation>
...
</s2:Record>
```

Viittauksia saa esiintyä myös luovutuspaketin ulkopuolisiin asiakirjoihin. Tällöin kaksisuuntaisten viittausten eheys tulee varmistaa pakettia muodostettaessa, koska sitä ei voi siirtotiedostoa tarkastelemalla tehdä.

#### 3.3.4.5 Immateriaalioikeuksiin liittyvät rajoitukset

Jos asiakirja sisältää tekijänoikeuksin tai muiden immateriaalioikeuksien perusteella käyttörajoitettuja osia, esimerkiksi taustamusiikkia tai kaupallisella lisenssillä käytettyjä kirjaimia, täytyy näistä rajoituksista mainita *Record.Rights* -solmussa.

```
<s2:Record>
...
```

## SÄHKE2

### Siirto-ohje

v. 1.0.2  
2.10.2009

17 (23)

```
<s2:Rights>Taustamusiikki (c) Love Records, 1977. Käytetty luvanvaraisesti,
edelleenlevitys kielletty.</s2:Rights>
...
</s2:Record>
```

#### 3.3.4.6 Tiedostojen metatiedot

Jos siirrettävä asiakirja on sähköinen, siihen liittyy yksi tai useampia tiedostoja. Tiedostojen metatiedot tulee pitää mahdollisimman yksinkertaisina: kaikki mahdollinen on parempi esittää asiakirjatasolla. Tiedostoja koskevat metatiedot liittyvät lähinnä tiedostomuotoihin ja tiedostojen sijaintiin luovutuspaketissa.

```
<s2:Document>
  <s2:NativeId>urn:oid:1.2.246.10.2048198.10.1.11.2009.89.1</s2:NativeId>
  <s2:UseType>Arkisto</s2:UseType>
  <s2:File>
    <s2:Name>R0900089.pdf</s2:Name>
    <s2:Path>2008245/pdf/R0900089.pdf</s2:Path>
  </s2:File>
  <s2:Format>
    <s2:Name>PDF/A</s2:Name>
    <s2:Version>1b</s2:Version>
  </s2:Format>
  <s2:HashAlgorithm>sha1</s2:HashAlgorithm>
  <s2:HashValue>pC4LUmP5L6DRL4gUYTx2y82JanG=</s2:HashValue>
  <s2:Encryption>Ei salattu</s2:Encryption>
</s2:Document>
```

- Tiedoston yksilöivä OID-tunnus (*Document.NativeId*) voidaan muodostaa liittämällä asiakirjan OID-tunnuksen perään pisteellä eroteltu juokseva numero.
- Tiedoston nimen tulee olla muotoa 8.3 - kahdeksanmerkkinen nimi ja kolmimerkkinen formaattitunnus. Niissä tulee käyttää merkkejä A-Z, a-z, 0-9 ja \_, ja loppuosan tulee vastata käytettyä tiedostoformaattia. Muita rajoituksia nimeämiskäytännölle ei ole.
- Tiedostopolku luovutuspaketissa ilmaistaan suhteessa XML-siirtotiedoston sijaintiin. Siirtotiedosto sijaitsee luovutuspaketin juuritasolla, jolloin polut viittavat tiedostorakenteessa syvemmällä oleviin tiedostoihin. Luovutuspaketin tiedostorakenne on kuvattu yllä, kohdassa "Luovutuspaketin tiedostorakenne."
- Tiedoston formaatti ja formaatin versio ilmoitetaan kohdassa *Document.Format.Name* ja *Document.Format.Version*. Luettelo sallituista arkistointimuodoista ja niiden esitystavoista siirtotiedostossa on yllä, kohdassa "Konversio pitkäaikaissäilytettävään muotoon." Natiivimuodon osalta tiedostoformaatin ja version ilmaisevat tiedot ovat vapaamuotoisia ja luonteeltaan kuvailevia.
- Tiedoston hajautussumma ja siihen käytetty algoritmi ilmaistaan kohdissa *Document.HashAlgorithm* ja *Document.HashValue*. Sallitut hajautusalgoritmit ovat:

## SÄHKE2

### Siirto-ohje

v. 1.0.2  
2.10.2009

18 (23)

Algoritmi	Ilmaisutapa tiedostossa
MD5	md5
SHA-1	sha1
SHA-256	sha256

- Salattujen tiedostojen siirtämistä arkistolaitokseen ei suositella, ja siitä täytyy erikseen sopia. Jos sopimus sallii salattujen tiedostojen siirron, käytetty salausten menetelmä ilmaistaan solmussa *Document.Encryption*. Muussa tapauksessa solmu saa arvon "Ei salattu".

#### 3.3.5 Käyttöoikeuksiin liittyvät metatiedot

Jos asiakirjan julkisuusluokaksi kohdassa *Record.Restriction.PublicityClass* on merkitty jokin muu kuin Julkinen, täytyy käyttöoikeudet kuvailla saman solmun lisämääreissä. Nämä tiedot ovat luonteeltaan kuvailevia, ja niitä käytetään auditointitarkoituksiin. Niillä ei siis ohjata teknisesti arkistoitujen asiakirjojen käyttöoikeuksia.

```
<s2:Restriction>
  <s2:PublicityClass>Salassa pidettävä</s2:PublicityClass>
  <s2:SecurityPeriod>20</s2:SecurityPeriod>
  <s2:SecurityPeriodEnd>2029-02-05</s2:SecurityPeriodEnd>
  <s2:SecurityReason>Kalastuslaki, 5§, 33 mom.</s2:SecurityReason>
  <s2:SecurityClass>Ei turvallisuusluokiteltu</s2:SecurityClass>
  <s2:PersonalData>Sisältää henkilötietoja</s2:PersonalData>
  <s2:Owner>Rauno Murju</s2:Owner>
  <s2:AccessRight>
    <s2:Name>Rauno Murju</s2:Name>
    <s2:Role>henkilöstöpäällikkö</s2:Role>
    <s2:AccessRightDescription>2</s2:AccessRightDescription>
  </s2:AccessRight>
</s2:Restriction>
```

On huomioitava, että järjestelmissä, joissa käyttöoikeuksien hallinta on politiikkaperusteista, SÄHKE2 -määräyksen mukaisesti yksilöityjen käyttöoikeuksien määrä saattaa nousta hyvin suureksi, mikä saattaa kasvattaa *Restriction.AccessRight* -solmujen lukumäärää hallitsemattomasti. Tällöin käyttöoikeuksien kuvailussa on syytä käyttää harkintaa, esimerkiksi kuvailemalla ainoastaan asiakirjan omistajan oikeudet.

#### 3.3.6 Tekninen validointi

Siirtotiedosto täytyy validoida teknisesti XML-schemaa vastaan tähän tarkoitukseen sopivalla työkalulla (esimerkiksi JHOVE). Teknistä validointia läpäisemätöntä siirtotiedostoa ei saa lähettää. Siirtotiedoston merkistökäsi suositellaan UTF-8:aa, ja merkintä merkistöstä täytyy sisällyttää XML-deklaraatioon:

## SÄHKE2

### Siirto-ohje

v. 1.0.2  
2.10.2009

19 (23)

```
<?xml version="1.0" encoding="UTF-8"?>
<s2:Metadata xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.narc.fi/Sahke2/2009-06 ../xsd/Sahke2_2009_06.xsd"
  xmlns:s2="http://www.narc.fi/Sahke2/2009-06">
  ...
</s2:Metadata>
```

#### 3.3.7 Semanttinen validointi

Teknisen validoinnin lisäksi siirtotiedosto täytyy validoida semanttisesti: tietojen teknisen oikeellisuuden lisäksi täytyy tarkastaa niiden merkityksellinen oikeellisuus. Mikäli mahdollista, validointi on syytä automatisoida ja tehdä aina luovutuspakettia kootessa. Jos tämä ei ole teknisesti toteutettavissa, semanttinen validointi täytyy tehdä ihmisvoimin ja pistokokein; aineistoa ei saa ryhtyä lähettämään arkistolaitokselle niin kauan kuin virheitä löytyy. Semanttisessa validoinnissa on tarkastettava:

- Siirron metatietojen oikeellisuus (*TransferInformation-* ja *ContactInformation-*solmut ja niiden sisältö)
- Tyhjät solmut: niitä ei tule sisällyttää siirtotiedostoon
- Kaksisuuntaisten viittausten eheys. Jos viittaukset tehdään ulkopuolisiin asiakirjoihin, ne tulee tarkastaa sikäli kuin tämä on teknisesti mahdollista.
- Rakenteiden oikeellisuus: tulee tarkastaa, että asiakirjojen metatiedot ovat niille kuuluvien asioiden metatietojen sisällä
- Tehtäväluokituksen oikeellisuus: ainoastaan se osa tehtäväluokituksesta, johon siirrettävä aineisto kuuluu, tulee sisällyttää siirtotiedostoon. Toisaalta täytyy tarkistaa, että viittaukset tehtäväluokkiin kohdistuvat luokkiin, jotka on kuvattu tehtäväluokitusolmussa.

#### 3.4 Aineiston lohkominen eri luovutuspaketteihin

Jos siirrettävää aineistoa on paljon, se on syytä lohkoa useampiin luovutuspaketteihin. Vaikka määräys sallii useampien siirtotiedostojen sisällyttämisen yhteen pakettiin, on yleensä selkeämpää lohkoa aineisto siten, että jokaisessa paketissa on tasan yksi siirtotiedosto sekä kaikki siihen liittyvä aineisto. Aineistoa lohkottaessa täytyy ottaa huomioon luovutuspakettien koolle syntyvät tekniset rajoitteet sekä siirrettävän aineiston semanttinen luonne ja rakenne. Näin vältetään pakettia muodostettaessa, siirrettäessä ja luettaessa syntyviä ongelmia, ja tehdään mahdollisten ongelmatilanteiden ratkaisu helpommaksi.

Pienempien luovutuspakettien allekirjoittaminen sekä tekninen ja semanttinen validointi, pistokokein tai ohjelmallisesti, on myös merkittävästi helpompaa kuin suurien. On siis suositeltavaa lohkoa aineisto mahdollisimman pieniin kokonaisuuksiin, kuitenkin siten, että kerrallaan siirretään asiallisesti mahdollisimman täydellisiä kokonaisuuksia. Useimmiten yhdessä luovutuspaketissa siirretään yhteen asiakirjallisen tiedon käsittelyprosessiin sisältyvä aineisto. Tällöin XML-siirtotiedostossa esiintyy täsmälleen yksi sitä vastaava *CaseFile*-solmu.

## SÄHKE2

### Siirto-ohje

v. 1.0.2  
2.10.2009

20 (23)

#### 3.4.1 Tekniset rajoitteet pakettien koolle

Luovutuspaketin kokoa rajoittaa ennen kaikkea siirtotiedoston koko. Koska se on XML-tiedosto, jota käsitellään ohjelmallisesti sekä sitä muodostettaessa että sitä luettaessa, on todennäköistä, että tiedosto tai siitä muodostettu objektimalli joudutaan lukemaan tietokoneen muistiin. XML-tiedostojen työstäminen on sitä tehokkaampaa, mitä pienempiä tiedostot ovat. Siksi aineisto on syytä lohkoa useampiin paketteihin, viimeistään jos XML-tiedoston koko ylittää yhden megatavun.

Toinen luovutuspaketin kokoa rajoittava tekninen tekijä on aineiston koko. Tätä rajoittaa ennen kaikkea valittu siirtomenetelmä: useiden gigatavujen kokonaisuuksia on helppo siirtää offline-siirtona esimerkiksi USB-kiintolevyllä. Online-siirroissa hyvin suurten kokonaisuuksien siirtäminen ei suositella, koska verkkoyhteyksien kapasiteetti saattaa muodostua pullonkaulaksi, ja virhetilanteista toipuminen on hankalaa jos koko paketti joudutaan siirtämään uudestaan. Online-siirrettävien luovutuspakettien koko ei tulisi ylittää 10 MT, ellei arkistolaitoksen kanssa toisin erityisesti sovita.

Alarajaa luovutuspaketin koolle ei ole. Yleisesti ottaen siirto on helpompaa, kun siirretään monta pientä luovutuspakettia kuin yksi suuri, kuitenkin niin, että tehtäväkokonaisuudet sekä asiakirjojen väliset viittaukset pidetään mahdollisuuksien mukaan yhden paketin sisällä.

#### 3.4.2 Lohkominen sisällön ja luokittelun perusteella

Suosittelava pienin yksikkö siirrettävälle kokonaisuudelle on *yksi asiakirjallisen tiedon käsittelyprosessi*. Siirtokokonaisuus voidaan siis muodostaa esimerkiksi yhteen rekrytointiprosessiin tai yhteen hankintaprosessiin liittyvästä tehtäväkokonaisuudesta. Koska asiakirjat tyypillisesti viittaavat toisiinsa nimenomaan asiakirjallisen tiedon käsittelyprosessin sisällä, näin toimittaessa asiakirjojen väliset viittaukset ovat pääsääntöisesti luovutuspaketin sisäisiä, ja siten melko helposti validoitavissa ohjelmallisesti. Myös prosessin auditoitavuus säilyy, koska koko aineisto siirretään yhtenä yksikkönä.

## 4 LUOVUTUSPAKETIN VALMISTELU

Ennen luovutuspaketin lähettämistä arkistolaitokselle, se täytyy vielä valmistella lähetystä varten. Tämä pitää sisällään aineiston sähköisen allekirjoittamisen, teknisen validoinnin sekä laadunvarmistuksen, ja mahdollisesti aineiston pakkaamisen ja salaamisen siirtoa varten.

#### 4.1 Aineiston sähköinen allekirjoittaminen

Siirrettävän aineiston eheys ja kiistämättömyys varmistetaan allekirjoittamalla se PKCS#7-allekirjoituksella. Allekirjoitus muodostetaan seuraavasti:

1. Aineiston muodostavista tiedostoista (asiakirjat sekä metatietotiedosto tai -tiedostot) lasketaan kustakin hajautussumma jollakin SÄHKE2-määräyksessä sallituista algoritmeista.

## SÄHKE2

### Siirto-ohje

v. 1.0.2  
2.10.2009

21 (23)

2. Hajautussummat kootaan tekstitiedostoon siten, että jokaiselle riville kirjataan hajautetun tiedoston hakemistopolku suhteessa luovutuspaketin juureen, käytetty hajautusalgoritmi, ja hajautussumma.
3. Tiedosto allekirjoitetaan PKCS#7-allekirjoituksella, käyttäen varmennetta, josta on sovittu arkistolaitoksen kanssa sopimusta tehtäessä.
4. Tiedosto tallennetaan luovutuspaketin juureen nimellä **varm1.sig**. Merkistönä käytetään UTF-8:aa.

Jos paketissa käytetään useampia allekirjoitustiedostoja, ne numeroidaan juoksevasti 1:stä alkaen.

Arkistolaitos ohjeistaa tarkemmin allekirjoitusprosessista sekä siihen käytettävistä työkaluista siirron teknisistä järjestelyistä sovittaessa.

#### Yksinkertaistettu esimerkki varm1.sig -tiedostosta:

```
MIME-Version: 1.0
Content-Type: multipart/signed; protocol="application/x-pkcs7-signature";
micalg=sha1; boundary="-----ms010103040403030808010908"

This is a cryptographically signed message in MIME format.

-----ms010103040403030808010908
Content-Type: text/plain; charset=UTF-8; format=flowed
Content-Transfer-Encoding: 8bit

file://sahkel.xml:sha1:DTA4MDUxNTEyMTkwNloXDTEzMDU=
file://244/pdf/R0900082.pdf:sha1:mP5L62Ja4LnGpCDRL4gUYTx2y8U=
file://244/pdf/R0900083.pdf:sha1:Ja4LnGpCDRL4mP5YTx2y8UL62gU=
file://245/pdf/R0900084.pdf:sha1:x2y8UmP5L62Ja4LnGpCDRL4gUYT=
file://245/pdf/R0900085.pdf:sha1:L4gUYTx2y8UmP5L62Ja4LnGpCDR=
file://246/pdf/R0900091.pdf:sha1:YTRL4gJanx2y82pUC4LUmP5L6DG=
file://246/pdf/R0900092.pdf:sha1:L62Ja4LxUYmL4gP5nGpCRTD2y8U=
-----ms010103040403030808010908
Content-Type: application/x-pkcs7-signature; name="smime.p7s"
Content-Transfer-Encoding: base64
Content-Disposition: attachment; filename="smime.p7s"
Content-Description: S/MIME Cryptographic Signature

MIAGCSqGS Ib3DQEHAqCAMIACAQExCzAJBgUrDgMCGgUAMIAGCSqGS Ib3DQEHAQAooIIMHDCC
...
1eV1bBgsCggCqMRV2U23/NSOXk+IB3BHCU0osWlw6h3eCiDTnndSJQ1YDxEAAAAAAAA=
-----ms010103040403030808010908--
```

## SÄHKE2

### Siirto-ohje

v. 1.0.2  
2.10.2009

22 (23)

#### 4.1.1 Validointi ja laadunvarmistus

Jokainen lähetettävä luovutuspaketti tulee validoida teknisesti ennen lähetystä. Tämä sisältää vähintään seuraavat toimenpiteet:

- XML-metatietotiedoston koko, joka ei saa ylittää arkistolaitoksen kanssa sovittua kattoa.
- XML-metatietotiedoston validoinnin SÄHKE-schemaa vastaan.
- Allekirjoitustiedostoon koottujen hajautussummien vertaaminen XML-metatietotiedostossa oleviin.
- Allekirjoituksen validointi.

Mikäli teknisesti mahdollista, on myös hyvä validoida seuraavat asiat:

- XML-metatietotiedostossa olevat asiakirjojen väliset viittaukset luovutuspaketin sisällä. Niiden tulee olla kaksisuuntaisia.
- Tehtäväluokkiin tehdyt viittaukset. Niiden tulee kohdistua luokkiin, jotka on ilmoitettu metatietotiedoston FunctionalClassification-osiossa.
- Se, että XML-metatietotiedostossa ei ole tyhjiä solmuja tai tehtäväluokkia, joihin ei viitata.

#### 4.1.2 Poikkeustilanteet

Online-siirtojen teknisistä järjestelyistä sopimisen yhteydessä arkistolaitos ilmoittaa siirrossa käytetyistä palautusarvoista. Samassa yhteydessä sovitaan yhteyshenkilöistä ja yhteydenottotavoista. Mikäli online-siirrossa tapahtuu virhe, aineiston lähettäjän tulee ottaa yhteyttä arkistolaitoksen yhteyshenkilöön.

Jos kyseessä on offline-siirto tai virhe havaitaan vasta arkistolaitoksen järjestelmissä, ottaa arkistolaitoksen edustaja yhteyttä aineiston lähettäjään, ensi sijassa XML-siirtotiedostossa ilmoitettuun henkilöön, mutta mikäli siirtotiedosto ei ole luettavissa, sopimuksessa määritellyllä tavalla.

Epäonnistunutta siirtoa ei saa yrittää uudestaan samalla aineistolla ennen kuin virheen aiheuttaja on selvitetty ja korjattu.

#### 4.1.3 Aineiston salaaminen ja pakkaaminen siirtoa varten

Ellei asiasta toisin sovita, aineisto on syytä pakata siirtoa varten. Pakkausmenetelmäksi suositellaan ZIP-pakkausta, versio 5.2 tai uudempi.

Jos aineisto siirretään arkistolaitoksen järjestelmiin online-siirtona, jossa yhteys on suojattu SSL-kerroksella (esimerkiksi SFTP tai HTTPS -protokollilla), luovutuspakettia ei tarvitse erikseen salata. Jos käytetään salaamatonta siirtoprotokollaa tai offline-siirtoa erillisellä medialla, joka toimitetaan arkistolaitokselle, on paketti salattava.

Menetelmäksi suositellaan AES-salattua ZIP-pakettia. Näitä paketteja voidaan luoda ja purkaa esimerkiksi seuraavilla sovelluksilla:

- WinZip (Microsoft Windows)

## SÄHKE2

### Siirto-ohje

v. 1.0.2  
2.10.2009

23 (23)

- 7Zip (Microsoft Windows, Linux ja muut Unix-pohjaiset käyttöjärjestelmät)
- Xceed (Microsoft Windows)
- DotNetZip (Microsoft Windows)

Tällöin arkistolaitoksen kanssa tulee sopia käytettävistä avaimista sekä niiden suhteen käytettävästä politiikasta. Suositeltavinta on käyttää salaukseen arkistolaitoksen tarkoitusta varten jakamaa julkista avainta.

## 5 LUOVUTUSPAKETIN SIIRTÄMINEN

Luovutuspaketin siirtotavasta ja käytännön järjestelyistä tulee sopia arkistolaitoksen kanssa pitkäaikaissäilytyksestä tehtävän sopimuksen yhteydessä. Online-siirto voi tapahtua lähettämällä paketti (S)FTP-protokollalla arkistolaitoksen osoittamalle palvelimelle tai Web Services -kutsulla VAPA-järjestelmään. Offline-siirto voidaan tehdä optisella levyllä, USB-kiintolevyllä tai muulla arkistolaitoksen kanssa sovitulla medially. Samassa yhteydessä tulee sopia siirron aikataulusta.

### 5.1 Online-siirto

Online-siirrossa on suositeltavaa suojata yhteys SSL-salauksella. Tällöin käytetään SFTP- tai HTTPS-protokollaa. Salaukseen käytettävä asiakasvarmenne täytyy hankkia luotetulta kolmannelta osapuolelta, esimerkiksi Väestörekisterikeskukselta. Käytetystä varmenteesta täytyy sopia pitkäaikaissäilytyksestä tehtävän sopimuksen yhteydessä. Arkistolaitos opastaa yhteyden muodostamisen teknisissä järjestelyissä kun palvelu otetaan käyttöön.

Mikäli käytetään salaamatonta yhteyttä, täytyy luovutuspaketti salata. Salaukseen käytetyt avaimet vaihdetaan palvelun käyttöönoton yhteydessä. Käytetystä salausmenetelmästä sovitaan ennen palvelun käyttöönottoa.

### 5.2 Offline-siirto

Offline-siirrossa luovutuspaketti kopioidaan siirrettävälle medially, esimerkiksi optiselle levyllä (CD-ROM tai DVD-ROM) tai USB-levyasemalle. Paketti on syytä salata siirron ajaksi. Menetelmäksi suositellaan ZIP-paketin salausta.

Offline-siirrettäessä aineistoa, on arkistolaitoksen kanssa sovittava:

- Siirtomediasta
- Salaustavasta ja avainten hallinnasta
- Siirtomenetelmästä (kuriiri/posti/tms.)

Siirtomediaa valmistellessa on huomioitava:

- Optisen median sulkeminen. Medially ei tule jättää avoimia sessioita.
- Siirrettävän kiintolevyn tiedostojärjestelmä. Suositeltuja järjestelmiä ovat:
  - NTFS
  - FAT32 (pienille medioille)

Muista tiedostojärjestelmistä voidaan tarvittaessa sopia erikseen.