Rakenteiset tietoaineistot – siirtopaketin muodostaminen

Sisällys

1. Ohjeen tarkoitus

Ohjeessa kuvataan, miten rakenteisesta datasta muodostetaan Kansallisarkistoon siirrettävä siirtopaketti.

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

Ohjetta tulee hyödyntää yhdessä taulukon 1 dokumenttien kanssa.

Taulukko 1. Viittaukset muihin ohjeisiin ja dokumentteihin

Dokumentti	**Tarkoitus**
Vastaanotettavia tiedostoja koskevat ohjeet ja sen LIITE 1 Luettelo Kansallisarkistoon vastaanotettavista tiedostomuodoista	Ohjeessa kuvataan Kansallisarkistoon vastaanotettavia tiedostoja koskevat yleiset vaatimukset, tekstitiedostoissa sallitut merkistöt ja siirtokelpoiset tiedostoformaatit. Ohje on Sähköisen arkistoinnin palvelun kotisivuilla. Ohjeeseen on eritelty, missä siirtorakenteessa mitäkin tiedostomuotoa voi siirtää.
Diaariaineistojen XML-rakenne (LIITE 1)	Alkuperäisestä diaarisovelluksesta saatavat datakoosteet tulee muodostaa ensisijaisesti tässä ohjeessa esitetyn rakenteen mukaisesti.
Metatietolomake	Siirtäjä ilmoittaa metatietolomakkeella Sähköisen arkistoinnin palvelulle siirtoerää koskevat metatiedot. Siirtäjä voi hyödyntää metatietolomaketta siirrettävän aineistokokonaisuuden rakenteen, arkistollisen kontekstin kuvailutietojen ja käyttörajoitusten kokonaisuuden hahmottamisessa. Lomake on Kansallisarkiston verkkosivuilla.

2. Siirrettävä tietoaineisto ja siirtoerät

Siirrettävä kokonaisuus: rakenteinen data

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

Siirrettävä kokonaisuus koostuu varsinaisesta rakenteisessa muodossa olevasta **data-aineistosta** ja siihen liittyvästä **oheisdokumentaatiosta** (katso kappale 4. Aineiston kuvailu ja dokumentointi). Varsinainen data-aineisto koostuu yleensä yhdestä tai useammasta tiedostosta (**datakooste tai SIARD-tiedosto**).

Kansallisarkisto Rauhankatu 17 PL 258, 00171 Helsinki Puh. 029 533 7000 kirjaamo@kansallisarkisto.fi
Riksarkivet Fredsgatan 17 PB 258, 00171 Helsingfors Tfn kansallisarkisto.fi

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

Datakoosteissa ja oheisdokumentaatiossa käytettyjen tiedostojen on oltava Kansallisarkiston hyväksymässä muodossa.

Aineiston poiminta

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

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

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

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

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

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

Siirtoerien muodostaminen

Siirrettävä tietoaineisto yksilöidään **siirtosuunnitelmassa**, jonka Siirtäjä on toimittanut Kansallisarkistoon. Siirrosta sopimisen ja siirtosuunnitelman perusteella siirrettävästä tietoaineistosta koostetaan yksi tai useampi **siirtoerä**. Yksittäinen

Kansallisarkisto Rauhankatu 17 PL 258, 00171 Helsinki Puh. 029 533 7000 kirjaamo@kansallisarkisto.fi
Riksarkivet Fredsgatan 17 PB 258, 00171 Helsingfors Tfn kansallisarkisto.fi

siirtoerä on looginen, ehyt ja valmis kokonaisuus, jonka aineistoihin tai tietoihin ei kohdistu enää muokkaustarpeita.

Siirtoerä jaetaan yhteen tai useampaan siirrettävään kokonaisuuteen, joka tallennetaan ohjeiden mukaiseen hakemistorakenteeseen ja paketoidaan siirtoa varten tiedostoksi (**siirtopaketti**), katso luku 6. Esimerkiksi diaariaineiston vuosittainen siirtoerä voidaan jakaa siirtopaketteihin tehtäväryhmittäin.

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

3. Formaattikohtaiset vaatimukset

XML (Extensible Markup Language)

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

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

XML:n rakenne tulee aina muodostamaan skeeman avulla. Näin varmistetaan siirrettävän tietoaineiston rakenteen säilyminen pitkäaikaissäilytyksessä. Tämän ohjeen liitteessä 1 on Kansallisarkiston diaariaineistolle määrittelemä XML-rakenne. Sen käyttäminen on suositeltavaa, mutta ei pakollista, mikäli Siirtäjällä esimerkiksi on valmiina jotain muuta tarkoitusta varten määritelty skeema.

Datakoosteessa viitataan käytettyyn skeemaan suoralla viittauksella siirtopaketin hakemistorakenteeseen, esim. xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="../schemas ../schemas/esimerkki.xsd"

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

Kansallisarkisto Rauhankatu 17 PL 258, 00171 Helsinki Puh. 029 533 7000 kirjaamo@kansallisarkisto.fi
Riksarkivet Fredsgatan 17 PB 258, 00171 Helsingfors Tfn kansallisarkisto.fi

XML:n rakenne validoidaan vastaanoton yhteydessä skeeman avulla. XML-rakennetta koskevat vaatimukset on esitetty taulukossa 2.

Taulukko 2. XML-rakennetta koskevat vaatimukset

Hyväksyttävä versio	PRONOM: fmt/101

Huom!
Tarkista ennen siirtoa, että XML:n merkistö on siirtokelpoisessa muodossa (ISO 8859-15, UTF-8, UTF-16, UTF-32) ja ilmoitettu tiedostossa oikein. Usein oletuksena on Windows-1252-merkistö, jolloin XML-deklaraatiossa on attribuutti ”encoding=Windows-1252”. Tiedoston merkistö tulee muuntaa siirtokelpoiseksi ja sen attribuutti käytetyn merkistön mukaiseksi.

CSV (Comma Separated Values)

CSV-rakennetta voidaan hyödyntää pääasiassa silloin, kun siirrettävä tietoaineisto muodostaa yksinkertaisen taulukkomaisen kokonaisuuden. CSV-rakenne ei sovellu monimutkaisten, sisäisesti hierarkkisten ja ristiviittauksia sisältävien tietorakenteiden tallentamiseen.

CSV-muodon käyttö edellyttää aina teknisen rakenteen dokumentointia. Ilman sitä datakoosteen rakennetta ja tietosisältöä ei ole mahdollista tulkita.

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

Taulukko 3. CSV-rakennetta koskevat vaatimukset

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

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

Kansallisarkisto Rauhankatu 17 PL 258, 00171 Helsinki Puh. 029 533 7000 kirjaamo@kansallisarkisto.fi
Riksarkivet Fredsgatan 17 PB 258, 00171 Helsingfors Tfn kansallisarkisto.fi

JSON (JavaScript Object Notation)

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

Tietosisällön esittäminen perustuu JSON-muodossa tieto-objektien ja niihin liittyvien arvojen listoihin. JSON-tiedostoon tallennettavat kentät ja arvot on aina määriteltävä aineistokohtaisesti. JSON-rakennetta koskevat vaatimukset on esitetty taulukossa 4.

Taulukko 4. JSON-rakennetta koskevat vaatimukset

Hyväksyttävä versio	PRONOM: x-fmt/817

SIARD 2.1 (Software Independent Archiving of Relational Databases)

SIARD on relaatiotietokantojen arkistointiin kehitetty avoimen lähdekoodin säilytysformaatti, joka perustuu avoimiin standardeihin (Unicode, XML ja SQL:2008) sekä ZIP64-tiedostoformaattiin. Relaatiotietokannan tietojen tallentaminen SIARD-muotoon tukee niiden käytettävyyttä merkittävästi paremmin kuin ratkaisut, joissa tietokannan taulujen sisältö arkistoidaan erillisinä tiedostoina. SIARD-rakennetta koskevat vaatimukset on koottu taulukkoon 5.

Rakenteeltaan SIARD 2.1 on määrämuotoinen, pakkaamaton ZIP-tiedosto, jonka sisältö on jaettu kahteen kansioon. Yksi kansio sisältää tietokannan taulujen sisällöistä muodostetut XML-tiedostot ja toinen XML-tiedostoon tallennettuna tietokannan rakenteen ja kuvauksen. SIARD ei tallenna käyttöliittymän ominaisuuksia eikä siten ole sellaisenaan käytettävissä oleva, operatiivinen tietokanta, mutta sen sisältö voidaan palauttaa käytettäväksi tietokannanhallintajärjestelmään.

Tällä hetkellä Kansallisarkisto ottaa vastaan SIARDin versiota 2.1.1. Tähän versioon ei voi sisällyttää tietokantaan tallennettuja erillisiä tiedostoja, esimerkiksi kuvia tai tekstitiedostoja, vaan ainoastaan tietokannan taulujen kenttiin tallennettua merkkipohjaista tietoa.

Taulukko 5. SIARD-tiedostoa koskevat vaatimukset

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

Kansallisarkisto Rauhankatu 17 PL 258, 00171 Helsinki Puh. 029 533 7000 kirjaamo@kansallisarkisto.fi
Riksarkivet Fredsgatan 17 PB 258, 00171 Helsingfors Tfn kansallisarkisto.fi

4. Aineiston kuvailu ja dokumentointi

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

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

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

Oheisdokumentaatio

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

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

Dokumentaatiossa voidaan hyödyntää kuvauksia ja ohjeita, jotka tietojärjestelmästä ja aineistoista on laadittu alkuperäisen käyttötarkoituksen aikana. Dokumentaatiota on kuitenkin yleensä laajennettava ja tarkennettava vastaamaan nimenomaan arkistosiirron ja arkistoinnin vaatimuksia. Tarvittava oheisdokumentaatio arvioidaan aineistosiirtojen yhteydessä. Taulukossa 6 on lueteltu yleisimpiä esimerkkejä tarvittavista kuvauksista.

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

Kansallisarkisto Rauhankatu 17 PL 258, 00171 Helsinki Puh. 029 533 7000 kirjaamo@kansallisarkisto.fi
Riksarkivet Fredsgatan 17 PB 258, 00171 Helsingfors Tfn kansallisarkisto.fi

Taulukko 6. Oheisdokumentaation avulla kuvattavia kokonaisuuksia

Kuvaus	**Selite**
Yleiskuvaus siirrettävästä aineistosta	• vapaa kuvaus • esimerkiksi asiakirjajulkisuuskuvaus ja muut valmiit kuvaukset • muut tarvittavat dokumentit
Data-aineiston syntykonteksti	Kuvaukset toiminnasta, jonka tuloksena aineisto on muodostunut • vapaa kuvaus • esimerkiksi liittyvät lait, asetukset, normit • toimintaohjeet, toimintakertomukset, raportit tms. • muut tarvittavat dokumentit
Kuvaus data-aineiston muodostumisprosessista ja tietojen alkuperästä	• mistä lähteistä tietoja on saatu tai saadaan • miten tiedot on kerätty eri lähteistä • mahdolliset ohjeet ja säännöt, jotka määrittelevät, miten tiedot on kirjattu/tallennettu alkuperäisessä ympäristössä • miten tietoja on käsitelty
Data-aineiston tietorakenteiden kuvaus	• aineiston sisäinen tietorakenne • yhteys Kansallisarkiston seulontapäätöksessä yksilöityihin, säilytettäväksi määrättyihin tietokokonaisuuksiin • seloste siirrettävästä tietoaineistosta pois jätetyistä kentistä tai tauluista ja niiden sisällöstä Kuvaustapana voidaan käyttää graafisia tietomallikuvauksia (esim. ER-malli) ja teksti- tai taulukkodokumentteja tietojen rakenteista ja suhteista.
Data-aineiston sisältämien koodistojen ja niiden arvojoukkojen kuvaukset	• data-aineistossa hyödynnetyt yleiset koodistot ja niiden arvojoukot

Kansallisarkisto Rauhankatu 17 PL 258, 00171 Helsinki Puh. 029 533 7000 kirjaamo@kansallisarkisto.fi
Riksarkivet Fredsgatan 17 PB 258, 00171 Helsingfors Tfn kansallisarkisto.fi

	• data-aineistossa hyödynnetyt organisaation sisäiset koodistot ja niiden arvojoukot
Data-aineistossa hyödynnetyt standardit ja suositukset	• hyödynnetyt yleiset standardit ja suositukset • hyödynnetyt alakohtaiset standardit ja suositukset
Kattavat esimerkit data-aineiston käyttötavoista, käytöstä ja hyödyntämisestä	• vapaa kuvaus • ohjeet • dokumentoivat kuvakaappaukset
Siirrettävään kokonaisuuteen kuuluvat datakoosteet	• seloste (esimerkiksi taulukko) datakoosteista • viittaus tiedoston nimen ja sisällön välillä
Datakoosteiden rakenne	• datakoosteiden sisäinen tietorakenne • datakoosteiden väliset suhteet ja rakenteet

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

Esimerkki

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

Oheisdokumentaatioon tulee liittää myös seloste siirtopaketin datakoosteista sekä viittaukset tiedoston nimen ja sisällön välillä. Seloste voi olla esimerkiksi taulukkotiedosto (katso taulukko 7.), jossa on yhdessä sarakkeessa esitetty tiedoston sisältö ja toisessa sarakkeessa tiedoston nimi.

Kansallisarkisto Rauhankatu 17 PL 258, 00171 Helsinki Puh. 029 533 7000 kirjaamo@kansallisarkisto.fi
Riksarkivet Fredsgatan 17 PB 258, 00171 Helsingfors Tfn kansallisarkisto.fi

Taulukko 7. Esimerkki viittauksesta tiedoston ja sen sisällön välillä

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

Kuvailevat metatiedot ja käyttörajoitukset

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

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

5. Tiedostoja ja niiden nimeämistä koskevat vaatimukset

Tiedostoja koskevat yleiset vaatimukset, luettelo vastaanotettavista tiedostomuodoista ja tiedostoissa sallituista merkistöistä ovat Kansallisarkistoon vastaanotettavia tiedostoja koskevassa ohjeessa.

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

Datakoosteet ja oheisdokumentaatioon kuuluvat tiedostot nimetään lukujonon avulla. Lähtökohtaisesti lukujonon tulee koostua neljästä numerosta. Numerointi on juokseva niin, että ensimmäinen tiedosto saa nimen 0001, toinen 0002 jne. Katso myös taulukko 7.

Skeemat nimetään XML:ssä mainittujen skeemasijaintien tiedostonimien mukaan (ilman polkua), katso luku 3. ja taulukko 7.

6. Siirtopaketin rakenne

Hakemistorakenne

Datakoosteet, oheisdokumentaatio ja mahdolliset skeemat, joita datakoosteissa hyödynnetään, pitää tallentaa kuvan 2. mukaiseen hakemistorakenteeseen.

Kansallisarkisto Rauhankatu 17 PL 258, 00171 Helsinki Puh. 029 533 7000 kirjaamo@kansallisarkisto.fi
Riksarkivet Fredsgatan 17 PB 258, 00171 Helsingfors Tfn kansallisarkisto.fi

Juurihakemisto pitää nimetä siirtopaketin yksilöllisellä tunnisteella. Siirtäjä määrittelee tämän tunnisteen siirtokäyttöliittymässä siirtopaketin kontekstimetatietojen luonnin yhteydessä. Jotta siirtopaketti siirtyy hyväksytysti säilytettäväksi, tulee tunnisteen vastata täysin siirtopaketille siirtokäyttöliittymässä määriteltyä nimeä.

Datakoosteet, oheisdokumentaatio ja mahdolliset skeemat tallennetaan omiin **alihakemistoihinsa**.

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

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

Hakemistojen nimet ja sisältö

Juurihakemisto

Hakemisto	Selite/sisältö
juurihakemisto	Juurihakemisto pitää nimetä siirtopaketin tunnisteella, esim. *Paketti1* tai *vuodet9195*. Siirtäjä määrittelee paketin tunnisteen siirtokäyttöliittymässä siirtopaketin kontekstimetatietojen luonnin yhteydessä. Olennaista on, että tunniste yksilöi siirtopaketin muista saman siirtoerän paketeista. • Tunniste saa sisältää seuraavia merkkejä: a–z, A–Z ja 0–9. • Siirtopaketin tunnisteena ei saa käyttää Siirtäjälle kontekstimetatietojen luomista varten toimitettua metatietotunnusta.

Juurihakemistoon tulee sisällyttää alihakemistojen lisäksi taulukkomuodossa datakoosteiden tarkistussummat.

Tiedosto	Selite/sisältö
siirtopaketin tunniste.csv	CSV-tiedosto sisältää master-hakemiston sisältämien tiedostojen tiedostonimet (sarake Filenumber) ja tiedostoille lasketut tarkistussummat (sarake Hashvalue). Filenumber-sarakkeeseen ei sisällytetä tiedostopäätettä. Tarkistussumma annetaan muodossa MD5. • Tiedosto tulee nimetä siirtopaketin tunnisteella, eli samalla merkkijonolla kuin juurihakemisto, esimerkiksi *Paketti1.csv* tai *vuodet9195.csv*. • Tiedoston merkistön pitää olla UTF-8. • Kenttien ympärillä ei saa olla lainausmerkkejä.

Kansallisarkisto Rauhankatu 17 PL 258, 00171 Helsinki Puh. 029 533 7000 kirjaamo@kansallisarkisto.fi
Riksarkivet Fredsgatan 17 PB 258, 00171 Helsingfors Tfn kansallisarkisto.fi

Datakoosteet ja oheisdokumentaatio

Seuraavien alihakemistojen (master ja documentation) kohdalla tiedostojen nimeämisen periaate on aina sama. Hakemistoissa olevat tiedostot nimetään lukujonon avulla. Lähtökohtaisesti lukujonon tulee koostua neljästä numerosta. Numerointi on itsenäinen ja juokseva niin, että jokaisen alihakemiston ensimmäinen tiedosto saa nimen 0001, toinen 0002 jne. Alihakemistot itsessään nimetään sisältönsä mukaisesti master tai documentation.

Hakemisto	**Selite/sisältö**
master	Alihakemisto sisältää siirrettävän tietoaineiston datakoosteet. Hakemisto on pakollinen, ja sen on sisällettävä tiedostoja. • Hakemistoon saa tallentaa CSV-, XML- tai JSON- ja SIARD-tiedostoja tai SIARD-tiedoston. • Tiedostot nimetään lukujonon avulla (0001, 0002, 0003, 0004). SIARD-tiedosto nimetään aina 0001.siard. • Jos master-hakemiston datakoosteissa viitataan skeemoihin, pitää skeemat tallentaa schemas-hakemistoon.
documentation	Alihakemisto sisältää mahdollisen oheisdokumentaation. • Hakemistoon ei saa tallentaa XML-, CSV- tai JSON-formaatissa olevia tiedostoja tai kuvatiedostoja (TIFF, JPEG). • Tiedostot nimetään lukujonon (0001, 0002, 0003, 0004) avulla. • Seloste datakoosteista esimerkiksi taulukkomuodossa (katso taulukko 7.)

Skeemat

Alihakemisto schemas lisätään juurihakemistoon, kun aineiston muodostamiseen on käytetty skeemaa. Viittaukset siirtopaketin ulkoisiin skeemoihin eivät ole riittäviä.

Hakemisto	**Selite/sisältö**
schemas	Alihakemisto sisältää mahdolliset skeemat. Skeemat ovat pakollisia silloin, kun datakooste on muodostettu skeeman mukaisesti ja skeemaan viitataan master-hakemiston datakoosteissa. • Skeemat nimetään skeemasijaintien tiedostonimien mukaan (ilman polkua), esim. esimerkki.xsd, jos skeemaviittaus on seuraava xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="../schemas ../schemas/esimerkki.xsd"

Kansallisarkisto Rauhankatu 17 PL 258, 00171 Helsinki Puh. 029 533 7000 kirjaamo@kansallisarkisto.fi
Riksarkivet Fredsgatan 17 PB 258, 00171 Helsingfors Tfn kansallisarkisto.fi

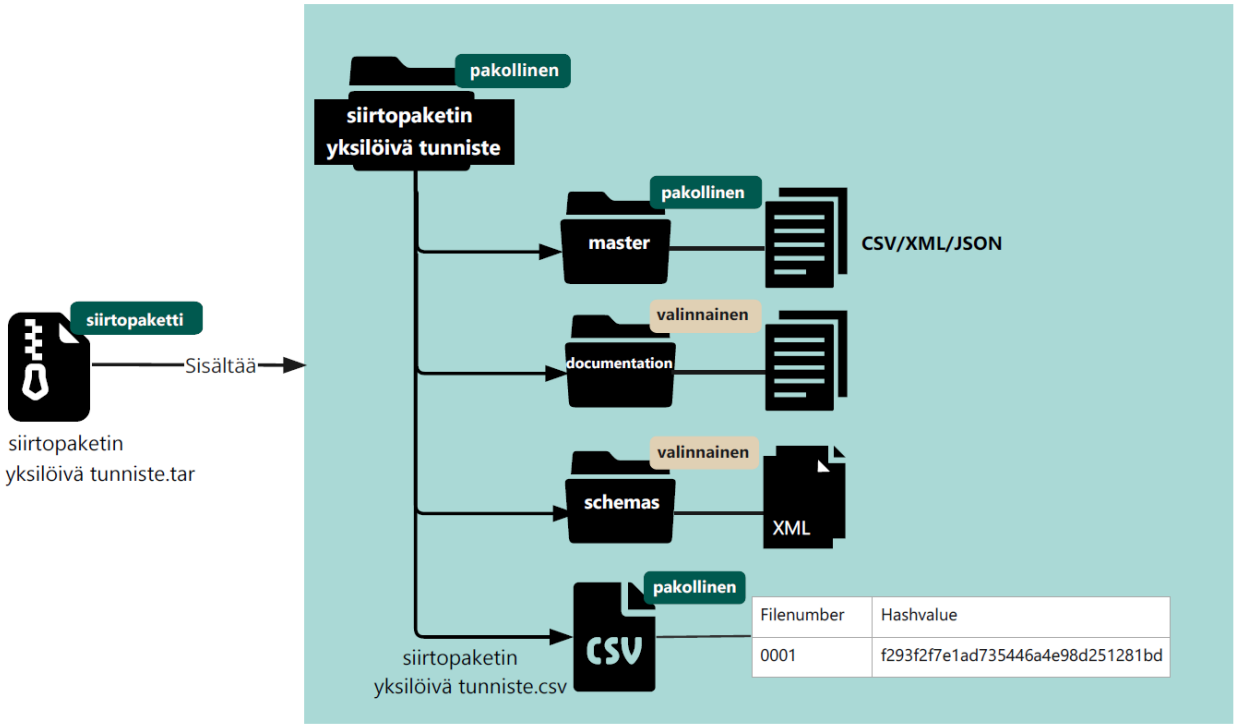
Vinkki!

Laskentataulukko-ohjelmistot muuntavat numeroita sisältävät merkkijonot usein automaattisesti luvuiksi (esim. 0001 -> 1). Solun muotoilu kannattaa muuttaa tekstiksi, jolloin merkkijono tallentuu syötetyssä muodossa.

Siirtopaketti

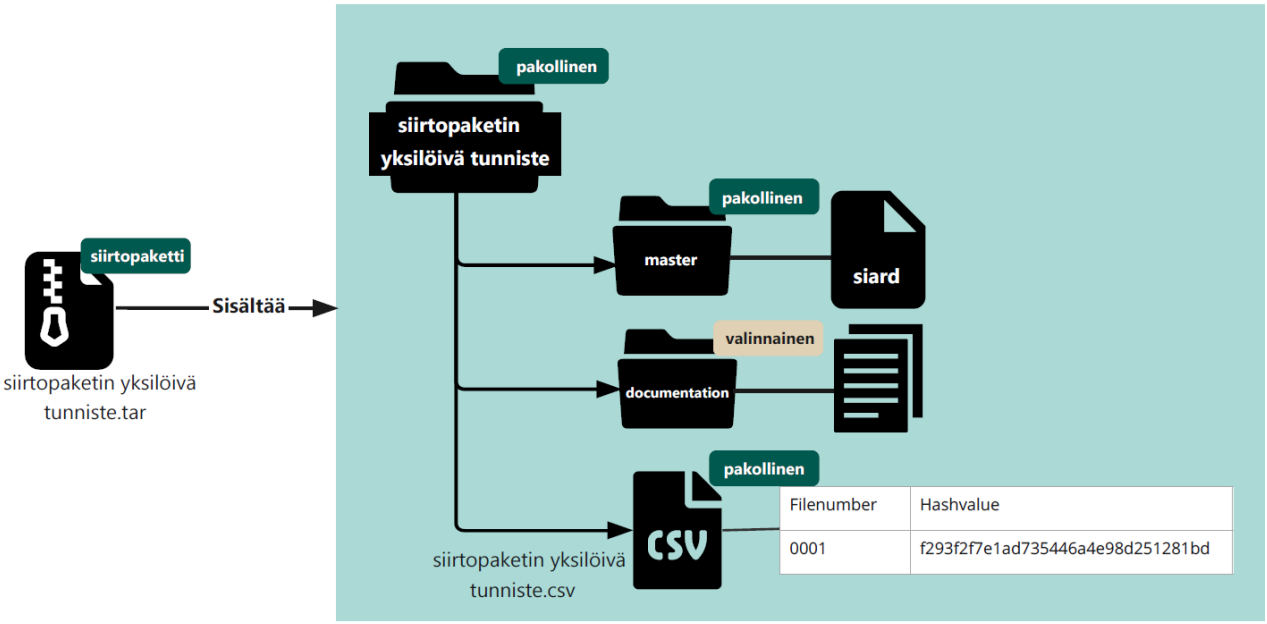
Juurihakemisto nimetään siirtopaketin tunnisteella. Hakemistot tiedostoineen paketoidaan siirtopaketiksi eli yhdeksi TAR-tiedostoksi. TAR-tiedoston saa lisäksi pakata häviöttömään GZIP- tai BZIP2-muotoon.

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



Kuva 2. Hakemistorakenne ja siirtopaketti. Tiedostot tallennetaan siirtopakettia varten kuvan mukaiseen hakemistorakenteeseen. Juurihakemisto nimetään siirtopaketin tunnisteella ja paketoidaan yhdeksi TAR-tiedostoksi.

Kansallisarkisto Rauhankatu 17 PL 258, 00171 Helsinki Puh. 029 533 7000 kirjaamo@kansallisarkisto.fi
Riksarkivet Fredsgatan 17 PB 258, 00171 Helsingfors Tfn kansallisarkisto.fi



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

Kansallisarkisto Rauhankatu 17 PL 258, 00171 Helsinki Puh. 029 533 7000 kirjaamo@kansallisarkisto.fi
Riksarkivet Fredsgatan 17 PB 258, 00171 Helsingfors Tfn kansallisarkisto.fi

Liitteet

LIITE 1 Diaariaineistot esimerkkinä rakenteisesta tietoaineistosta ja diaariaineistojen XML-rakenne

Diaariaineistoilla[2] tarkoitetaan tässä ohjeessa digitaalisessa muodossa olevia asioiden ja asiakirjojen rekisteröintitietoja. Ohjetta sovelletaan esimerkiksi silloin, kun viranomaisen vanhasta diaarisovelluksesta ollaan luopumassa.

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

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

Diaariaineistojen XML-rakenne

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

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

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

Rakenteessa asiakirjan voi liittää asialle joko toimenpiteen kautta tai ilman toimenpidettä suoraan asialle. Samalle asialle on mahdollista liittää asiakirjoja sekä asialle että toimenpiteelle. Mikäli asiakirja on liitetty suoraan asialle, merkitään se asialle toimenpiteiden tietojen jälkeen ilman toimenpidekytköstä.

```
<ASIA>
 <Toimenpiteet>
                <Asiakirjat> asiakirjat, jotka liittyvät toimenpiteeseen
                </Asiakirjat>
 </Toimenpiteet>
```

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

Kansallisarkisto Rauhankatu 17 PL 258, 00171 Helsinki Puh. 029 533 7000 kirjaamo@kansallisarkisto.fi
Riksarkivet Fredsgatan 17 PB 258, 00171 Helsingfors Tfn kansallisarkisto.fi

<Asiakirjat> asiakirjat, jotka liittyvät suoraan asialle
</Asiakirjat>
</ASIA>

Taulukoissa 1., 2. ja 3. esitetään XML-rakenne asiatasolla (taulukko 1.), toimenpidetasolla (taulukko 2.) ja asiakirjatasolla (taulukko 3.). Taulukoiden rivit kuvaavat rakenteeseen liittyvän XML-skeeman elementtejä. Pakolliseksi merkityt tiedot on sisällytettävä datakoosteeseen aina, kun ne ovat tiedossa.

Taulukko 1. Asian tiedot

Metatiedon nimi ja tunniste	**Selitys**	**Tietotyyppi**	**Pakollisuus ja toistettavuus**
Numero <nr>	Asian rekisteröinnin yhteydessä annettu numero, joka esiintyy osana diaarinumeroa	Vapaa teksti	Pakollinen
Tehtävä <function>	Tehtävän tai asiaryhmän tunniste, joka yleensä esiintyy osana diaarinumeroa	Vapaa teksti	Pakollinen
Vuosi <year>	Diaarinumerossa esiintyvä vuosiluku, joka yleensä merkitsee vuotta, jolloin asia on avattu	Vapaa teksti	Pakollinen
Diaarinumero <caseid>	Asianumero, diaarinumero tai rekisterinumero, esim. 13/43/1998 Vapaa teksti, joka yhdistää kolmen yllä kuvatun elementin (<nr>, <function> ja >year>) tiedot	Vapaa teksti	Pakollinen
Tehtäväluokka <functionclass>	Tehtäväluokitus on mahdollista esittää monitasoisena hierarkiana	Vapaa teksti	Valinnainen
Asian lähettäjä/vastaanottaja <actor>	Voi olla myös asian vireillepanija, avaaja jne. Rooli esitetään attribuutissa	Vapaa teksti	Pakollinen

Kansallisarkisto Rauhankatu 17 PL 258, 00171 Helsinki Puh. 029 533 7000 kirjaamo@kansallisarkisto.fi
Riksarkivet Fredsgatan 17 PB 258, 00171 Helsingfors Tfn kansallisarkisto.fi

Asian otsikko <title>	Asian otsikko	Vapaa teksti	Pakollinen
Asian kuvaus <description>	Asian otsikkoa tarkentava asian sisällön kuvaus. Voidaan käyttää myös vapaamuotoisen lisätiedon esittämiseen.	Vapaa teksti	Valinnainen
Päivämäärä <casedate>	Asian avaamisen tai päättymisen päivämäärä Rooli esitetään attribuuttina avaus tai päätös	Päivämäärä vvvv-kk-pv Datatyyppi määrittää	Pakollinen
Asian suunta <direction>	Kuvaa, onko asia saapuva vai lähtevä	Saapuva, Lähtevä Arvojoukko	Valinnainen
Julkisuus <publicity>	Tieto siitä, onko asia julkinen vai salassa pidettävä	Julkinen, Osittain salassa pidettävä, Salassa pidettävä Arvojoukko	Pakollinen
Henkilötietoluonne <personaldata>	Tieto siitä, sisältyykö asian tietoihin henkilötietoja vai ei	Sisältää henkilötietoja, Ei sisällä henkilötietoja Arvojoukko	~~Valinnainen~~, Pakollinen, ei toistettava
Käyttörajoituksen peruste <restriction>	Lainkohta, johon käyttörajoitus perustuu (julkisuuslaki tai erityislaki)	Vapaa teksti	Pakollinen, jos julkisuus on salassa pidettävä tai osittain salassa pidettävä Toistettava
Sisäiset lisätiedot <custom>	Muut organisaation asianhallintaan liittyvät lisätiedot, jotka eivät sovellu muihin kohtiin.	Mikä tahansa (sekä tekstiä että muuta xml:ää)	Valinnainen, ei toistettava

Kansallisarkisto Rauhankatu 17 PL 258, 00171 Helsinki Puh. 029 533 7000 kirjaamo@kansallisarkisto.fi
Riksarkivet Fredsgatan 17 PB 258, 00171 Helsingfors Tfn kansallisarkisto.fi

Taulukko 2. Toimenpiteen tiedot

Metatiedon nimi ja tunniste	**Selitys**	**Tietotyyppi**	**Pakollisuus ja toistettavuus**
Toimenpiteen otsikko <title>	Toimenpiteen otsikko.	Vapaa teksti	Valinnainen
Toimenpiteen kuvaus <description>	Toimenpiteen otsikkoa tarkentava kuvaus. Voidaan käyttää myös vapaamuotoisen lisätiedon esittämiseen.	Vapaa teksti	Valinnainen
Toimenpiteen laatija <actor>	Voi olla lähettäjä, vastaanottaja, hyväksyjä jne. Rooli esitetään attribuutissa. Rooli on vapaasti määriteltävissä.	Vapaa teksti	Valinnainen, toistettava
Toimenpiteen tyyppi <actiontype>	Toimenpiteet voivat olla esimerkiksi asian välitoimenpiteitä tai lopputoimenpiteitä.	Vapaa teksti	Valinnainen
Toimenpiteen päivämäärä <date>	Saapumispäivämäärä tai laatimispäivämäärä. Rooli esitetään attribuutissa. Rooli on vapaasti määriteltävissä.	Päivämäärä vvvv-kk-pv Datatyyppi määrittää	Valinnainen, toistettava
Toimenpiteen suunta <direction>	Tieto siitä, onko kyseessä saapuva vai lähtevä toimenpide.	Saapuva, Lähtevä Arvojoukko	Valinnainen
Sisäiset lisätiedot <custom>	Muut organisaation asianhallintaan liittyvät lisätiedot, jotka eivät sovellu muihin kohtiin.	Mikä tahansa (sekä tekstiä että muuta xml:ää)	Valinnainen, ei toistettava

Taulukko 3. Asiakirjan tiedot

Metatiedon nimi ja tunniste	**Selitys**	**Tietotyyppi**	**Pakollisuus ja toistettavuus**
Asiakirjan otsikko <title>	Asiakirjan otsikko.	Vapaa teksti	~~Valinnainen~~ Pakollinen

Kansallisarkisto Rauhankatu 17 PL 258, 00171 Helsinki Puh. 029 533 7000 kirjaamo@kansallisarkisto.fi
Riksarkivet Fredsgatan 17 PB 258, 00171 Helsingfors Tfn kansallisarkisto.fi

Asiakirjan kuvaus <description>	Asiakirjan otsikkoa tarkentava sisällön kuvaus. Voidaan käyttää myös vapaamuotoisen lisätiedon esittämiseen.	Vapaa teksti	Valinnainen
Asiakirjan lähettäjä- tai vastaanottaja <actor>	Voi olla myös asiakirjan laatija, hyväksyjä tai allekirjoittaja. Rooli esitetään attribuuttina.	Vapaa teksti	Valinnainen, toistettava
Asiakirjatyyppi <doctype>	Asiakirjatyyppi esim. päätös, hakemus, ilmoitus.	Vapaa teksti	Valinnainen
Asiakirjan päivämäärä <date>	Asiakirjan laatimisen, lähettämisen, hyväksymisen tai allekirjoittamisen päivämäärä. Rooli esitetään attribuuttina. Rooli on vapaasti määriteltävissä.	Päivämäärä vvvv-kk-pv Datatyyppi määrittää	Valinnainen, toistettava
Julkisuus <publicity>	Tieto siitä, onko asiakirja julkinen vai salassa pidettävä.	Julkinen, Osittain salassa pidettävä, Salassa pidettävä Arvojoukko	Valinnainen
Henkilötietoluonne <personaldata>	Tieto siitä, sisältyykö asiakirjan tietoihin henkilötietoja vai ei.	Sisältää henkilötietoja, Ei sisällä henkilötietoja Arvojoukko	Valinnainen, ei toistettava
Käyttörajoituksen peruste <restriction>	Lainkohta, johon käyttörajoitus perustuu (julkisuuslaki tai erityislaki).	Vapaa teksti	Pakollinen, jos asiakirja on salassa pidettävä tai osittain salassa pidettävä
Sisäiset lisätiedot <custom>	Muut organisaation asianhallintaan liittyvät lisätiedot, jotka eivät sovellu muihin kohtiin.	Mikä tahansa (sekä tekstiä että muuta xml:ää)	Valinnainen, ei toistettava

Kansallisarkisto Rauhankatu 17 PL 258, 00171 Helsinki Puh. 029 533 7000 kirjaamo@kansallisarkisto.fi
Riksarkivet Fredsgatan 17 PB 258, 00171 Helsingfors Tfn kansallisarkisto.fi

LIITE 2 SIARD 2.1 -tiedoston muodostaminen

SIARD-tiedoston muodostamiseen on saatavilla ilmaisia työkaluja. Näitä ovat Keep Solutionsin Database Preservation Toolkit (https://database-preservation.com/) sekä Sveitsin kansallisarkiston SIARDSuite (https://www.bar.admin.ch/bar/en/home/archiving/tools/siard-suite.html). Keskeinen ero työkaluissa on, että ensin mainittu sisältää myös luodun SIARD-tiedoston validointiominaisuuden.

Database Preservation toolkitillä SIARDin voi luoda seuraavista tietokannanhallintajärjestelmistä:

- MySQL/MariaDB
- PostgreSQL
- Oracle
- Microsoft SQL Server
- Microsoft Access
- Progress OpenEdge
- Sybase ASA
- muut tietokannat, jotka hyödyntävät JDBC:tä

(lähde: https://github.com/keeps/dbptk-developer)

SIARD Suite mahdollistaa SIARDin muodostamisen seuraavista tietokantatyypeistä:

- MS Access 2007 tai uudempi
- DB/2 tai uudempi
- MySQL (tai MariaDB) 5.5 tai uudempi
- Oracle 10 tai uudempi
- PostgreSQL 11 tai uudempi
- SQL Server 2012 tai uudempi

(lähde: https://www.bar.admin.ch/bar/en/home/archiving/tools/siard-suite.html)

Kansallisarkisto ei tarjoa työkalujen käyttöön käyttäjätukea eikä voi vastata yksityiskohtaisiin kysymyksiin työkalujen käyttöön liittyen. Ongelmatilanteissa Siirtäjän on otettava yhteyttä työkalun käyttäjätukeen.

SIARD-tiedoston sisältö

Muodostettaessa SIARD-tiedostoa on mahdollista valita, mitä alkuperäisen tietokannan tauluja ja taulujen kenttiä siihen sisällytetään. Valinta voi perustua joko seulontapäätökseen tai siirtoneuvottelujen yhteydessä käytäviin tarkentaviin keskusteluihin tietokannan arkistoitavasta sisällöstä. Tiedot, jotka

Kansallisarkisto Rauhankatu 17 PL 258, 00171 Helsinki Puh. 029 533 7000 kirjaamo@kansallisarkisto.fi
Riksarkivet Fredsgatan 17 PB 258, 00171 Helsingfors Tfn kansallisarkisto.fi

ovat olennaisia tutkimuksen ja toisaalta aineiston käytettävyyden, ymmärrettävyyden ja todistusvoimaisuuden kannalta, arkistoidaan. Taulut ja kentät, joita ei ole otettu käyttöön, sekä tekniset taulut ja kentät jätetään pääsääntöisesti arkistoimatta.

SIARD-työkalu tallentaa tietokannan sisällön tauluittain tekstinä XML-muodossa SIARD-tiedoston content-kansioon.

Tietosisällön ohella kaikki tietokannan relaatioiden ja tiedon rakenteen säilymiseen liittyvä tieto tulee säilyttää. Nämä metatiedot kuvaavat muun muassa käytettyä skeemaa, tietokannan tauluja ja kenttiä sekä niiden tietotyyppejä, käytettyjä avaimia, tietokannan käyttäjien rooleja sekä roolien käyttöoikeuksia.

SIARD-2.1 Format Specification[3] -määrityksessä luetellaan pakollisia ja vapaaehtoisia metatietoja, jotka sisältyvät XML-muodossa SIARD-tiedoston header-kansioon. Näihin voidaan harkinnan mukaan tuottaa tarkkojen tietojen sijaan geneeristä sisältöä, esimerkiksi tietokannan käyttäjien listaaminen (SIARD-2.1 Format Specification, vaatimus 5.1 Database level metadata) ei tietoaineiston arkistointitarkoituksessa ole välttämättä oleellista, vaan voidaan todeta käyttäjäryhmän yleistasoisemman luonnehdinnan olevan riittävä taso.

Arkistoitavaksi voidaan valita myös järjestelmään luotuja näkymiä. Näkymät ovat tietokantaan tallennettuja SQL-kyselyjä, joiden suorittamisen tuloksena syntyy uusia tauluja. Vaikka näkymien sisältämät tiedot käyvät ilmi myös muista tietokannan tauluista, rikastavat ne arkistoidun tietoaineiston sisältöä tarjoamalla siihen vastaavanlaisia näkymiä, jollaisia oli tarjolla järjestelmän alkuperäisille käyttäjille sen alkuperäisessä käyttötarkoituksessa. Kyselyitä on mahdollista tuottaa myös vastaamaan tunnistettuihin tietopalvelun tarpeisiin.

Tiedot tietokantaan määritellyistä ehdoista (vaatimus 5.12 Check constraint metadata), rutiineista/proseduureista (vaatimus 5.15 Routine level metadata) ja tietokantatoteuttimista (vaatimus 5.13 Trigger level metadata) on myös mahdollista ottaa talteen SIARD-tiedostoon. Näitä ei arkistoidun tietoaineiston kohdalla ole enää tarkoitus hyödyntää aineiston kertymisessä ja käsittelyssä, vaan ne kertovat tietokannan käyttäytymisestä eli siitä, millaisia automaattisia tarkastuksia ja toimenpiteitä tietokantaan on toteutettu.

SIARD-tiedostoon sisällytettävät kuvailutiedot

Luotaessa SIARD-tiedostoa on olennaista kuvailla arkistoitavat tietokannan taulut ja niiden kentät, sillä näiden tekniset nimet eivät tavallisesti riitä tietokannan sisällön tulkitsemiseen, kun sitä käytetään alkuperäisen käyttöympäristön ja käyttäjäkunnan ulkopuolella tutkimustarkoituksissa. Mikäli kuvauksia ei ole tuotettu tietokantaan, voidaan ne lisätä jälkikäteen SIARD-työkalussa. Myös tietokannan pääpiirteiden luonnehdinta sisältyy itse SIARD-tiedostoon (SIARD-2.1 Format Specification, vaatimus 5.1 Database level

3 https://raw.githubusercontent.com/DILCISBoard/SIARD/master/specification/2018-12-04_SIARD_Format_Version-2_1-English.pdf

Kansallisarkisto Rauhankatu 17 PL 258, 00171 Helsinki Puh. 029 533 7000 kirjaamo@kansallisarkisto.fi
Riksarkivet Fredsgatan 17 PB 258, 00171 Helsingfors Tfn kansallisarkisto.fi

metadata). Tietoaineiston hallinnolliset ja kuvailevat metatiedot tuotetaan sovitulla tasolla siirron valmistelun yhteydessä erilliselle lomakkeelle ja pakettikohtaisesti SAPA-siirtokäyttöliittymässä.

Siirtopakettiin sisältyvään oheisdokumentaatioon on syytä sisällyttää erityisesti aineiston tietorakenteita ja näiden välisiä suhteita sekä aineiston käyttötapoja alkuperäisessä käyttöyhteydessä dokumentoivia kuvauksia. Tällaisia ovat esimerkiksi ER-kaaviot, käyttöliittymäkuvaukset, tietokannan sisältöä selventävät ohjeet ja käytetyt koodistot selityksineen, mikäli ne eivät sisälly itse tietokantaan.

SIARD-tiedoston laadunvarmistus

SIARD-tiedosto validoidaan teknisesti siirtovaiheessa SAPA-käyttöliittymässä. Lisäksi monet SIARD-työkalut sisältävät validointiominaisuuden. Onnistunutkaan tekninen validointi ei kuitenkaan takaa, että tiedosto olisi muodostunut täydellisesti, eikä siitä olisi jäänyt mitään tietoja pois.

SIARD-tiedoston sisältöä voi validoida esimerkiksi seuraavilla tavoilla:

- Muodostetaan SIARD-tiedosto kahdella erillisellä kerralla ja lasketaan molemmille tiedostoille tarkistussumma. Mikäli tarkistussummat poikkeavat toisistaan, on tiedostojen sisällöissä eroja. Näin on mahdollista saada kiinni hetkellisen häiriön tai inhimillisen virheen tuottama laatupoikkeama muodostetussa tietosisällössä.
    - Huomioi, että jotkin SIARD-työkalut voivat muodostaa tiedostoon tietoja automaattisesti (esim. luontihetki). Tällöin tarkistussummat eivät tule vastaamaan toisiaan.
- Muodostetaan tietokannan tietosisällöstä SIARD-tiedosto ja ladataan se tietokannanhallintajärjestelmään. Tehdään samat haut sekä alkuperäiseen tietokantaan että SIARD:iin ja verrataan hakujen tuloksia toisiinsa. Hakutulosten tulee olla identtiset.

Kansallisarkisto Rauhankatu 17 PL 258, 00171 Helsinki Puh. 029 533 7000 kirjaamo@kansallisarkisto.fi
Riksarkivet Fredsgatan 17 PB 258, 00171 Helsingfors Tfn kansallisarkisto.fi