



SÄHKE2-tietoaaineistot – siirtopaketin muodostaminen

Sisällys

1.	OHJEEN TARKOITUS.....	1
2.	SIIRRETTÄVÄ TIETOAINEISTO JA SIIRTOERÄT	2
	SIIRTOERIEN MUODOSTAMINEN	2
	TIETOAINEISTON METATIEDOT.....	3
	TIEDOSTOT JA TIEDOSTOMUODOT.....	3
	Siirrettävät tiedostot	3
	Tiedostojen metatiedot.....	5
	SIIRRON METATIEDOT.....	6
	Kontaktitiedot.....	6
	Siirtokokonaisuuden identifiointitiedot	7
	MUUT METATIEDOJA KOSKEVAT LISÄVAATIMUKSET.....	7
3.	SIIRTOPAKETIN MUODOSTAMINEN	8
	HAKEMISTORAKENNE	8
	SIIRTOPAKETTI	9

1. Ohjeen tarkoitus

Ohjeessa kuvataan, miten Kansallisarkistoon siirrettävä SÄHKE2-siirtopaketti muodostetaan. Ohje on yleiskäyttöinen, sillä tietojärjestelmien siirtotoiminnallisuudet ja -ratkaisut poikkeavat toisistaan. Ohjetta tulee hyödyntää yhdessä taulukossa 1 lueteltujen dokumenttien kanssa.

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

Taulukko 1 Viittaukset muihin ohjeisiin ja dokumentteihin.

Dokumentti	Sisältö
Suositus SÄHKE2-metatietomallin hyödyntämisestä	Kansallisarkiston SÄHKE2-suositus antaa metatietoihin ja metatietomalliin (SÄHKE2-metatietomalli, versio 1.3, 18.3.2020) liittyviä soveltamisohjeita. Suositus käsittelee SÄHKE2-määrittelyjä erityisesti tiedon arkistoinnin näkökulmasta sekä kommentoi SÄHKE2-metatietomallia lainsäädännön muutoksiin liittyvillä huomioilla.
SÄHKE2-siirtotiedoston XML-skeema	Siirrettävien SÄHKE2-metatietojen esittäminen XML-muodossa - versio 2011 - versio 2019
Vastaanotettavia tiedostoja koskevat ohjeet -ohje ja sen LIITE 1 Luettelo Kansallisarkistoon vastaanotettavista tiedostomuodoista sekä LIITE 2 Tekniset vaatimukset AV-tiedostoille	Ohjeessa kuvataan Kansallisarkistoon vastaanotettavia tiedostoja koskevat yleiset vaatimukset, tekstitiedostoissa sallitut merkistöt ja siirtokelpoiset tiedostoformaatit. Ohjeeseen on eritelty, missä siirtorakenteessa mitään tiedostomuotoa voi siirtää. Ohje on Kansallisarkiston verkkosivuilla.
Metatietolomake	Siirtäjä ilmoittaa metatietolomakkeella sähköisen arkistoinnin palvelulle siirtoerää koskevat kuvailutiedot. Siirtäjä voi hyödyntää metatietolomaketta siirrettävän aineistokokonaisuuden rakenteen, arkistollisen kontekstin kuvailutietojen ja käyttörajoitusten kokonaisuuden hahmottamisessa. Lomake on Kansallisarkiston verkkosivuilla.

2. Siirrettävä tietoaaineisto ja siirtoerät

Siirrettävä tietoaaineisto yksilöidään Kansallisarkiston ja Siirtäjän välisessä **siirtosuunnitelmassa**, jonka Siirtäjä on toimittanut Kansallisarkistoon. Siirrosta sopimisen ja siirtosuunnitelman perusteella siirrettävästä tietoaaineistosta koostetaan yksi tai useampi **siirtoerä**. Yksittäinen siirtoerä on looginen, ehyt ja valmis kokonaisuus, jonka aineistoihin tai tietoihin ei kohdistu enää muokkaustarpeita.

Siirtoerä muodostuu siirrettävään tietoaaineistoon kuuluvista tiedostoista sekä metatiedoista, jotka koskevat tietoaaineistoa, tiedostoja ja siirtoa. Metatiedoista tuotetaan XML-tiedosto (**sahke.xml**).

Siirtoerä tallennetaan ohjeen mukaiseen hakemistorakenteeseen ja paketoidaan siirtoa varten yhdeksi tiedostoksi (**siirtopaketti**), katso luku 3.

Siirtäjä saa Kansallisarkistolta hakemiston ja siirtopaketin nimeämisessä käytettävän tunnisteiden (**metatietotunniste**). Tunniste ohjaa siirrettävän aineiston osaksi oikeaa arkistollista kuvailukokonaisuutta Kansallisarkiston metatietovarannossa.



Siirtoerän koostamisessa voidaan hyödyntää esimerkiksi tietojärjestelmän omaa siirto-toiminnallisuutta tai siirtoa varten erikseen luotua ratkaisua. Kansallisarkisto ei tarjoa työkaluja siirtoerän koostamiseen, vaan Siirtäjän on tehtävä **aineistopoiminta** käytös-sään olevilla työkaluilla tai luotava tarvittava ratkaisu sovellustoimittajansa/palvelun-tarjoajansa tai muun teknisen kumppanin kanssa.

Tietoaaineisto voidaan jakaa siirtoeriksi esimerkiksi tehtävuokittain (tai asiaryhmittäin) ja erilaisten aikarajausten mukaisesti. Siirtäjä voi siirtosuunnittelun yhteydessä perus-tella mahdollisia tapoja jakaa aineisto siirtoeriin. Aineiston ryhmittelyyn ja kuvailuun liittyvät vaatimukset Kansallisarkistossa asettavat kuitenkin reunaehdot siirtoerien muodostamiselle.

Kun tietoaainestoa siirretään vaiheittain (erissä) Siirtäjän tulee huolehtia siitä, että jokaisen siirtoerän osalta kaikki arkistoitavaksi määrätty tietoaainesto tulee siirretyksi.

Huom!

- Siirtoerä on esimerkiksi tehtäväluokan tai asiaryhmän ja ajan mukaan rajattu eheä kokonaisuus.
- Siirtoerään kuuluvien asiakirjojen säilytysaika on pysyvä säilytys (ArkL.)/arkistointi (TihL.).
- Siirrettävien asioiden käsittely on päättynyt ja asiakirjat kuuluvat ratkaistuun asiaan.
- Asiakirjat ovat tilatiedoltaan valmiita.
- Asiakirjojen tietosisältöä ei enää muokata siirtävässä järjestelmässä.

Tietoaaineston metatiedot

Tietoaaineston metatiedot (SÄHKE2-metatietomalli, versio 1.3, 18.3.2020, 2–5) kuvaavat aineiston rakennetta ja asiakirjojen sisältöjä. Metatiedot koostuvat:

- kunkin siirrettävän asiakirjallisen tiedon käsittelyprosessin (asian) metatiedoista
- kuhunkin käsittelyprosessiin liittyvien toimenpiteiden metatiedoista
- kuhunkin toimenpiteeseen liittyvien asiakirjojen metatiedoista (kuten käyttörajoituksista) ja
- asiakirjatiedostojen metatiedoista.

Metatietojen käyttötarkoitus, sisältö ja rakenne on määritelty SÄHKE2-metatietomallissa. Metatiedot tuotetaan SÄHKE2.xsd-skeeman mukaiseksi XML-rakenteeksi.

SÄHKE2-tietoaainestojen arkistollisiksi metatiedoiksi edellytetään vähimmäismetatietoja, jotka on lueteltu suosituksessa SÄHKE2-metatietomallin hyödyntämisestä. Mikäli Kansallisarkistoon siirrettävälle aineistolle ei ole muodostettu riittäviä metatietoja jo sen käsittelyn aikana, voi Siirtäjä täydentää niitä ennen siirtoa tai tietyn rajoituksen vielä siirron yhteydessä siirtokäyttöliittymässä. Tarvittaessa Siirtäjää ohjeistetaan metatietojen tuottamisessa ennen siirtoa.

Tiedostot ja tiedostomuodot

Siirrettävien tiedostojen tulee olla Kansallisarkiston hyväksymissä tiedostomuodoissa. Ajantasaiset tiedostoja koskevat vaatimukset sekä luettelo vastaanotettavista tiedostomuodoista ovat Kansallisarkiston verkkosivuilla tiedostoja koskevassa ohjeessa.

Siirrettävät tiedostot

Siirrettävä asiakirja voi koostua yhdestä tai useammasta tiedostosta.

Samasta asiakirjasta voi olla useita eri tiedostomuotoja, joiden käyttötarkoitus yksilöidään SÄHKE2-metatiedoissa UseType-elementillä (SÄHKE2-metatietomalli, versio 1.3, 18.3.2020, 6.1.2: Käyttötarkoitus). Natiivi-tiedostoksi kutsutaan alkuperäistä tiedostomuotoa, jossa tiedosto on luotu ja jossa sitä on käsitelty. Natiivi-tiedostosta on voitu tiedon elinkaaren jossain vaiheessa tuottaa pitkäaikaissäilytykseen tarkoitettu Arkistomuoto (usein PDF/A-tiedosto) tai asiakirjan jakeluun tai julkaisuun tarkoitettu Julkaisumuoto (esimerkiksi PDF/UA).

Asiakirja otetaan Kansallisarkistoon vastaan pääsääntöisesti vain yhdessä tiedostomuodossa, eli ei useita samansisältöisiä tiedostoja, ja kustakin tiedostosta otetaan vain yksi, valmis versio. On suositeltavaa siirtää asiakirja alkuperäisessä tiedostomuodossaan aina, kun se on mahdollista. Konversio arkistointia varten tulisi tehdä ainoastaan silloin, kun alkuperäinen tiedostomuoto ei ole Kansallisarkistoon vastaanotettavien tiedostoformaattien joukossa.

Asiakirjaan on aina liityttävä Natiivi- ja/tai Arkisto-tiedostomuoto. Siirtopaketista otetaan talteen ensisijaisesti Natiivi-tiedostot, mutta mikäli asiakirjalla ei ole Natiivi-tiedostoa, otetaan talteen Arkisto-tiedosto. Julkaisu-käyttötarkoituksella merkityt tiedostot ei siirrossa oteta lainkaan talteen Kansallisarkiston järjestelmiin.

Jos asiakirjaan liittyy useita tiedostoja, joilla kaikilla on sama käyttötarkoitus (Natiivi tai Arkisto), asiakirjan oletetaan koostuvan kaikista näistä tiedostoista, ja ne kaikki otetaan siirrossa talteen.

Tiedostojen vastaanottamista on havainnollistettu alla taulukoissa 2 ja 3.

Taulukko 2. Samasta tiedostosta on kaksi sisällöltään ei-identtistä tiedostoversiota eri Use-Type-merkinnällä. Vain Arkisto-versio halutaan siirtää arkistoon.

Siirtopaketin tiedostot	Mitä siirrossa tapahtuu?	Toimintatapa
Tiedosto A: xyz.doc (allekirjoittamaton Natiivi-versio)	Siirtosovellus valitsee siirrossa automaattisesti Natiivi-tiedoston (Word-tiedosto) ja poistaa Arkisto-tiedoston (PDF-tiedosto).	Natiivi-tiedostoa ei saa sisällyttää siirtopakettiin lainkaan.
Tiedosto B: xyz.pdf (allekirjoitettu Arkisto-versio)		

Taulukko 3. Samasta tiedostosta on kaksi sisällöltään identtistä, mutta käytettävyydeltään erilaatuista tiedostoversiota eri UseType-merkinnällä. Vain Natiivi-versio halutaan siirtää arkistoon.

Siirtopaketin tiedostot	Mitä siirrossa tapahtuu?	Toimintatapa
Tiedosto A: xyz.xlsx (Natiivi-versio)	Siirtosovellus valitsee siirrossa automaattisesti Natiivi-tiedoston (Excel-tiedosto) ja poistaa Arkisto-tiedoston (PDF/A-tiedoston).	Molemmat tiedostot voivat sisältyä siirtopakettiin.
Tiedosto B: xyz.pdf/a (Arkisto-versio)		

Mikäli samasta asiakirjasta on tietojärjestelmässä useampi tiedosto eri käyttötarkoituksilla (UseType), on Siirtäjän vastuulla siirtopakettia muodostaessaan huomioida ajantasaainen ohjeistus ja varmistaa, että Kansallisarkistoon siirretään haluttu tiedosto. Tarvittaessa Siirtäjää ohjeistetaan siirtopaketin tiedostomuodoista ennen paketin muodostamista.

Tiedostojen metatiedot

Tiedostojen metatiedot (SÄHKE2-metatietomalli, versio 1.3, 18.3.2020, 6.4: Siirrettävän kohteen metatiedot) koskevat tiedostoja ja niiden eheyttä sekä sijaintia siirtopaketissa. Tietojen avulla luodaan yhteys tiedostojen ja niitä koskevien aineistometatietojen (sahke.xml) välille. Tiedostojen ja aineistometatietojen on vastattava toisiaan. Toisin sanoen sahke.xml-tiedostossa on oltava kaikkien siirtopakettiin kuuluvien tiedostojen tiedot, eikä siirtopaketissa saa olla tiedostoja, joiden metatietoja ei ole xml-tiedostossa.

SÄHKE2-rakenteessa siirrettävien tiedostojen metatiedot esitetään *Dokumentti*-tasolla. Pakolliset metatiedot on koottu taulukkoon 4.

Taulukko 4. Tiedostojen pakolliset metatiedot

SÄHKE2-elementti	Selitys	Huomautus	Esimerkki
File.Path (Hakemistopolku)	Tiedoston fyysinen hakemistopolku suhteessa Sähke-rakenteen juureen	Absoluuttista viittausta ei saa käyttää (esim. c:\materiaalit\siirto\aineistot\123\abc)	PDF\F0055720.PDF PDF\F0055720.PDF
HashAlgorithm (Eheystunnisteen algoritmi)	Eheyden varmistamiseen käytetty algoritmi	Sallitut: MD5 SHA-1 SHA-256	md5 sha1 sha256
HashValue (Eheystunnisteen arvo)	Eheystunnisteen arvo		24sfstetew33dfd
NativeId (Kohteen yksilöivä tunniste)	Yksilöi siirrettävän tiedoston	Tunnisteen on oltava yksilöllinen sahke.xml-tiedoston sisällä.	

SÄHKE2-aineistojen ja -siirtopakettien tiedostonimissä sekä hakemistoissa pitää huomioida taulukossa 5 yksilöidyt rajaukset.

Taulukko 5. Tiedostonimiä ja hakemistoja koskevat rajoitukset

Nimeämisen rajoitteet	Vaatus/selitys
Tiedostonimen maksimipituus	256 merkkiä
Sallitut merkit	kirjaimet a-ö, A-Ö numerot 0-9 erikoismerkit - _ . () # välilyönti
Kirjainkoko	Hakemistoissa ja sahke.xml-tiedostossa esitettyt tiedot, myös tiedostonimet tiedostopäätteineen, ovat merkkikokoriippuvaisia (engl. <i>case-sensitive</i>). ESIMERKKI: Tiedoston nimi on <i>F0055720.pdf</i> , joten siihen kohdistuva viittaus <i>F0055720.PDF</i> on virheellinen.
XML-tiedoston tiedostonimi	Metatiedoista tuotetun XML-tiedoston nimi on aina sahke.xml

Siirron metatiedot

Siirron metatiedot (SÄHKE2-metatietomalli, versio 1.3, 18.3.2020, 6.1 ja 6.2) yksilöivät siirtoerän, siitä tuotetun siirtopaketin ja Siirtäjän. Tiedot perustuvat Siirtäjän ja Kansallisarkiston väliseen sopimukseen, ja ne tuotetaan siirtoerän muodostamisen yhteydessä.

Kansallisarkisto toimittaa tarvittavat tunnisteet (kts. taulukko 7) Siirtäjälle siirtosuunnitelun yhteydessä.

Kontaktitiedot

Kontaktitietoihin (taulukko 6) sisältyvät siirtävän tahon organisaatitiedon lisäksi siirrosta vastaavan teknisen yhteyshenkilön tiedot. Tiedot tukevat siirtotapahtumaa, eikä niitä hyödynnetä aineiston myöhemmässä käytössä.

Taulukko 6. Siirtoon liittyvät kontaktitiedot (SÄHKE2 6.2)

SÄHKE2-elementti	Vaatus/Selitys
Organisaatio (SÄHKE 6.2.1)	Siirtoerän siirtäneen organisaation (viranomaisen) nimi.
Kontaktihenkilö (SÄHKE 6.2.2)	Siirrosta vastaavan organisaation tekninen kontaktihenkilö yhteystietoineen.

Siirtokokonaisuuden identifiointitiedot

Siirtokokonaisuuden identifiointitiedot (taulukko 7) ovat pakollisia teknisiä metatietoja, jotka yksilöivät siirron ja siirrettävän kokonaisuuden.

Taulukko 7. Siirtoerän pakolliset tunnistetiedot (SÄHKE 6.1)

SÄHKE2-elementti	Vaatus/Selitys
Identifiointitunnus (SÄHKE 6.1.1 <i>Identifier.Nativeld</i>)	Uniikki identifiointitunnus yksilöi siirtoerän sähköisen arkistoinnin palvelussa. Identifiointitunnuksen pitää olla OID-muotoinen, ja se tulee muodostaa Kansallisarkiston esittämässä rakenteessa. OID-tunnus perustuu sekä Kansallisarkiston antamaan siirtosuunnitelman numeroon että Siirtäjän tuottamiin siirron yksilöintitunnuksiin. Tunnuksen rakenne on muotoa urn:oid:1.2.246.582.200.[siirtosuunnitelman asianumero + vuosi]. [siirron vuosi].[nnnn]. Tunnuksen alkuosa on kiinteä ja aina sama. Kansallisarkisto toimittaa Siirtäjälle hyödynnettävän siirtosuunnitelman numeron, vuosiosa ilmoittaa siirron kuluvan vuoden kokonaislukuna ja juokseva numero on Siirtäjän määrittämä merkkijono, joka yksilöi siirtopaketin. Merkkijono voi perustua esim. siirrettävän asialuokan tunnukseen. Esimerkki: <Nativeld>urn:oid:1.2.246.582.200.12352023.2024.290101</Nativeld>
Aineiston metatietojen skeema (SÄHKE 6.1.5)	Osoite, jota on käytetty siirrettävän aineiston metatietojen validointiin. 2019 SÄHKE2: http://www.arkisto.fi/skeemat/Sahke2_2019_03.xsd 2011 SÄHKE2: http://www.arkisto.fi/skeemat/Sahke2_2011_12.xsd
Siirto- ja säilytys sopimuksen tunnus (SÄHKE 6.1.4 <i>TransferContractId</i>)	Kansallisarkiston toimittama siirtosopimuksen diaarinumero, esim. KA/1234/29.01.01/2023.
Nimeke (SÄHKE 6.1.3 <i>Title</i>)	Tietojärjestelmä, josta aineisto siirretään.

Muut metatietoja koskevat lisävaatimukset

Alla olevaan taulukkoon (taulukko 8) on koottu siirtoon liittyvät SÄHKE2-skeeman metatietoja koskevat tarkennukset.

Taulukko 8. Asia- ja asiakirjatason pakolliset tiedot

SÄHKE2-elementti	Vaatus/Selitys
Identifiointitunnus (SÄHKE 2.3 Identifier.Nativeld)	Tieto ei saa olla sahke.xml:n CaseFile- tai Record-tasolla tyhjä.
Nimeke (SÄHKE 2.7 Title)	Tieto ei saa olla tyhjä. sahke.xml:n CaseFile- ja Record -tasolla nimekkeen enimmäispituus on 255 merkkiä.

Salassapitoperuste (SÄHKE 2.6.4 Restriction.SecurityReason) ei SÄHKE2-skeemassa ole pakollinen, mutta metatietomallin mukaan se on ehdollisesti pakollinen, eli jos Julkisuusluokka on muu kuin *Julkinen*, tulee salassapitoperuste antaa. Mikäli salassapitoperusteita ei ole sahke.xml:ään tuotettu, on ne määriteltävä asioille ja asiakirjoille viimeistään siirron yhteydessä.

3. Siirtopaketin muodostaminen

Hakemistorakenne

sahke.xml-tiedosto ja asiakirjatiedostot pitää tallentaa taulukon 9 mukaiseen **hakemistorakenteeseen**.

- sahke.xml-tiedosto on pakollinen.
- Asiakirjatiedostot, joihin sahke.xml-tiedostossa viitataan, ovat pakollisia.

sahke.xml-tiedosto tallennetaan juurihakemistoon. Juurihakemisto pitää nimetä Kansallisarkiston toimittamalla tunnisteella (metatietotunniste). Tunniste on numerosarja, joka on esimerkiksi muotoa 3222566740. Tunniste ei saa sisältää virheitä, joten se on hyvä kopioida ja liittää hakemiston nimeen.

Mahdolliset asiakirjatiedostot tallennetaan yhteen tai useampaan alihakemistoon.

Taulukko 9. Hakemistojen nimet ja sisältö

Hakemisto	Selite/sisältö
juurihakemisto, siirtopaketin tunniste	Hakemisto sisältää: • sahke.xml-tiedosto (pakollinen) Juurihakemisto pitää nimetä Kansallisarkiston toimittamalla tunnisteella (metatietotunniste).
alihakemisto tai alihakemistot	Hakemisto sisältää: • asiakirjatiedostot (ehdollisesti pakollisia) Hakemistossa on oltava kaikki tiedostot, joihin sahke.xml-tiedostossa viitataan, eikä siihen saa sisältyä tiedostoja, joihin sahke.xml:ssä ei viitata.

Siirtopaketti

Juurihakemisto nimetään siirtopaketin tunnisteella. Hakemistot tiedostoineen paketoidaan yhdeksi TAR-tiedostoksi. TAR-tiedoston saa lisäksi pakata häviöttömään GZIP- tai BZIP2-muotoon.

Kun siirtopaketti ladataan siirtokäyttöliittymään, sen nimenä käytetty tunniste ohjaa tietoaineiston osaksi oikeaa kuvailukokonaisuutta Kansallisarkiston metatietovarannossa.

Kuva 2. Siirtopaketti ja hakemistorakenne. Tiedostot tallennetaan siirtopakettia varten kuvan mukaiseen hakemistorakenteeseen. Hakemistot tiedostoineen paketoidaan yhdeksi TAR-tiedostoksi. Juurihakemisto ja siirtopaketti pitää nimetä Kansallisarkiston toimittamalla tunnisteella (metatietotunniste).

