



Kansallisarkisto

SÄHKE2-aineistot - siirtopaketin muodostaminen



Sähköisen arkistoinnin palvelu

6.2.2023

Sisällys

1. OHJEEN TARKOITUS	2
1.1. VIITTAUKSET MUIHIN OHJEISIIN JA DOKUMENTTEIHIN	2
2. SIIRRETTÄVÄ AINEISTO JA SIIRTOERÄT	2
2.1. SIIRTOERIEN MUODOSTAMINEN	3
2.2. AINEISTON METATIEDOT	4
2.3. TIEDOSTOT JA TIEDOSTOMUODOT	4
2.4. SIIRRON METATIEDOT	7
2.5. MUUT METATIETOJA KOSKEVAT LISÄVAATIMUKSET	8
3. SIIRTOPAKETIN MUODOSTAMINEN	8
3.1. HAKEMISTORAKENNE.....	8
3.2. 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 kohdassa 1.1. lueteltujen dokumenttien kanssa (taulukko 1).

1.1. Viittaukset muihin ohjeisiin ja dokumentteihin

Taulukko 1. Viittaukset muihin ohjeisiin ja dokumentteihin.

Dokumentti	Sisältö
Suositus SÄHKE2-metatietomallin hyödyntämisestä	Kansallisarkiston suositus antaa SÄHKE2-määräyksen metatietoihin ja metatietomalliin (SÄHKE2-metatietomalli, versio 1.3, 18.3.2020) liittyviä soveltamisohjeita. Suositus käsittelee SÄHKE2-määrityksiä erityisesti tiedon arkistoinnin näkökulmasta sekä kommentoi SÄHKE2-metatietomallia lainsäädännön muutoksiin liittyvillä huomioilla.
SÄHKE2-siirtotiedoston XML-skeema	Siirrettävien SÄHKE2-metatietojen esittäminen XML-muodossa <ul style="list-style-type: none">- versio 2011- versio 2019
Vastaanotettavia tiedostoja koskevat ohjeet ja sen LIITE 1 Luetelo 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 siirtoarokenteessa mitään tiedostomuotoa voi siirtää. Ohje on Sähköisen arkistoinnin palvelun kotisivuilla.

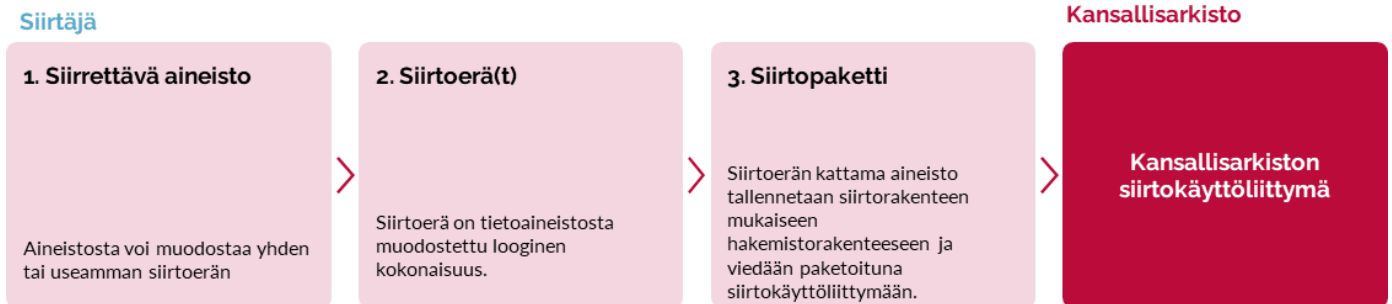
2. Siirrettävä aineisto ja siirtoerät

Siirrettävä aineisto yksilöidään Kansallisarkiston ja **Siirtäjän** välisessä **siirtosuunnitelmassa**, jonka Siirtäjä on toimittanut Kansallisarkistoon. Siirrosta sopimisen ja siirtosuunnitelman perusteella siirrettävästä aineistosta koostetaan yksi tai useampi **siirtoerä**.

Siirtoerä muodostuu siirrettävään aineistoon kuuluvista tiedostoista sekä metatiedoista, jotka koskevat aineistoa, 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 tunniste (metatietotunniste). Tunniste ohjaa siirrettävän aineiston osaksi oikeaa arkistollista kuvailukokonaisuutta Kansallisarkiston metatietovarannossa.



Kuva 1. Siirrettävä aineisto siirretään yhdessä tai useammassa siirtoerässä. Jokaisesta siirtoerästä muodostuu oma siirtokäyttöliittymään vietävä siirtopakettinsa.

2.1. Siirtoerien muodostaminen

Siirtämistä selkeyttää ja helpottaa siirrettävän aineiston jakaminen useisiin itsenäisiin siirtoeriin.

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** käytössään olevilla työkaluilla tai luotava tarvittava ratkaisu sovellustoimittajansa/palveluntarjoajansa tai muun teknisen kumppanin kanssa.

Yksittäinen siirtoerä on looginen, ehyt ja valmis kokonaisuus, jonka aineistoihin tai tietoihin ei kohdistu enää muokkaustarpeita. Pienin siirrettävä kokonaisuus on yksi asiakirjallisen tiedon käsittelyprosessi (**asia**). Yksittäistä asiakirjaa ei voi siirtää ilman sen käsittelyyn liittyvää käsittelyprosessia.

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

- Siirtoerä on esimerkiksi tehtävuokan 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.

Kun aineistoa siirretään vaiheittain (erissä) Siirtäjän tulee huolehtia siitä, että jokaisen siirtoerän osalta kaikki arkistoitavaksi määrätty aineisto tulee siirrettyksi.

2.2. Aineiston metatiedot

Aineiston metatiedot (**SÄHKE2 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 (katso taulukko 1).

Sähke-elementit tuotetaan XML-rakenteeksi SÄHKE2.xsd-skeeman mukaisesti.

SÄHKE2-tietoaineistojen arkistollisiksi metatiedoiksi edellytetään suosituksessa SÄHKE2-metatietomallin hyödyntämisestä lueteltuja vähimmäismetatietoja (katso taulukko 1). Mikäli Kansallisarkistoon siirrettävälle aineistolle ei ole muodostettu riittäviä metatietoja jo sen käsittelyn aikana, voi siirtäjä täydentää niitä ennen siirtoa tai tietyin rajoituksin vielä siirron yhteydessä siirtokäyttöliittymässä. Tarvittaessa Siirtäjää ohjeistetaan metatietojen tuottamisessa ennen siirtoa.

2.3. Tiedostot ja tiedostomuodot

Siirrettävien tiedostojen tulee olla Kansallisarkiston hyväksymissä tiedostomuodoissa. Ajantasaiset tiedostoja koskevat vaatimukset ja luettelo vastaanotettavista tiedostomuodoista ovat Kansallisarkiston verkkosivuilla tiedostoja koskevassa ohjeessa (katso taulukko 1).

2.3.1. 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 6.1.2, Käyttötarkoitus). Natiivi-tiedostoksi kutsutaan alkuperäistä tiedostomuotoa, jossa tiedosto on luotu ja jossa sitä on käsitelty. Natiivi-tiedostosta on voitu tiedon linkaaren jossain vaiheessa tuottaa pitkäaikaissäilytykseen tarkoitettu Arkisto-muoto (usein PDF/A-tiedosto) tai asiakirjan jakeluun tai julkaisuun tarkoitettu Julkaisu-muoto (esimerkiksi PDF/UA)

Asiakirja otetaan Kansallisarkistoon vastaan pääsääntöisesti vain yhdessä tiedostomuodossa. On suositeltavaa siirtää asiakirja alkuperäisessä tiedostomuodossaan aina, kun se on mahdollista; yleisesti ottaen 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 oteta siirrossa talteen Kansallisarkiston järjestelmiin lainkaan.

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-version halutaan siirtyvän arkistoon.

Siirtopaketin tiedostot	Mitä siirrossa tapahtuu?	Oikea toimintatapa
Tiedosto A: xyz.doc (allekirjoittamaton Natiivi-versio)	Siirto-sovellus 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-version halutaan siirtyvän arkistoon.

Siirtopaketin tiedostot	Mitä siirrossa tapahtuu?	Oikea toimintatapa
Tiedosto A: xyz.xlsx (Natiivi-versio)	Siirto-sovellus 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)		

Siirtäjän vastuulla on siirtopakettia muodostaessaan huomioida ajantasainen ohjeistus ja varmistaa, että Kansallisarkistoon siirretään haluttu tiedosto, mikäli samasta asiakirjasta on tietojärjestelmässä useampi tiedostoversio. Siirtopaketin tiedostomuodoista ohjeistetaan Siirtäjää tarvittaessa ennen paketin muodostamista.

2.3.2. Tiedostojen metatiedot

Tiedostojen metatiedot (**SÄHKE2 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-tiedossa on oltava kaikkien siirtopakettiin kuuluvien tiedostojen tiedot, eikä siirtopaketissa saa olla tiedostoja, joiden metatietoja ei ole xml-tiedostossa.

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

Taulukko 4. Tiedostojen pakolliset metatiedot

SÄHKE2-elementti	Selitys	Huomaus	Esimerkki
File.Path (Hakemistopolku)	Tiedoston fyysinen hakemistopolku suhteessa Sähke-rakenteen juureen	Absoluuttista viittausta (esim. c:\materiaalit\siirto\aineistot\123\abc) ei saa käyttää.	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
Nativeld (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 rajoitukset.

Taulukko 5. Tiedostonimiä ja hakemistoja koskevat rajoitukset

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

2.4. Siirron metatiedot

Siirron metatiedot (**SÄHKE2 6.1 ja 6.2**) yksilöivät siirtoerän, siitä tuotetun siirtopaketin ja Siirtäjän. Tiedot perustuvat palvelun käyttäjän ja Kansallisarkiston väliseen sopimukseen.

Siirron metatiedot tuotetaan siirtoerän muodostamisen yhteydessä. Kansallisarkisto toimittaa tarvittavat tunnisteet Siirtäjälle siirtosuunnittelun yhteydessä.

2.4.1. 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.

2.4.2. 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>)	<p>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 sopimusnumeroon että Siirtäjän tuottamiin siirron yksilöintitunnuksiin.</p> <p>Tunnuksen rakenne on muoto <code>urn:oid:1.2.246.582.200.[sopimusnumero].[vuosi].[nynn1]</code>.</p> <p>Tunnuksen alkuosa on kiinteä ja aina sama. Kansallisarkisto toimittaa Siirtäjälle hyödynnettävän sopimusnumeron, vuosiosa ilmoittaa siirron kuluvan vuoden kokonaislukuna ja juokseva numero yksilöi yksittäisen siirtopaketin.</p> <p>Esimerkki: <Nativeld><code>urn:oid:1.2.246.582.200.123123.2018.0001</code></Nativeld></p>
Aineiston metatietojen skeema (SÄHKE 6.1.5)	<p>Osoite, jota on käytetty siirrettävän aineiston metatietojen validointiin.</p> <p>2019 SÄHKE2: http://www.arkisto.fi/skeemat/Sahke2_2019_03.xsd</p>

	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 sopimuksen yksilöivä tunniste, esim. 10012021.
Nimeke (SÄHKE 6.1.3 <i>Title</i>)	Tietojärjestelmä, josta aineisto siirretään.

2.5. 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 <i>Identifier.NativeId</i>)	Tieto ei saa olla CaseFile- tai Record-tasolla tyhjä.
Nimeke (SÄHKE 2.7 <i>Title</i>)	Tieto ei saa olla tyhjä. 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

3.1. Hakemistorakenne

sahke.xml-tiedosto ja asiakirjatiedostot pitää tallentaa kuvan 2 mukaiseen **hakemistorakenteeseen** (kansiorakenne).

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

sahke.xml-tiedosto tallennetaan juurihakemistoon (pääkansio). Juurihakemisto pitää nimetä Kansallisarkiston toimittamalla tunnisteella (metatietotunnus). 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 (aliansio).

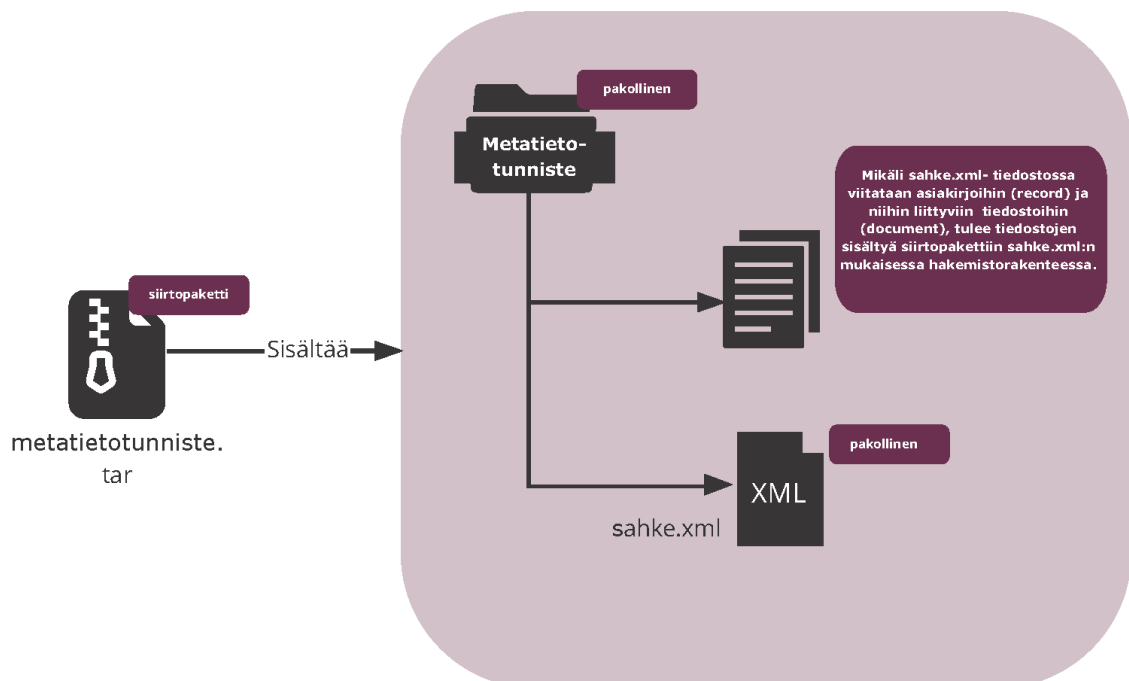
Taulukko 9. Hakemistojen nimet ja sisältö

Hakemisto	Selite/sisältö
juurihakemisto (pääkansio), siirtopaketin tunniste	Hakemisto sisältää: <ul style="list-style-type: none"> sahke.xml-tiedosto (pakollinen) Juurihakemisto pitää nimetä Kansallisarkiston toimittamalla tunnisteella (metatietotunnus).
alihakemisto tai alihakemistot	Hakemisto sisältää: <ul style="list-style-type: none"> asiakirjatiedostot (ehdollisesti pakollisia) Hakemistossa on oltava kaikki tiedostot, joihin sahke.xml-tiedostossa viitataan.

3.2. Siirtopaketti

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

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



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