



# PAS-rajapinnat työpaja 9.11.2021



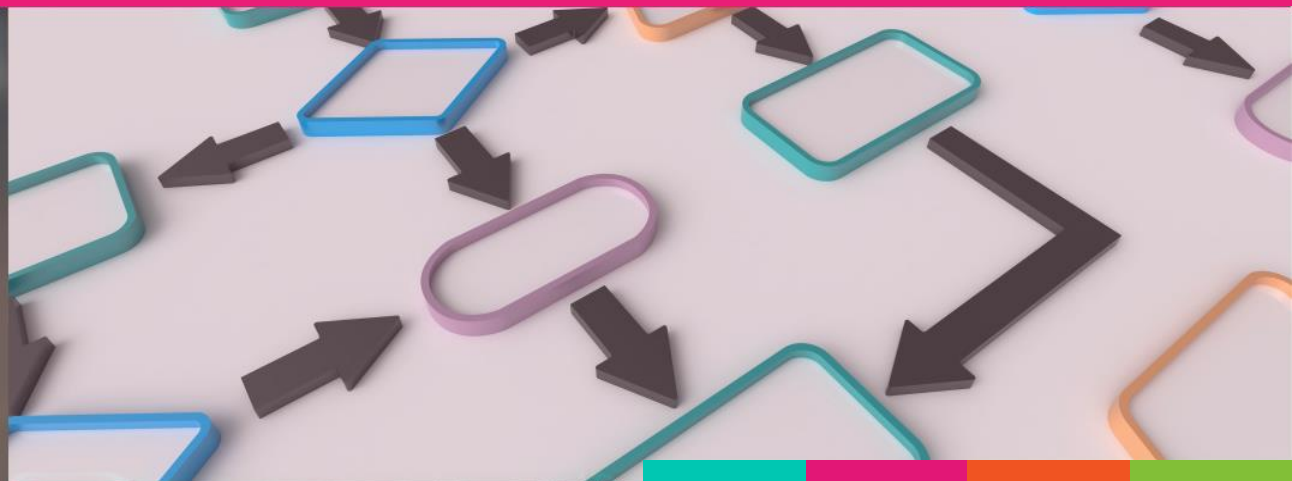
# Työpajan tavoitteet

- Päästä (alustavaan) yhteisymmärrykseen PAS-palvelun rajapinnan uusista ominaisuuksista konkreettisella tasolla
  - Aineiston siirtäminen PAS-palveluun
  - Jakelupaketin muodostaminen muusta kokonaisuudesta kuin yhdestä kokonaisesta säilytyspaketistä
  - Tilastotiedot
- Ja keskustella aikataulusta

## PARENTAL ADVISORY

Tänään esitetyt uusien toiminnallisuuden toteutustavat ovat alustavia esimerkkejä, joiden tarkoitus on hahmotella tulevaa rajapintaa. Nämä eivät tässä vaiheessa välttämättä täysin vastaa toteutuvaa rajapintaa.

# Nykytila



# Nykytila: Aineiston siirto PAS-palveluun

- Aineiston siirto PAS-palveluun tapahtuu tällä hetkellä SFTP-rajapinnan kautta
- Raportti aineiston vastaanotosta ilmestyy hakemistoihin accepted (hyväksytty) tai rejected (hylätty)
- Eri raporttien yhdistäminen eri siirtoihin SFTP-rajapintaa hyödyntäen on kankeaa: Käytännössä hyödyntävä organisaatio joutuu tekemään tähän automaation
- REST-rajapinnassa on toiminto, jolla voi noutaa raportin siirtopaketin tunnisteiden avulla
  - Raportteja on siirtopaketille useita, jos sama siirtopaketti on siirretty useasti
  - Raportin nouto vain SFTP:llä, jos siirtopaketin tunnistetta ei löytynyt paketista

# Nykytila: Aineiston nouto PAS-palvelusta

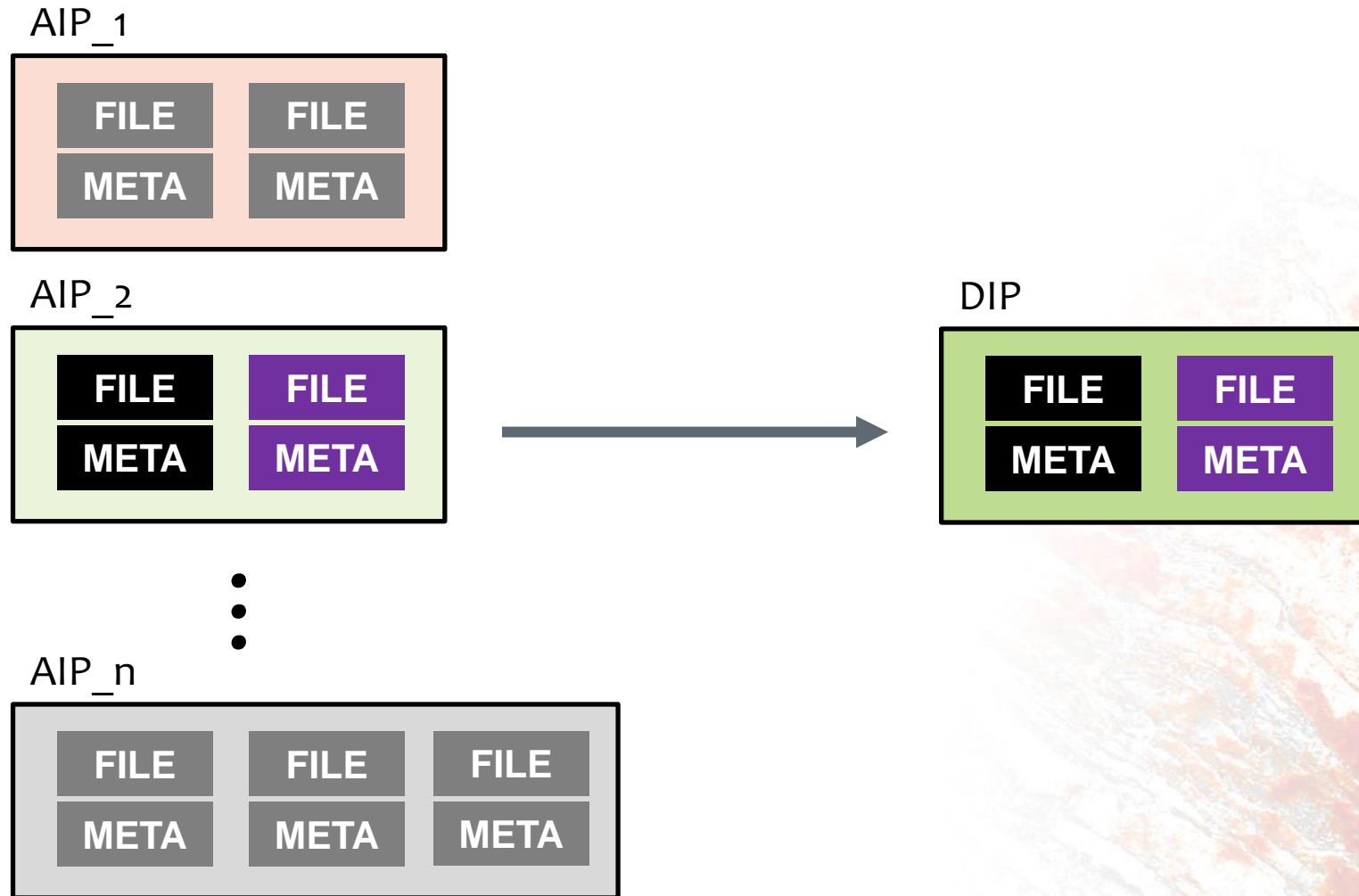
- Hyödyntävä organisaatio voi etsiä aineistoa haluamallaan hakukriteereillä
  - Useiden hakukriteerien yhdistely on mahdollista
  - METS-dokumentit on kokonaan indeksoitu hakutietokantaan
  - “Jos tieto on olemassa METS-dokumentin sisällä, sitä voi käyttää hakukriteerinä”
- Hakukriteerien tuloksista voi automatisoidusti poimia haluamansa paketit (tai vaikka kaikki paketit) ja noutaa ne
- Valmis rajapintatyökalu Githubissa:

```
access-client search --query <hakuehdot>
```

TOISTA kaikille hakutuloksien <AIP-tunniste>-alkioille:

```
access-client download <AIP-tunniste>
```

# Nykytila: Aineiston nouto PAS-palvelusta



# Nykytila: Aineiston nouto PAS-palvelusta ilman työkalua (1/3)

- Aineiston etsintä:

GET <base>/<contract>/search?<parameters>

- Vastauksena hakua vastaavien säilytyspakettien resurssit, kukin muodossa: <base>/<contract>/preserved/<aip-id>

- Säilytyspaketin muodostus jakelupaketiksi:

POST <base>/<contract>/preserved/<aip-id>/disseminate

- Kertauksena...

- <base> on URL-merkkijono, osoite, joka on ohjaa kyselyt PAS-palveluiden rajapinnalle:
  - Nykyisin: <https://pas.csc.fi/api/2.0>
- <contract> on aineiston sopimustunnus

Yksi alkio osumista:

```
{  
  "location": "<base>/<contract>/preserved/<aip-id>",  
  "createdate": "<create-date>",  
  "lastmoddate": "<last-mod-date>",  
  "match": "<search-match>",  
  "id": "<aip-id>",  
  "pkg_type": "AIP"  
}
```

## Nykytila: Aineiston nouto PAS-palvelusta ilman työkalua (2/3)

- Jakelupaketin muodostuspyynnön vastauksena hyväksyntä ja osoite jakelupaketin resurssiin:

```
HTTP/1.1 202 Accepted
Location: <base>/<contract>/disseminated/<dip-id>
...
{
  "status": "success",
  "data": {
    "disseminated": "<base>/<contract>/disseminated/<dip-id>"
  }
}
```



# Nykytila: Aineiston nouto PAS-palvelusta ilman työkalua (3/3)

- Valmistumisen seuranta ja nouto:  
GET <base>/<contract>/disseminated/<dip-id>

- Vastaus:

```
HTTP/1.1 200 OK
```

```
...
```

```
{
```

```
  "status": "success",
```

```
  "data": {
```

```
    "complete": "true",
```

```
    "actions": {
```

```
      "download": "<base>/<contract>/disseminated/<dip-id>/download",
```

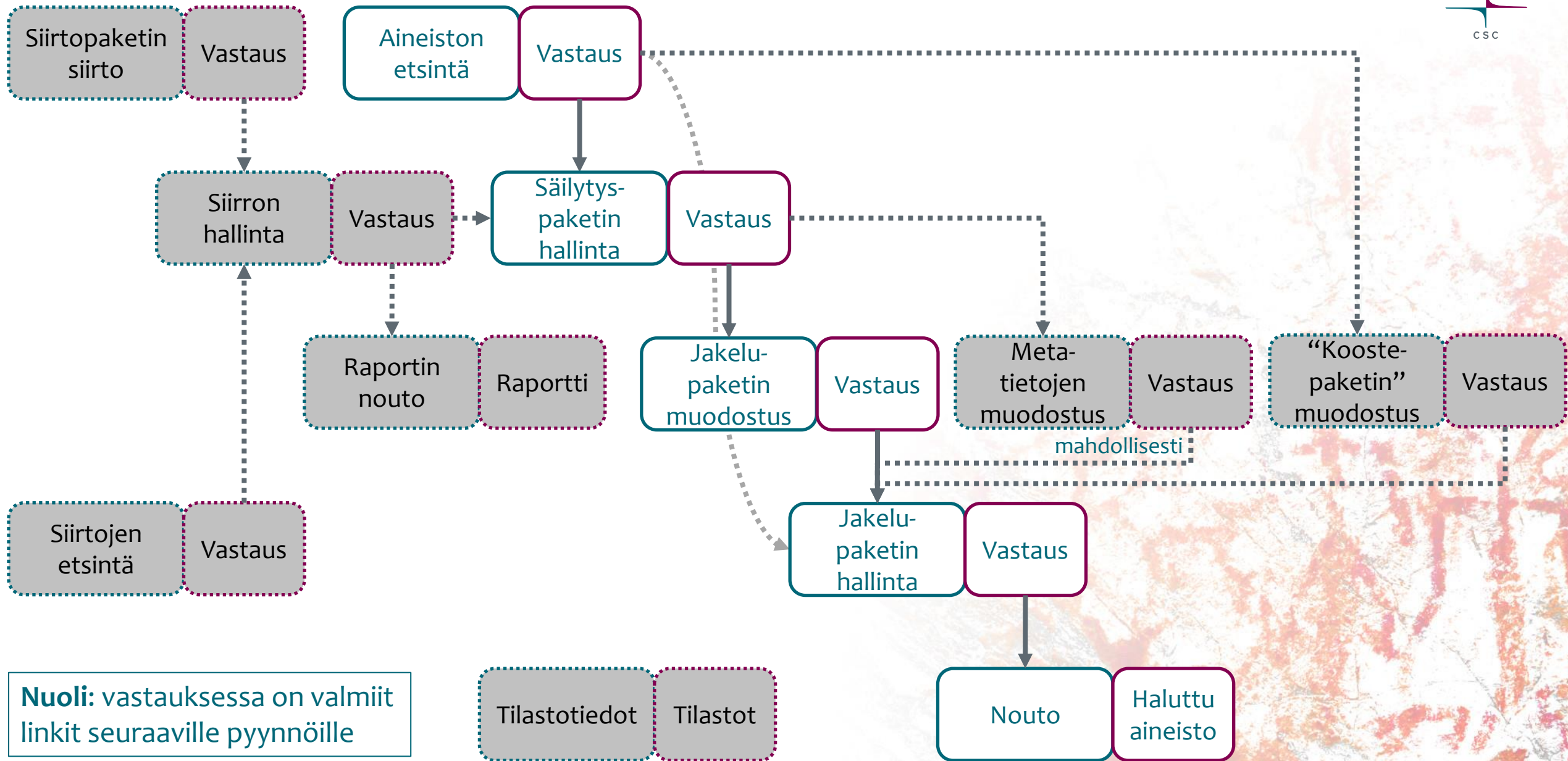
```
      "metadata": "<base>/<contract>/disseminated/<dip-id>/metadata",
```

```
      "history": "<base>/<contract>/disseminated/<dip-id>/history"
```

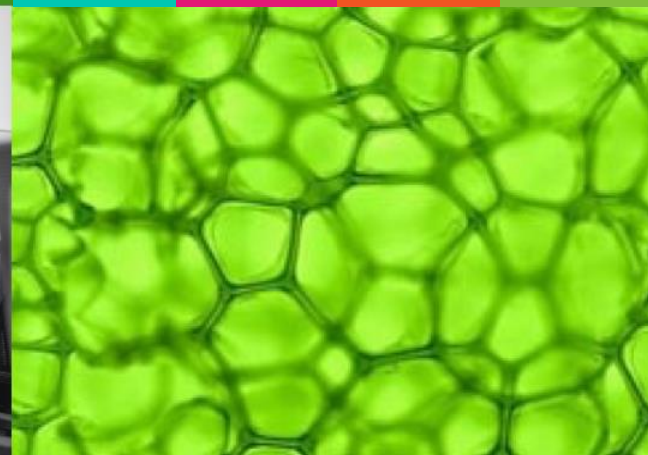
```
    }
```

```
  }
```

```
}
```



# Tavoitetila: Aineiston siirtäminen PAS-palveluun



# Aineiston siirtäminen PAS-palveluun (REST)

- Mahdollistaa helpommin aidon automaattisen kommunikoinnin järjestelmien välillä
  - Jokainen siirto on yksilöllisesti seurattavissa omalla siirtotunnisteellaan
  - Valmiit työkalut kommunikointiin olemassa
- Vastaanoton tarkastusraporttia ei tarvitse etsiä kotihakemistoista, vaan sen saa REST-keskustelun tuloksena tarvittaessa
- Helpompi ja kevyt käyttää, mahdollistaa ketterämmin eri tyyppisten ratkaisujen toteuttamisen
- HTTPS/REST-automatiikka yksinkertaistaa arkkitehtuuria verrattuna SFTP:hen
- Hylättyä aineistoa ei tarjota REST-rajapinnasta
  - Hylätyllä aineistolla ei ole säilytysvastuuta, jolloin aineisto on joka tapauksessa oltava hyödyntävällä organisaatiolla hyväksymiseen asti

# Aineiston siirtäminen PAS-palveluun (REST)

- Aineisto lähetetään REST-rajapintaan POST-tyyppisellä viestillä
- Siirtopaketti (zip/tar) liitetään HTTP-pyyntöön runkoon binäärimuodossa sellaisenaan
  - Content-type: application/zip tai application/x-tar
  - Ei Base64-koodausta
  - Ei multipart sisältöjä
- Siirron palautusviestissä ilmoitetaan siirrolle luotu (PAS-palvelun generoima) tunniste
  - Erityisesti on huomattava, että aineistoa ei vielä tässä vaiheessa ole hyväksytty säilytykseen
- Tunnisteen avulla seurataan siirron valmistumista
- Tunnisteen avulla noudetaan siirtopaketin tarkastusraportti sitten kun aineisto on hyväksytty säilytykseen tai hylätty

# Esimerkki: Aineiston siirto PAS-palveluun

- Siirtoesimerkki:
  - POST `<base>/<contract>/ingest/upload`
  - Viestin sisältönä siirtopaketti
- Onnistuneen siirron palautusviestissä ilmoitetaan siirron resurssi:

HTTP/1.1 202 Accepted

Location: `<base>/<contract>/ingested/transfers/<transfer-id>`

...

{

"status": "success",

"data": {

"transfer": "`<base>/<contract>/ingested/transfers/<transfer-id>` ",

...

}

}

# Esimerkki: Aineiston siirto PAS-palveluun

- Siirron resurssin avulla voidaan seurata siirron tilaa sekä lopulta noutaa raportti:

GET <base>/<contract>/ingested/transfers/<transfer-id>

- Esimerkkivastaus:

```
{
  "status": "success",
  "data": {
    "complete": "true",
    "result": "accepted",
    ...
    "actions": {
      "report_html": "<base>/<contract>/ingested/transfers/<transfer-id>/html",
      "report": "<base>/<contract>/ingested/transfers/<transfer-id>/report",
      "preserved": ["<base>/<contract>/preserved/<aip-id>", ...]
    }
  }
}
```

# Siirto rajapintatyökalun avulla (REST)

- Rajapintatyökaluun voidaan lisätä toiminnallisuus, joka automatisoi koko siirtoprosessin
- Toiminto voisi toimia esimerkiksi seuraavasti:
  - Hyödyntävän organisaation järjestelmä (esim. kokoelmanhallintajärjestelmä, tai mikä tahansa muu ohjelmisto) antaa rajapintatyökalulle siirtopaketin tiedostonimen polkuineen
  - Tämän seurauksena työkalu automaattisesti vie annetun siirtopaketin PAS-palveluun, seuraa siirron valmistumisen loppuun ja palauttaa tarkastusraportin hyödyntävän organisaation järjestelmään
  - Työkalu voi prosessin päättyessä ilmoittaa, onko aineisto nyt säilytyksessä vai ei
  - Esimerkiksi:  
`rest-client upload siirtopaketti.zip`

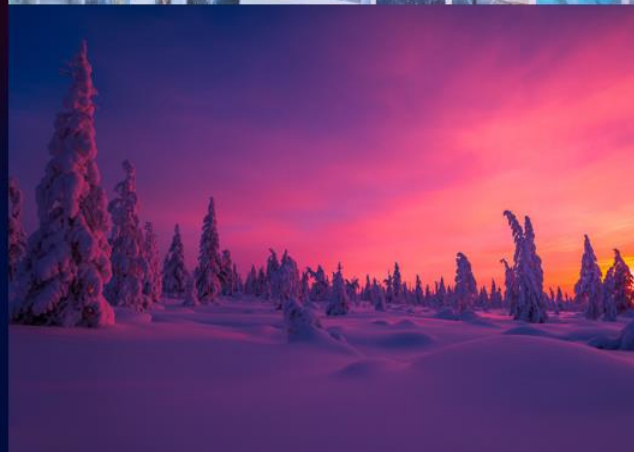


# Muutettava(?) toiminto: Raportin nouto (REST)

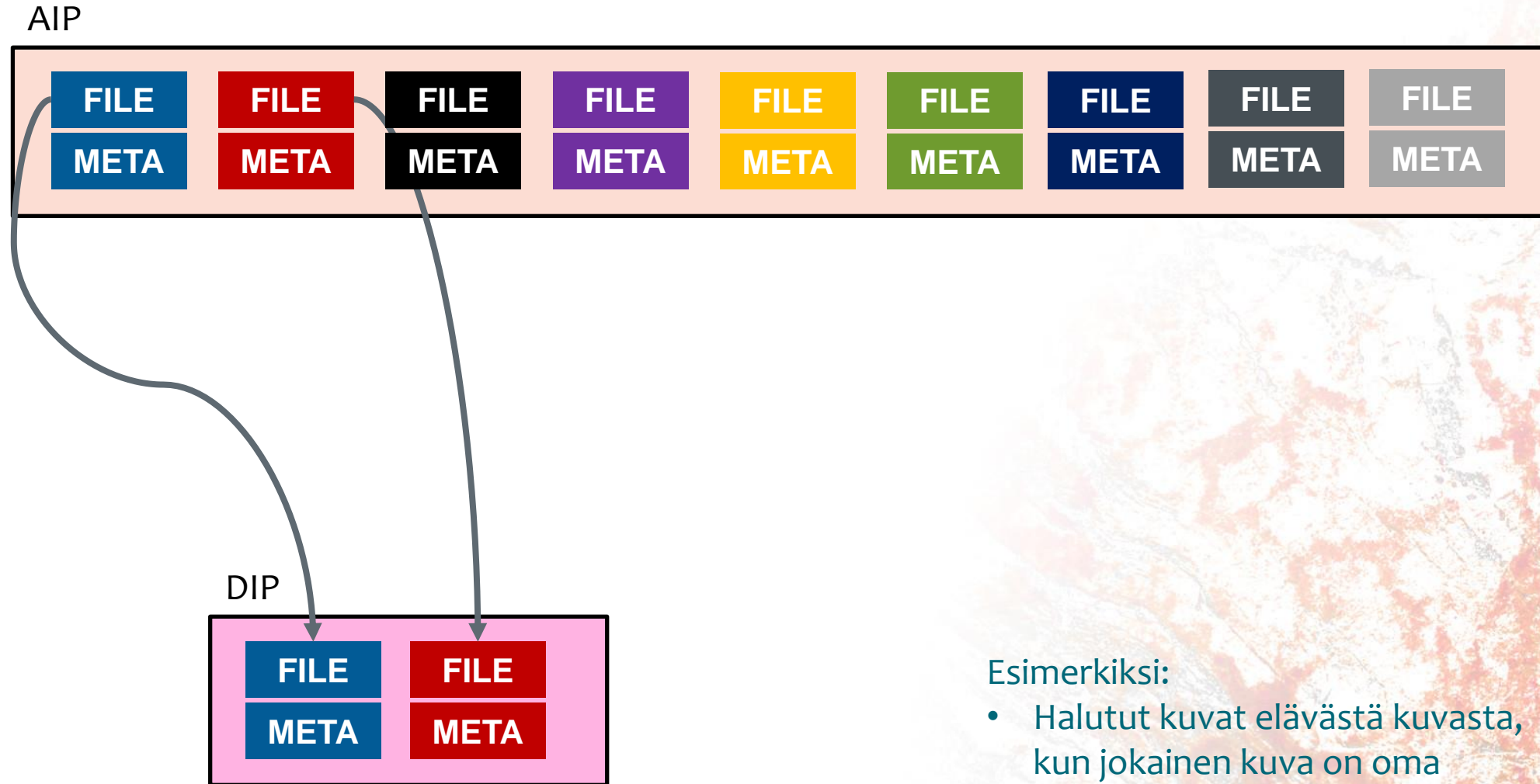
- Nykytila: Raportin nouto siirtopaketin tunnisteiden avulla (REST)
  - Tarvitaan vähintään SFTP-tuen päättymiseen asti
  - Raportti on noudettava luettelon kautta, koska sama siirtopaketti on voitu siirtää useasti (esim. yksi hylätty ja yksi hyväksytty siirto)
  - Raporttia ei ole saatavilla, jos siirtopaketin tunnistetta ei pystytty lukemaan (ZIP-purku epäonnistui, METS puuttuu, ...)
- Tavoitetila: Raportin nouto siirron tunnisteiden avulla
  - Siirto on aina yksikäsitteinen ja raportin saa suoraan palautusviestinä
  - Raportin saa, vaikka siirtopaketin tunnistetta ei pystytty lukemaan
  - Olemassa olevien siirtojen tunnisteista voi olla saatavilla luettelo

# Tavoitetila: Jakelupaketin muodostaminen

(muusta kokonaisuudesta kuin yhdestä kokonaisesta säilytyspaketista)



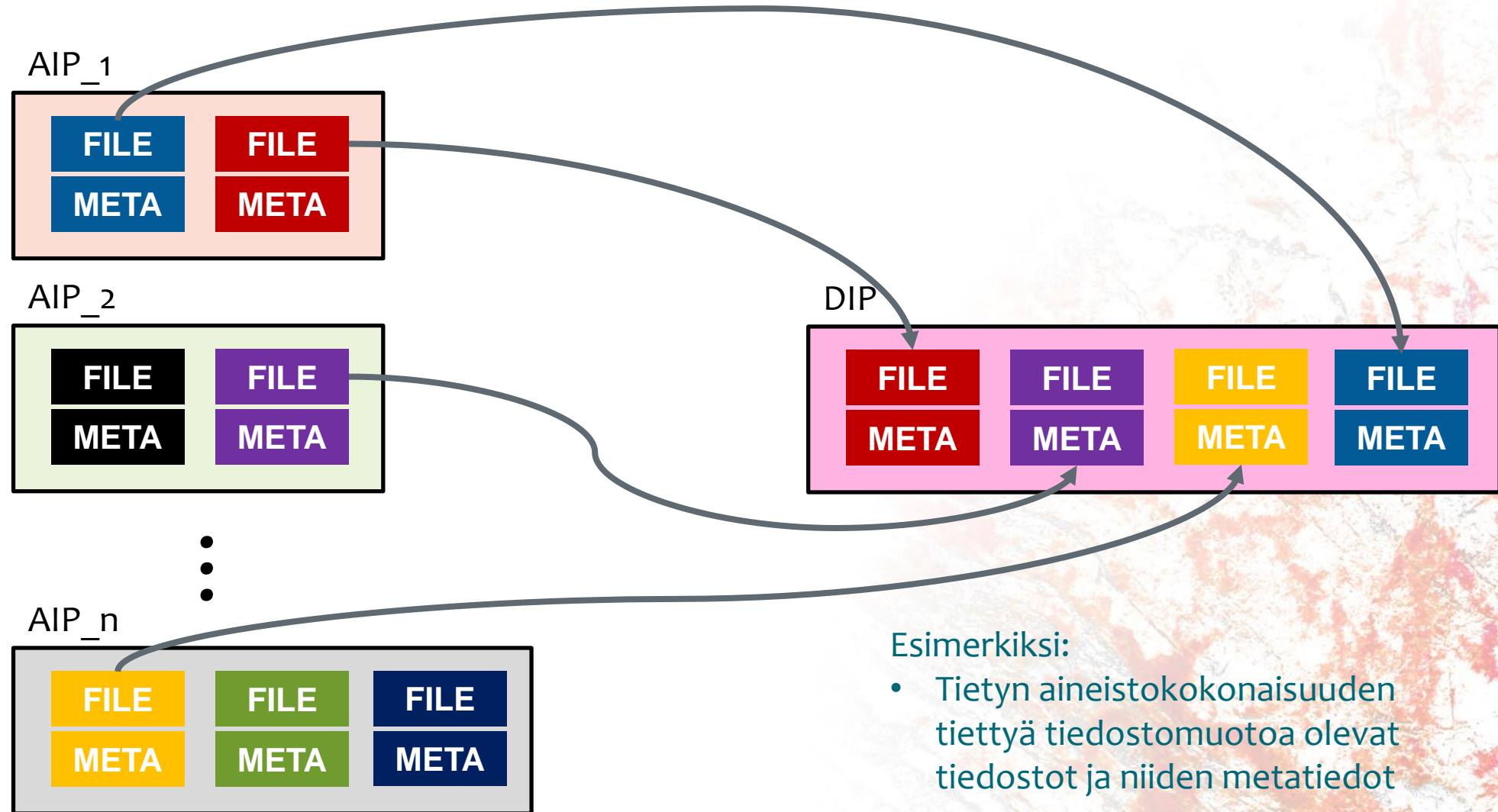
# Vain osa isosta säilytyspaketista



Esimerkiksi:

- Halutut kuvat elävästä kuvasta, kun jokainen kuva on oma tiedostonsa

# Tietopakettien yhdistäminen



Esimerkiksi:

- Tietyn aineistokokonaisuuden tiettyä tiedostomuotoa olevat tiedostot ja niiden metatiedot

# Tavoitetila

- Haluttujen säilytyspakettien tunnisteluettelo saadaan hakutoiminnon vastauksena, kuten nykyäänkin. Tätä voi käyttää apuna halutun kokonaisuuden koostamisessa
- Säilytyspakettien tunnisteluettelon voi lähettää PAS-palvelulle haluamiensa tiedostojen tunnistetietojen kera, jolloin PAS-palvelu kerää näistä säilytyspaketeista halutut tiedostot ja muodostaa niistä yhden jakelupaketin.

- Esimerkiksi seuraavankaltainen luettelo:

**Säilytyspaketti A:** tiedosto 1, tiedosto 2, tiedosto 123

**Säilytyspaketti D:** tiedosto 6, tiedosto 7

Tuottaa yhden jakelupaketin, jossa koottuna tiedostot 1, 2 ja 123 säilytyspaketista A sekä tiedostot 6 ja 7 säilytyspaketista D, sekä näiden metatiedot

# Esimerkki: Sisällön muodostaminen jakelupaketiksi (1/2)

- Esimerkki: POST <base>/<contract>/content\_request
- Mukana luettelo tiedostoista:

```
{  
  "content": [  
    {  
      "package": "A",  
      "files": ["1", "2", "123"]  
    },  
    {  
      "package": "D",  
      "files": ["6", "7"]  
    }  
  ]  
}
```

## Esimerkki: Sisällön muodostaminen jakelupaketiksi (2/2)

- Vastausviesti on sama kuin yksittäisen säilytyspaketin muodostamisessa jakelupaketiksi:

```
HTTP/1.1 202 Accepted
Location: <base>/<contract>/disseminated/<dip-id>
...

{
  "status": "success",
  "data": {
    "disseminated": "<base>/<contract>/disseminated/<dip-id>"
  }
}
```

- Viesti sisältää jakelupaketin resurssin
- Ja tästä edetään, kuten jakelupakettien osalta tähänkin asti

# Metatietojen nouto

- Nykyisin jakelupaketti muodostetaan komennolla

```
POST <base>/<contract>/preserved/<aip-id>/disseminate
```

- Tämän rinnalle lisätään toiminnot pelkän metatiedon tai historian noudolle (ilman digitaalisia objekteja):

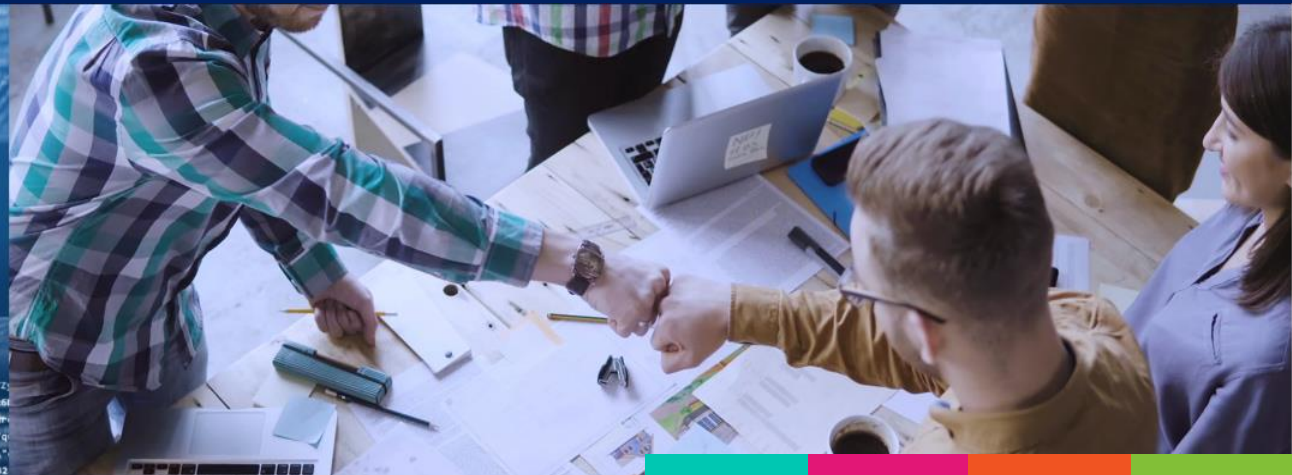
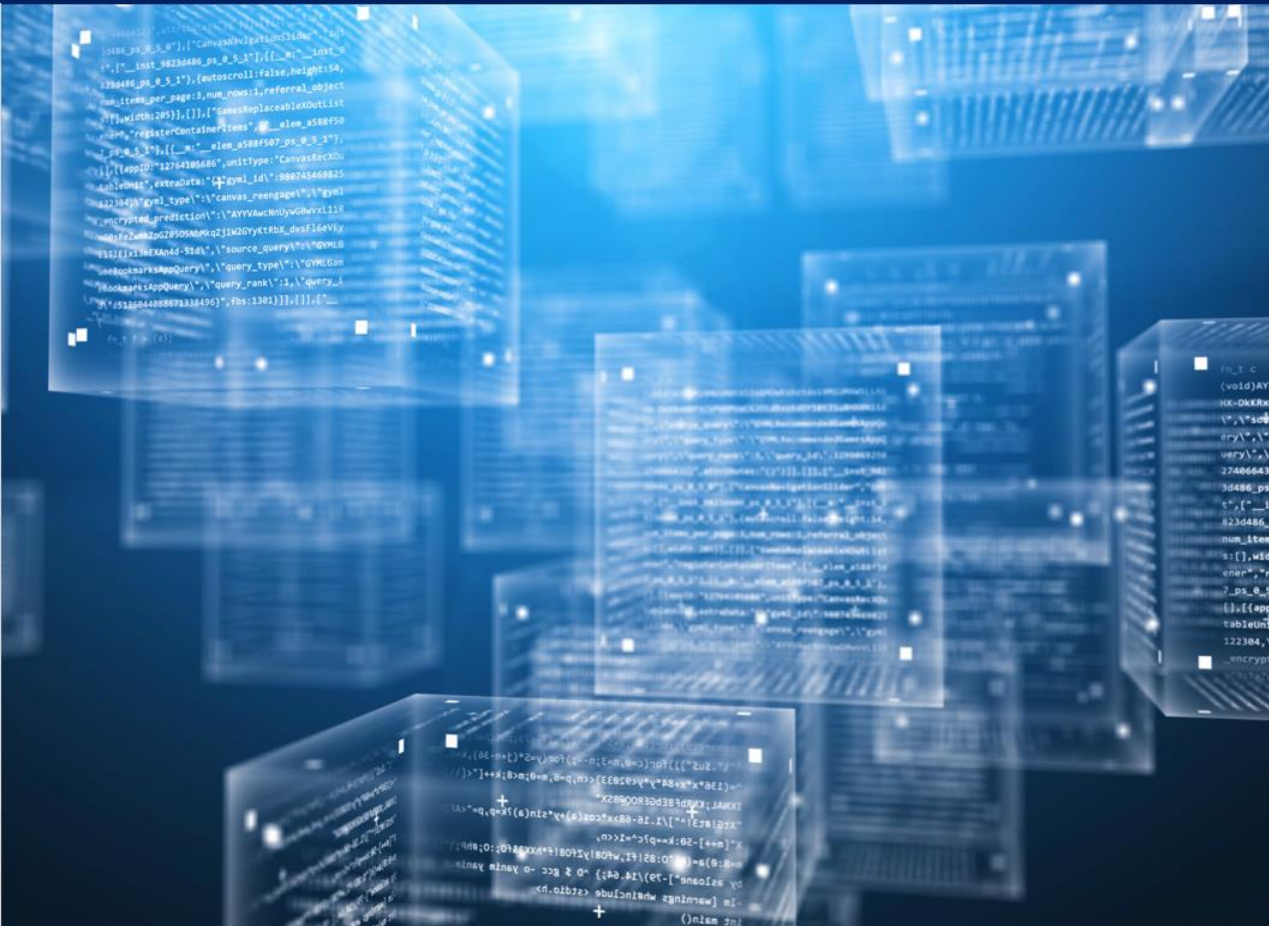
- POST <base>/<contract>/preserved/<aip-id>/disseminate\_metadata

- POST <base>/<contract>/preserved/<aip-id>/disseminate\_history

- Mahdollisesti näistäkin muodostuu asynkroniset operaatiot, koska metatiedot pitää ensin koostaa



# Tavoitetila: Tilastointi



# Tilastotiedot

- Alustava priorisoitu lista ensi vuodelle
  1. Kapasiteetti (yhteensä, käytettävissä)
  2. Hyväksytyjen/hylättyjen siirtopakettien määrä
  3. Säilytyksessä olevien tiedostojen määrä kokonaisuudessaan
  4. Tiedostomuoto tyypeittäin tiedostojen määrä (kpl, data); mime tai pronom
  5. Säilytyspaketin muodostamiseen käytetty aika
- Kohdat 2-4 myös ajan funktiona
- Kaikki tietenkin vain ja ainoastaan sopimuskohtaisesti

# Aikataulusta



# Alustava aikataulu

- Määrittelyihin ensi vuoden päivityskierroksella
- Upload REST toteutus Q4/2022
  - SFTP:n deprekointi Q1/2023
    - Tuki pois kokonaan Q4/2024
  - Rajapinnan version 1 tuen lopettaminen Q4/2024
  - Paketointimäärittelyiden 1.7.x vanhempien versioiden tuen lopettaminen Q4/2024
- DIP toteutus
  - Osa yhdestä AIPista vuonna 2023
  - DIP:n muodostaminen useamman AIP:n osista vuonna 2024