



SAPA-alusta valmisteluvaiheen työkaluna

31.3.2022

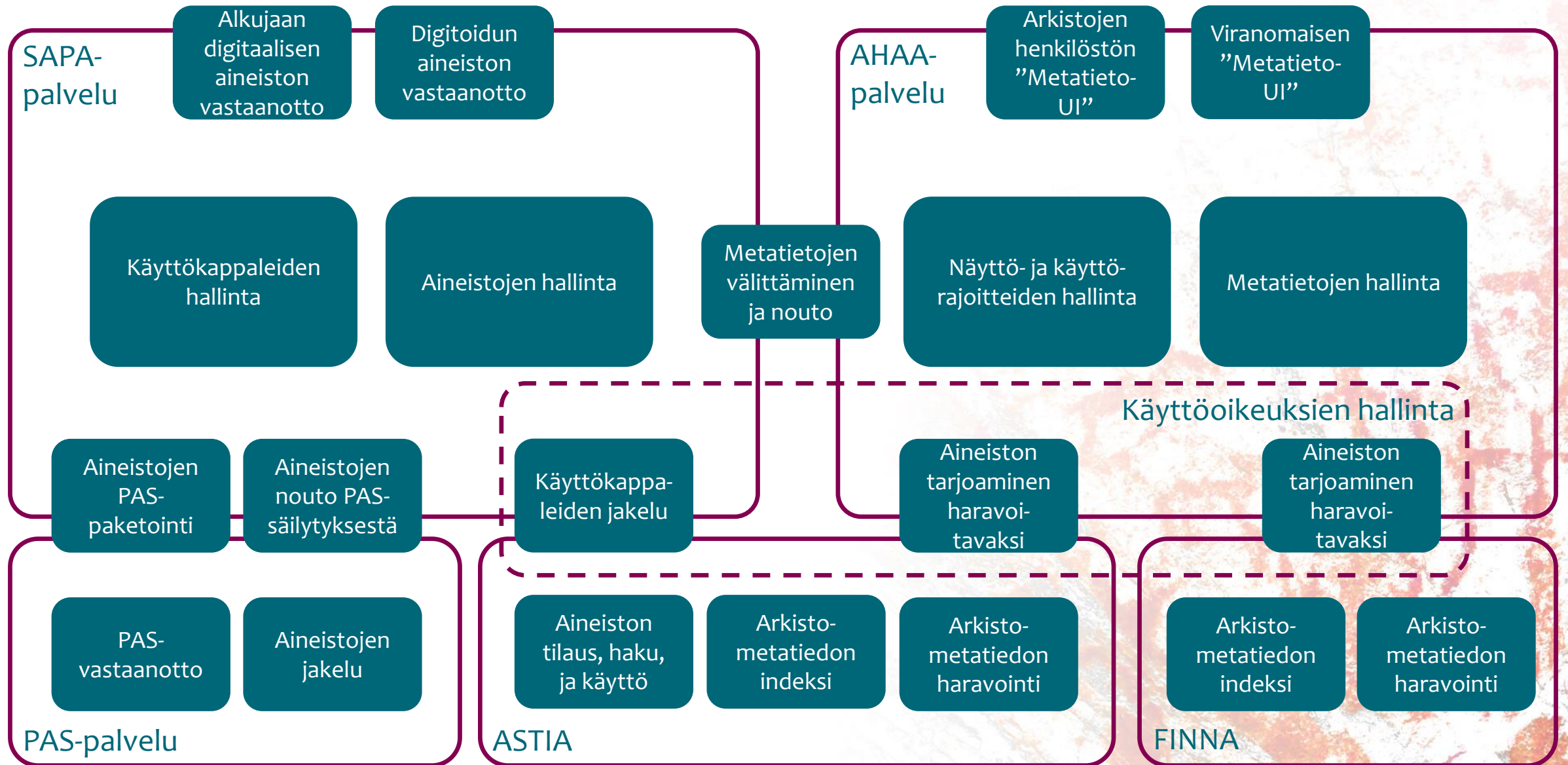
Mikko Laukkanen & Johan Kylander



SAPA-hankkeen taustaa

- Hanke perustuu hallituksen 24.9.2016 strategiaistunnossaan tekemään linjaukseen, jonka mukaan *“uudet arkistoitavat aineistot arkistoidaan jatkossa vain sähköisenä”* sekä linjaukseen pohjautuvaan valtioneuvoston 21.6.2017 antamaan periaatepäätökseen, jonka mukaan asiakirjallisten aineistojen sähköisellä arkistoinnilla ja digitoinnilla tavoitellaan niiden saatavuuden ja käytettävyyden olennaista parantumista sekä viranomaisten toiminnan tehostumista
- Valtioneuvoston 21.6.2017 periaatepäätös asiakirjallisen aineiston digitoinnista ja arkistoinnista toteaa SAPA-tietojärjestelmästä: *“Opetus- ja kulttuuriministeriö määrittää yhteistyössä valtiovarainministeriön ja Kansallisarkiston kanssa asiakirjallisen aineiston digitoinnissa ja arkistoinnissa tarvittavan kokonaisarkkitehtuurin sekä toimintamallit ja toimenpiteet tietojärjestelmien kehittämiseksi.”*
- SAPA-hanketta ohjattiin vuoden 2019 loppuun saakka OKM:n ja VM:n toimesta, mutta vuodesta 2020 eteenpäin hanketta toteutettiin yhdessä CSC:n ja Kansallisarkiston kanssa, OKM:n ja VM:n siirryttyä seurantaryhmään.
- Rahoitus hankkeelle tulee OKM:ltä vuoden 2022 loppuun saakka, jonka jälkeen Kansallisarkisto tilaa jatkotyöt suoraan CSC:ltä.

SAPA-kokonaisarkkitehtuuri

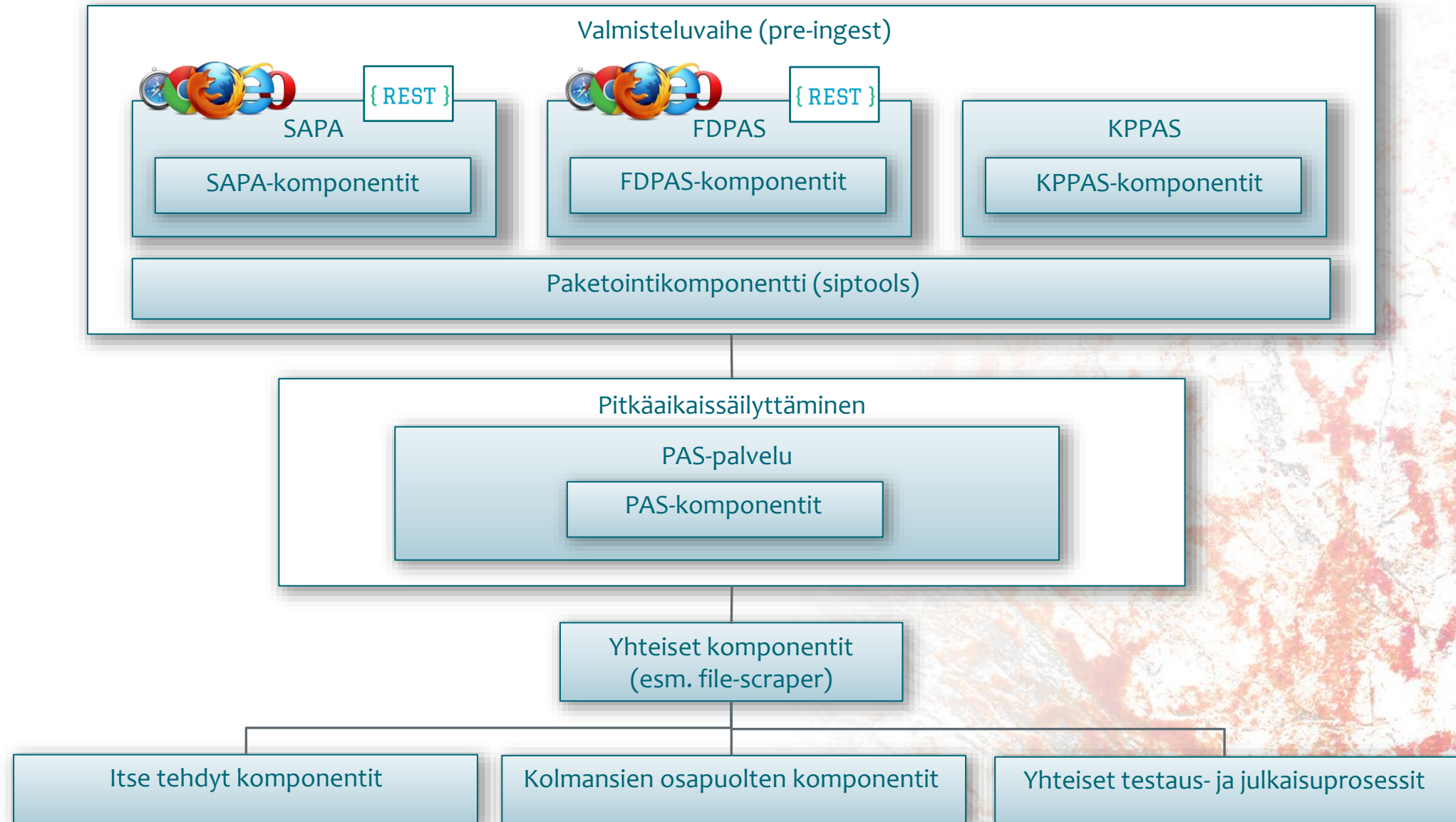


SAPA-hanke – edistyminen ja käytännön työ tähän mennessä



- Tällä hetkellä SAPA tarjoaa sekä rajapinnat että käyttöliittymät siirroille, sekä rajapinnat aineistojen käytölle, ml. luvitukset (käyttökappaleiden haku sekä aineistojen palauttaminen)
- Vuoden 2022 aikana valmistuu:
 - Kontekstimetatietojen luominen/hallinta, sekä yleisesti käyttöliittymien toimintojen lisääminen
 - Eheytestarkistukset
 - Asiakkuuksien ja sopimusten hallinta
- Kehitystyötä on tehty tiiviissä yhteistyössä Kansallisarkiston ja CSC:n kesken, ketterästi Scrum-mallilla, Kansallisarkistolta tullessa tuoteomistajuus, ja CSC:ltä kehitystyöstä vastaaminen (ScrumMaster).

SAPA osana PAS-arkkitehtuuria

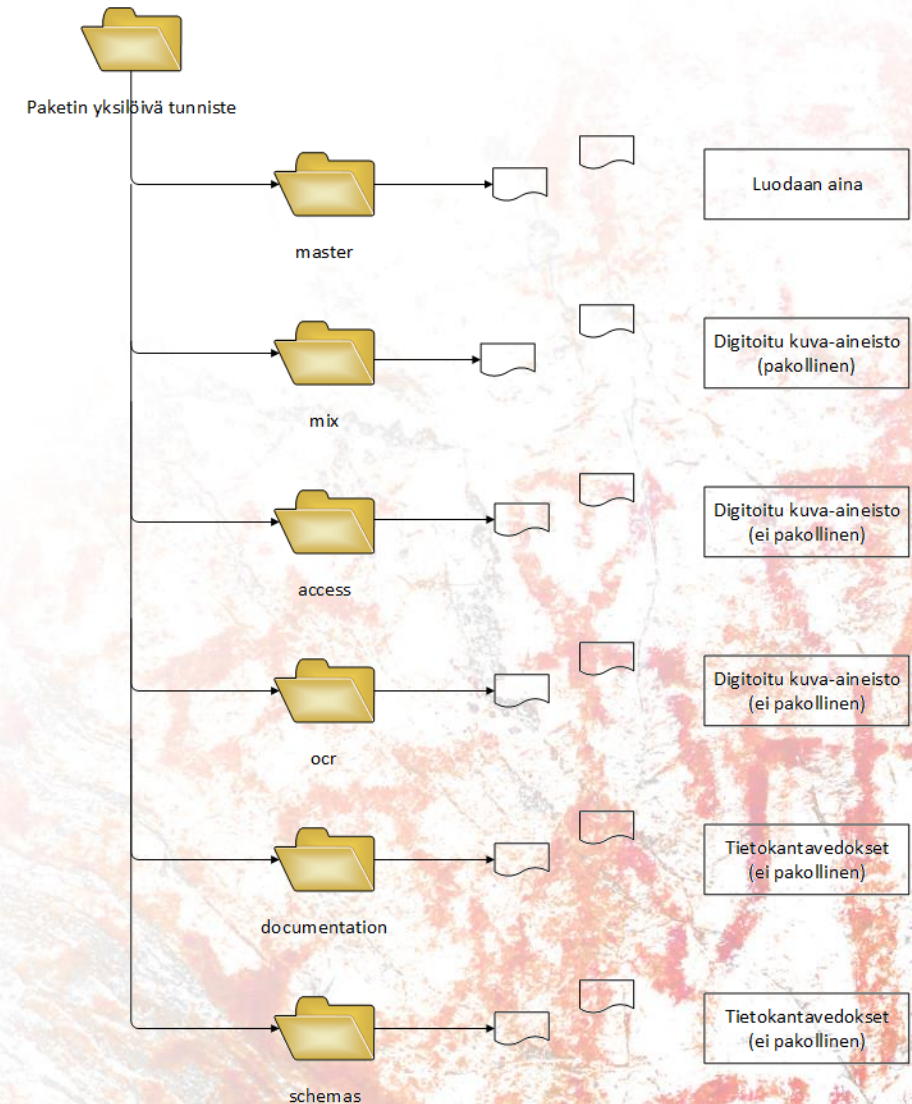


SAPA – vastaanotto ja paketointi

- SAPA ottaa vastaan siirtoja määritetyssä rakenteessa
 - Tuettuja rakenteita: digitoitu kuva-aineisto, tietokantavedokset ja sähkö-siirrot
- SAPA tuottaa siirretyistä aineistoista siirtopaketteja (SIP) ja lähettää ne PAS-palveluun säilytettäväksi
- SAPA hakee kuvailevat metatiedot sekä tiedostojen tunnisteet AHAA:sta ja liittää ne osaksi SIP:ejä
 - Tiedostojen lisäksi SAPA luo AHAA:seen digitaalisia ilmentymiä ja linkaaritapahtumia, jotka kuvaavat digitaalisen aineiston siirtoa ja pitkäaikaissäilyttämistä
- SAPA luo ja hallinnoi siirretyille aineistoille käyttökopioita
 - Käyttökopiot ovat käytettävissä Kansallisarkiston järjestelmien (kuten Astia) kautta
- Myös säilytettävän aineiston jakelu (SAPA:ssa "Aineiston palautus") on toteutettu

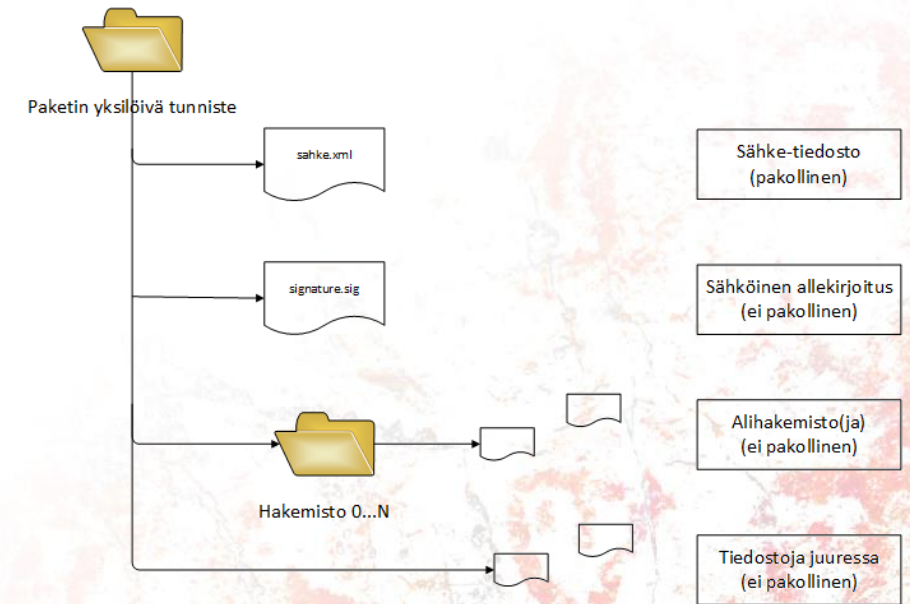
Sisääntulevien siirtojen rakenne (ei sähke)

- Paketilla yksilöivä tunniste (esim. AHAA-tunniste) juurihakemiston nimenä
- Tallekappaleet "master"-nimisessä kansiossa -> tiedostot paketoidaan SIP:iin
 - digitoitu-kuva aineisto: tiff, jpeg
 - tietokantavedos: csv, xml, json, (tulevaisuudessa myös SIARD)
- Digitoidun kuva-aineiston tekninen metatieto "mix"-nimisessä kansiossa
- Käyttökappaleet, dokumentaatio ja paikalliset XML-skeemat omiin kansioihinsa (jos niitä on)
 - dokumentaatio ja XML-skeemat mukaan SIP:iin
 - käyttökappaleet (myös ocr) säilytetään SAPA:ssa



Sisääntulevien siirtojen rakenne (sähke)

- Paketilla yksilöivä tunniste (esim. AHAA-tunniste) juurihakemiston nimenä
- sahke.xml juurihakemistossa
- Paketin rakenne muuten vapaa, koska sahke.xml sisältää kaikkien tiedostojen suhteelliset polut
- Tallekappaleet
 - sähke: teoriassa (lähes) mikä tahansa säilytys- tai siirtokelpoinen muoto



SAPA-paketointi – aineisto SIP:eiksi

- Automaattinen paketointityöväu
 - toimii aivan kuten PAS-palvelun automaattinen vastaanotto (python-luigi)
 - rakennettu paketointikomponentin (siptools) ympärille
- Vastaanotetut siirrot tarkistetaan ja validoidaan
 - siirron eheys
 - paketin rakenne
 - tiedostojen tunnistaminen ja osittainen validointi
- Teknistä metatietoa luodaan
 - hyödynnetään paketointikomponenttia
 - "lisätään" tiedostot AHAA-metatietovarantoon (AHAA:n tuottamat tiedostojen tunnisteet SIP:iin)
- Kuvailevaa metatietoa haetaan AHAA:sta EAD-muodossa
- Tapahtumahistoria, joka kuvaa aineiston paketoinnin, luodaan



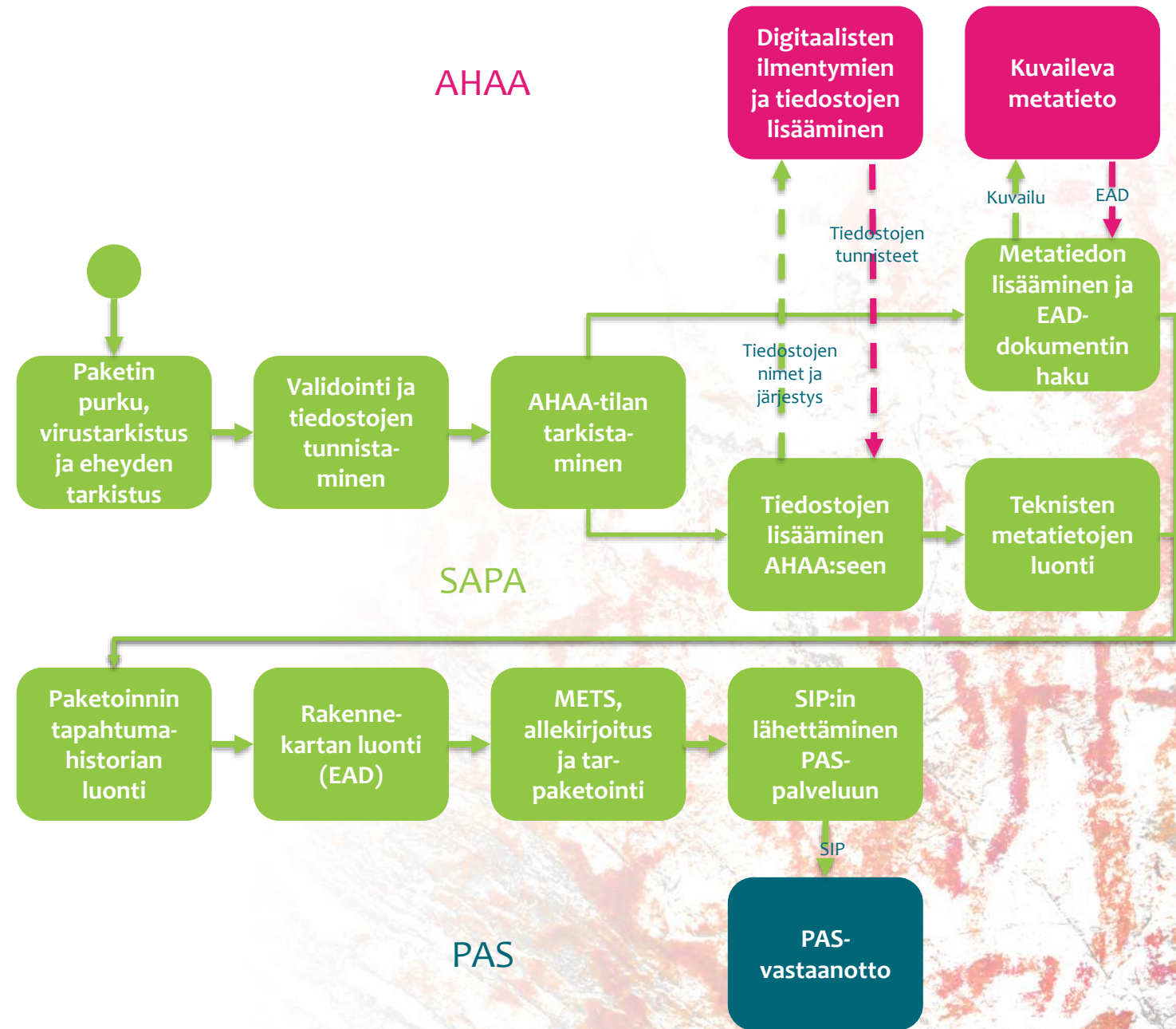
SAPA-paketointi – aineisto SIP:eiksi

- Rakennekartta **luodaan** hierarkkisen EAD-metatiedon pohjalta
- METS-dokumentti koostetaan
- SIP:in allekirjoittaminen, tar-paketointi ja lähettäminen PAS-palvelun vastaanottoon
- PAS-vastaanotto validoi SIP:in ja sen sisällön ja ottaa aineiston säilytykseen
- SAPA pollaa PAS-vastaanoton raporttia
- Sillä aikaa SAPA luo käyttökopiotiedostoja, vie niiden tiedot AHAA:seen ja tallentaa ne käytettäväksi jaetulle alustalle



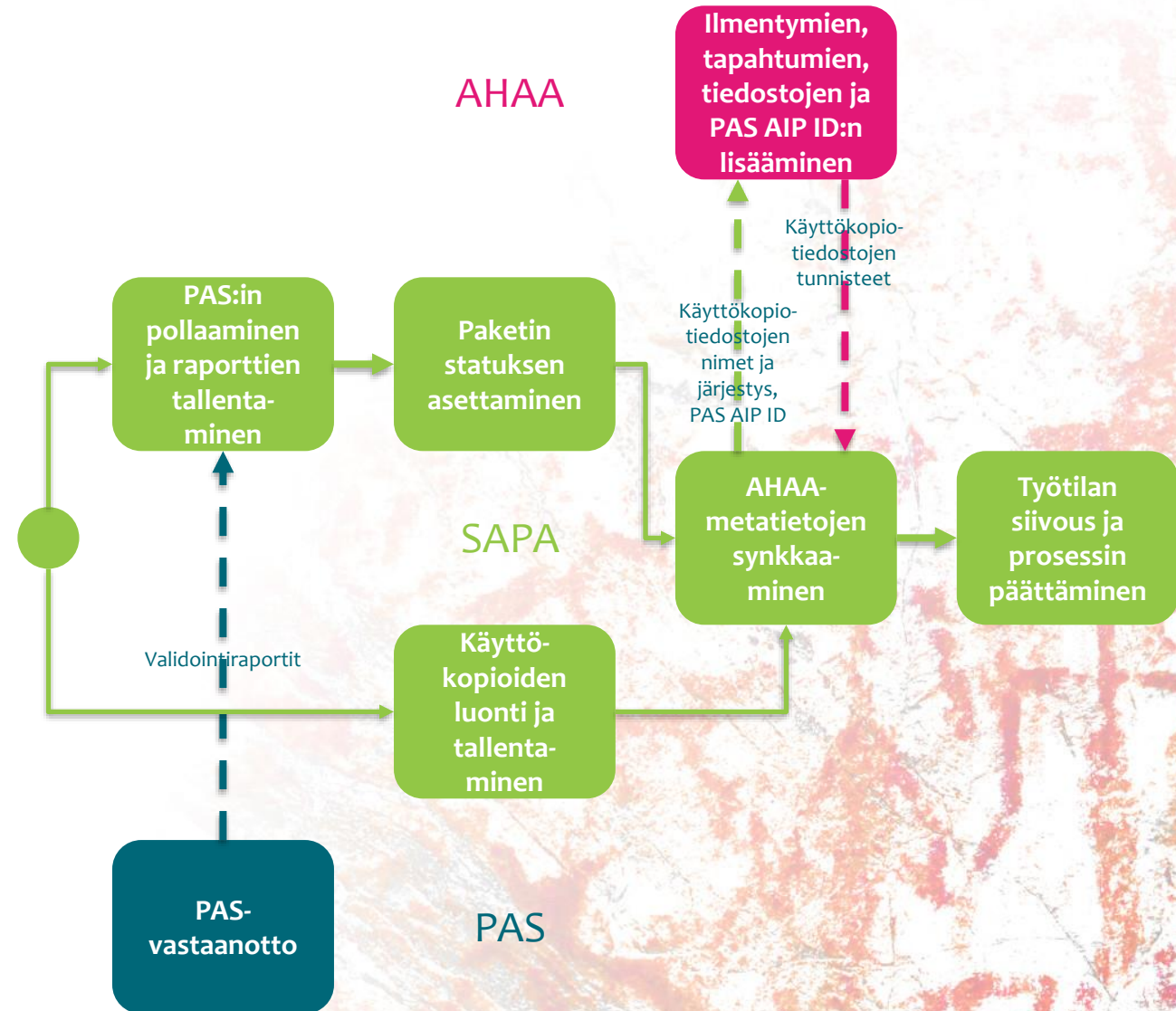
SAPA-paketointi

- SAPA:aan lähettyt siirrot tallentuvat työtilaan, josta ne otetaan käsittelyyn automaattisesti
 - siirtojen rinnakkainen käsittely
 - työvuoto on modulaarinen
- Paketoinnin työvuoto tarkistaa sisään tulevan aineiston, prosessoi sen ja lähettää aineiston PAS-vastaanottoon
- Työvuoto juttelee AHAA:n ja PAS-palvelun kanssa
- Työvuonon vaiheet tallentuvat tietokantaan, josta niiden statukset ovat haettavissa



SAPA:n viimeistely

- Työvuon toisessa osassa pollataan PAS-vastaanottoa ja sen raporttia
- Rinnakkain luodaan käyttökopioita ja tallennetaan tilastotietoja
- PAS-vastaanoton raportti tulkitaan, ja jos aineisto on hyväksytty säilytykseen vietään AHAA:seen pyyntöjä lisätä tapahtumia liittyen digitaalisen aineiston vastaanottoon ja säilytykseen
- Hylätyt paketit ilmoitetaan siirtäjälle virhekoodilla ja metatiedot siivotaan AHAA:sta



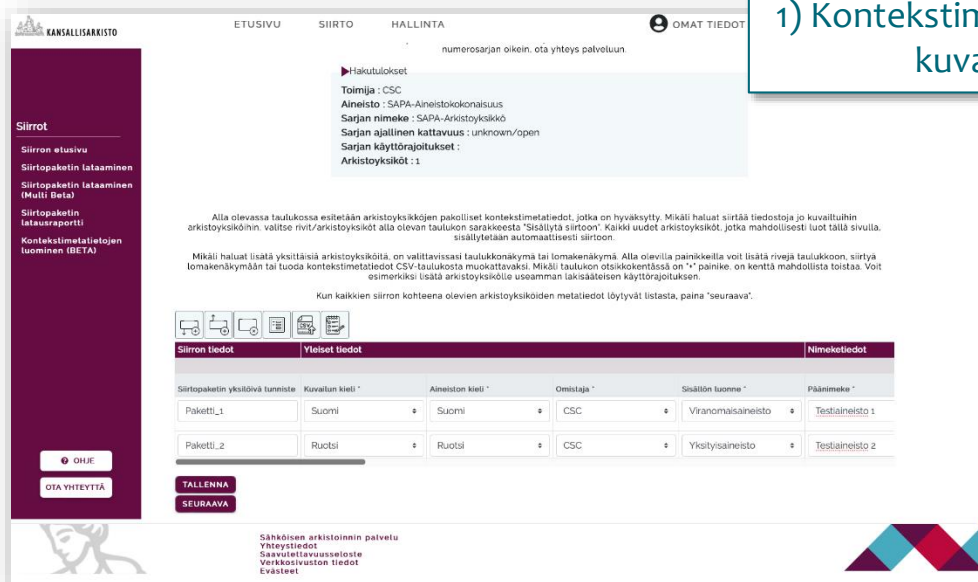
SAPA:n rajapinnat

- SAPA hyödyntää HTTP REST rajapintoja aineiston lähettämässä ja siirtojen statuksien kyselyssä
 - siirtojen lähetyksessä hyödynnetään TUS-protokollaa
- Paketointi pilkottu vaiheisiin, jotka päivittävät statuksen
- Myös aineiston jakelu ja käyttökopioiden haku tapahtuu REST-rajapintojen kautta
- SAPA:n rajapinnat ovat samankaltaisia kuin PAS-palvelun rajapinnat
 - logiikka sama
 - vastausviestien rakenne samankaltainen

```
{
  "data": {
    "id": "1_csv-4374b19a-9fcc-48ca-b640-b56162496a3e",
    "statuses": {
      "finalizing": {
        "completed_at": "2022-03-30T11:44:31.168000+00:00",
        "initialized_at": "2022-03-30T11:44:10.502000+00:00",
        "message": "Package processing finalized",
        "outcome_successful": true
      },
      "packaging": {
        "completed_at": "2022-03-30T11:44:10.331000+00:00",
        "initialized_at": "2022-03-30T11:32:37.014000+00:00",
        "message": "Transfer contents processing and packaging completed.",
        "outcome_successful": true
      },
      "upload": {
        "completed_at": "2022-03-30T11:27:42.215000+00:00",
        "message": "Transfer uploaded to service.",
        "outcome_successful": true
      },
      "validation": {
        "completed_at": "2022-03-30T11:32:36.806000+00:00",
        "initialized_at": "2022-03-30T11:32:14.494000+00:00",
        "message": "Transfer contents structure validation completed.",
        "outcome_successful": true
      },
      "verification": {
        "completed_at": "2022-03-30T11:32:14.334000+00:00",
        "initialized_at": "2022-03-30T11:28:05.539000+00:00",
        "message": "Transfer archive file decompression and verification completed.",
        "outcome_successful": true
      }
    }
  },
  "status": "success"
}
```

Aineistojen kuvailu ja siirto SAPA-käyttöliittymän kautta

1) Kontekstimetatietojen kuvailu



KANSALLISARKISTO ETUSIVU SIIRTO HALLINTA OMAT TIEDOT

numerosarjan oikein. etä yhteys palveluun.

Hakutulokset

Toimija : CSC
Aineisto : SAPA-Aineistokokonaisuus
Sarjan nimeke : SAPA-Arkistoyksikkö
Sarjan ajallinen kattavuus : unknown/open
Sarjan käyttörajoitukset :
Arkistoyksiköt : 1

Alla olevassa taulukossa esitellään arkistoyksiköiden pakolliset kontekstimetatiedot, jotka on hyväksytty. Mikäli haluat siirtää tiedostoja jo kuvailtuihin arkistoyksikköihin, valitse rivit/arkistoyksiköt alla olevan taulukon sarakkeesta "Siällytä siirtoon". Kaikki uudet arkistoyksiköt, jotka mahdollisesti luot tällä sivulla, sisällytetään automaattisesti siirtoon.

Mikäli haluat lisätä yksittäisiä arkistoyksiköitä, on valittavissa taulukonäkymä tai lomakenäkymä. Alla olevilla painikkeilla voit lisätä rivejä taulukkoon, siirtää lomakenäkymään tai tuoda kontekstimetatiedot CSV-taulukosta muokattavaksi. Mikäli taulukon etsikkokentässä on "*" painike, on kenttä mahdollista toistaa. Voit esimerkiksi lisätä arkistoyksiköille useamman lakisääteisen käyttörajoituksen.

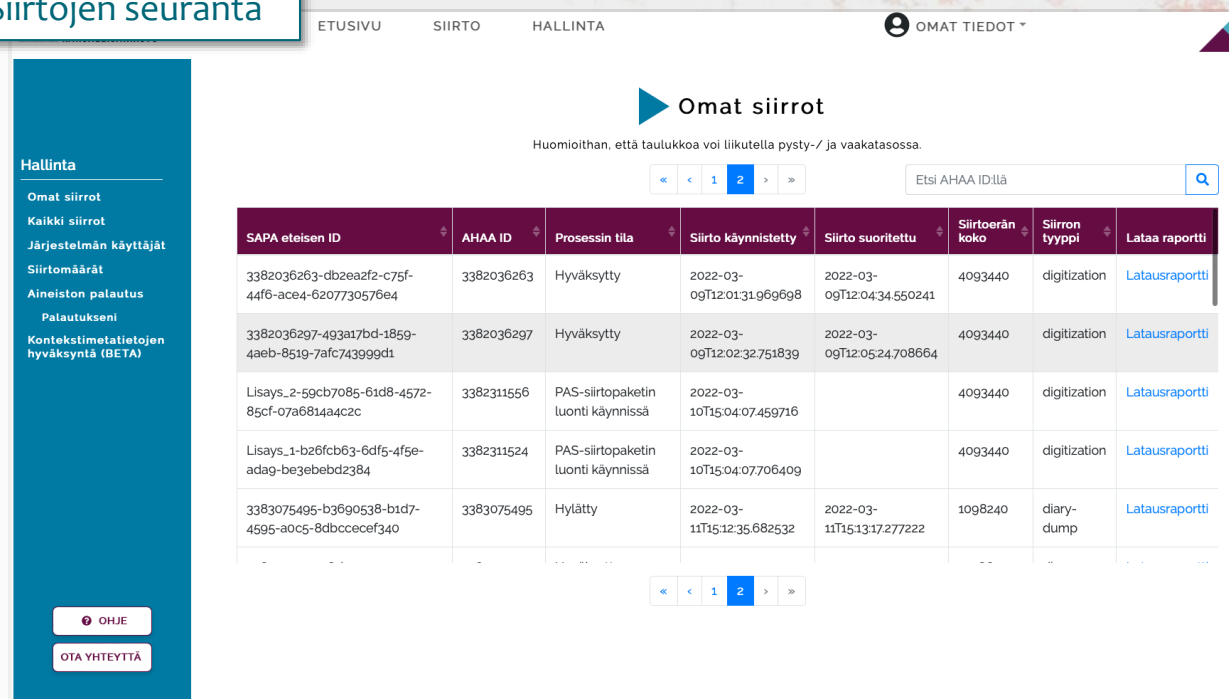
Kun kaikkien siirron kohteena olevien arkistoyksiköiden metatiedot löytyvät listasta, paina "seuraava".

Siirron tiedot	Yhteiset tiedot					Nimekettiedot
Siirtopaketin yksilöivä tunnistus	Kuvailun kieli *	Aineiston kieli *	Omistaja *	Sisällön luonne *	Päänimeke *	
Paketti_1	Suomi	Suomi	CSC	Viranomaisaineisto	Testiaineisto 1	
Paketti_2	Ruotsi	Ruotsi	CSC	Yksitysisaineisto	Testiaineisto 2	

TALLENNNA SEURAAVA

Sähköisen arkistoinnin palvelu
Yhteystiedot
Saavutettavuusseloste
Verkkosivuston tiedot
Evästeet

3) Siirtojen seuranta



KANSALLISARKISTO ETUSIVU SIIRTO HALLINTA OMAT TIEDOT

Omat siirrot

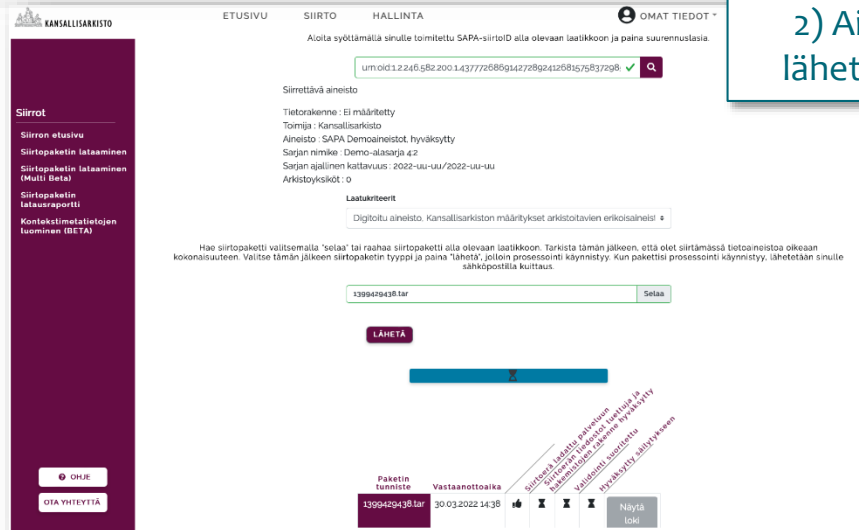
Huomioithan, että taulukko voi liikutella pysty-/ ja vaakatasossa.

Etsi AHAA ID:llä

SAPA eteisen ID	AHAA ID	Prosessin tila	Siirto käynnistetty	Siirto suoritettu	Siirtoerän koko	Siirron tyyppi	Lataa raportti
3382036263-db2eazf2-c75f-44f6-ace4-6207730576e4	3382036263	Hyväksytty	2022-03-09T12:01:31.969698	2022-03-09T12:04:34.550241	4093440	digitization	Latausraportti
3382036297-493a17bd-1859-4aeb-8519-7afc743999d1	3382036297	Hyväksytty	2022-03-09T12:02:32.751839	2022-03-09T12:05:24.708664	4093440	digitization	Latausraportti
Lisays_2-59cb7085-61d8-4572-85cf-07a6814a42c	3382311556	PAS-siirtopaketin luonti käynnissä	2022-03-10T15:04:07.459716		4093440	digitization	Latausraportti
Lisays_1-b26fcb63-6df5-4f5e-ada9-be3ebeb2384	3382311524	PAS-siirtopaketin luonti käynnissä	2022-03-10T15:04:07.706409		4093440	digitization	Latausraportti
3383075495-b3690538-b1d7-4595-a0c5-8dbcecef340	3383075495	Hylätty	2022-03-11T15:12:35.682532	2022-03-11T15:13:17.277222	1098240	diary-dump	Latausraportti

OHJE OTA YHTEYTTÄ

2) Aineiston lähettäminen



KANSALLISARKISTO ETUSIVU SIIRTO HALLINTA OMAT TIEDOT

Aiota syöttämällä sivulle toimitettu SAPA-siirtoID alla olevaan laatikkoon ja paina suurenustulosta.

Urmoid: 1.2.2.4.6.5.82.200.1.4377268091427289241268157637298

Siirrettävä aineisto

Tietoturvanne: Ei määritetty
Toimija: Kansallisarkisto
Aineisto: SAPA Demoaineistot, hyväksytty
Sarjan nimeke: Demo-alasarja 42
Sarjan ajallinen kattavuus: 2022-uu-uu/2022-uu-uu
Arkistoyksiköt: 0

Laatuksittien

Digitoitu aineisto, Kansallisarkiston määritykset arkistoitavien erikoisaineistoi *

Hae siirtopaketti valitsemalla "selaa" tai raahaa siirtopaketti alla olevaan laatikkoon. Tarkista tämän jälkeen, että olet siirtämässä tietoa oikeaan kokonaisuuteen. Valitse tämän jälkeen siirtopaketin tyyppi ja paina "lähetä", jolloin prosessiini käynnistyy. Kun paketti on prosessiini käynnistynyt, lähetetään sinulle sähköpostilla kuittaus.

1399429438 tar Selaa

LÄHETÄ

Paketti tunnistus: 1399429438 tar Vastaanottoaika: 30.03.2022 14:38

Siirretty lähetettävä pakettiin
Siirretty lähetettävään kokonaisuuteen
Valittu siirtopaketin tyyppi
Valittu siirtopaketin ajallinen kattavuus

Näytä loki

OHJE OTA YHTEYTTÄ



Kiitos!

<http://digitalpreservation.fi/>

Twitter @dpres_fi

<https://github.com/Digital-Preservation-Finland/>