



PAS-yhteistyöryhmä

30.3.2021

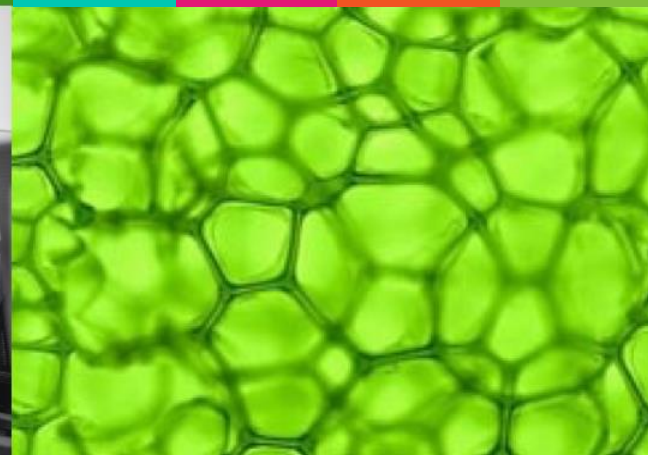


Kokouksen avaus

- Todetaan läsnäolijat
- Esityslistan hyväksyminen
- Edellisen kokouksen muistio



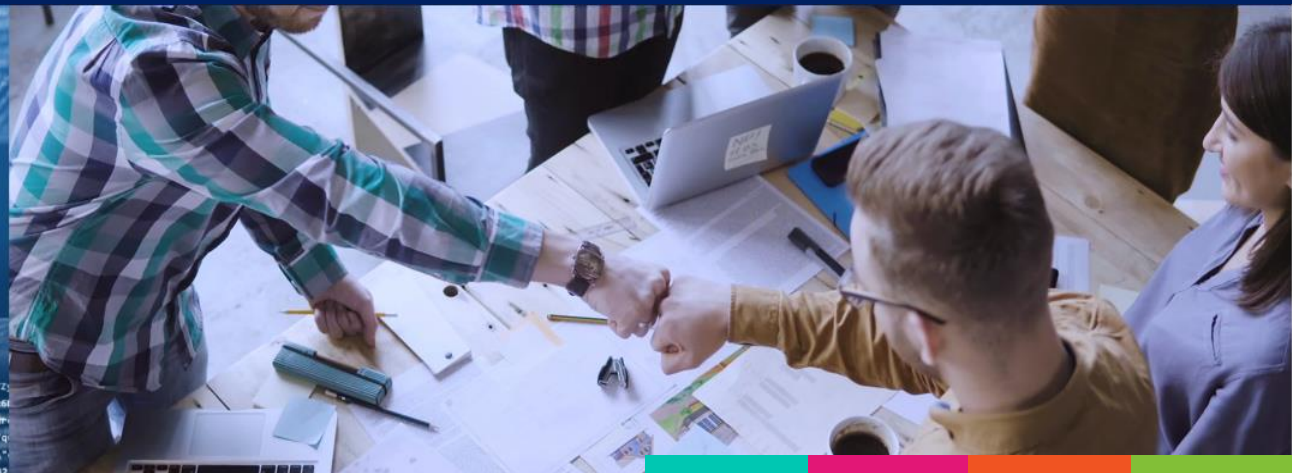
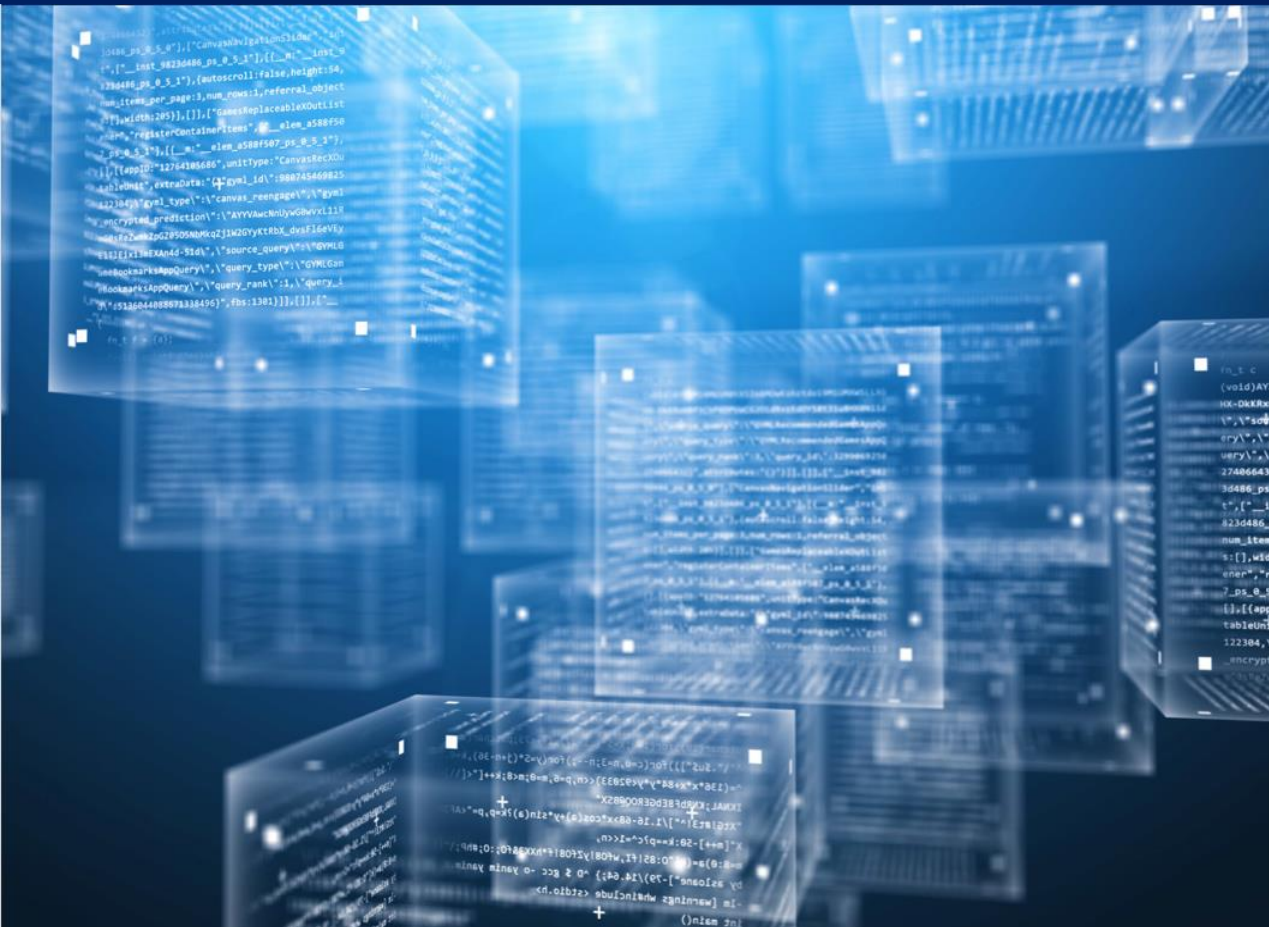
Looginen säilyttäminen



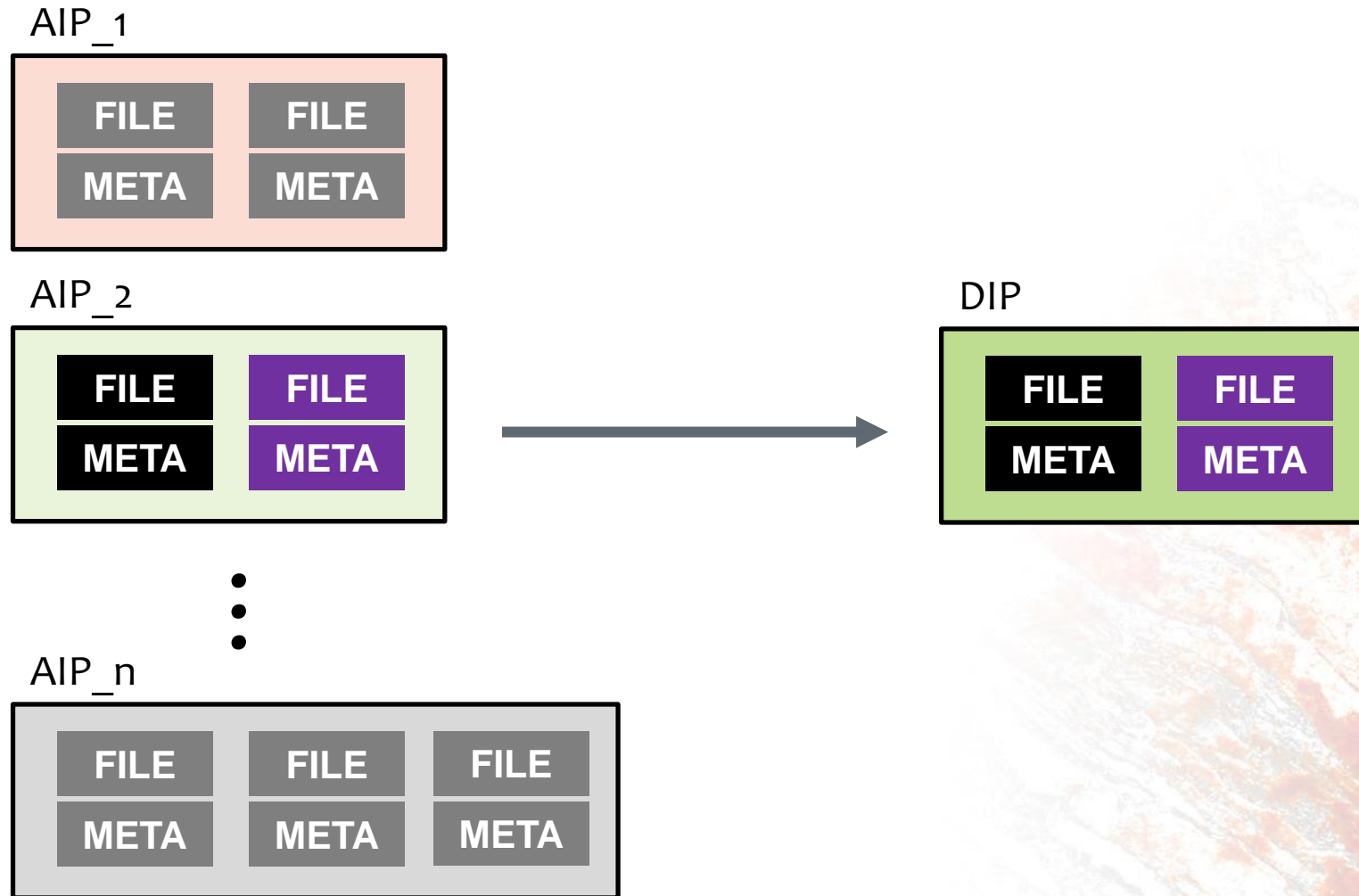
Looginen säilyttäminen – Vaatimusmäärittely

- Dokumentin versio 0.94 oli esityslistan liitteenä
- Dokumenttiin saatu palautetta eri tavoilla ja muutokset on huomioitu nykyisessä versiossa
 - Muutokset lähinnä täsmennyksiä ja kielioppi korjauksia
- Dokumentti vastannee organisaatioiden tämän hetkistä käsitystä loogisesta säilyttämisestä?
- Dokumentin perusteella voitaneen edetä kohti tarkempaa ominaisuuksien suunnittelua?

Rajapintojen jatkokehitys



Nykytila



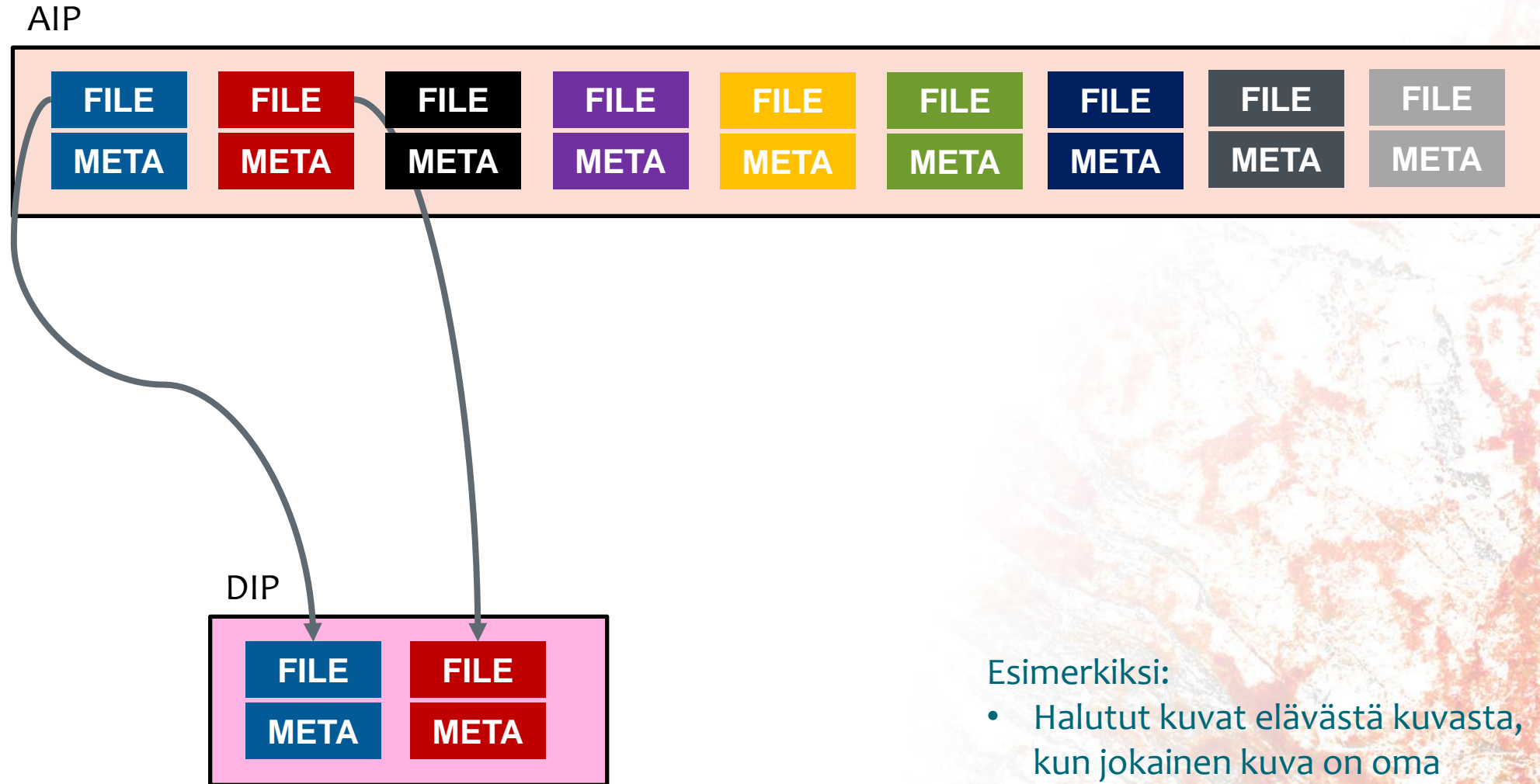
Nykytila

- Hyödyntävän organisaation ohjelmisto voi etsiä aineistoa halutuilla hakukriteereillä rajapinnasta
 - Metatiedot on indeksoitu, joten mihin tahansa aineiston metatietoihin voi kohdistaa hakukyselyt
 - Vastauksena tulee tunnisteluettelo säilytyspaketeista, joissa kyseinen hakuehto toteutuu
- Vastauksena saatua luetteloä voi käyttää suoraan hyväksi aineiston muodostamisessa jakelupaketeiksi
 - Halutuista aineistopaketeista voi muodostaa ja noutaa jakelupaketit käymällä luettelon koneellisesti läpi
- Rajapinnasta kuitenkin puuttuu toiminnallisuus säilytyspaketin osan muodostamiseksi jakelupaketiksi, kuten myös usean säilytyspaketin osien muodostaminen yhdeksi jakelupaketiksi.

Tavoitetila: Uusi ominaisuus

- Säilytyspakettien tunnisteluettelon voi lähettää PAS-palvelulle haluamiensa tiedostojen tunnistetietojen kera, jolloin PAS-palvelu kerää näistä säilytyspaketeista halutut tiedostot ja muodostaa niistä yhden jakelupaketin.
 - Säilytyspakettien tunnisteluettelon esimerkiksi saa vastauksena etsimällä aineistoa halutuilla hakukriteereillä rajapinnasta.
- Uusi ominaisuus ei muuta nykyään olemassa olevia ominaisuuksia

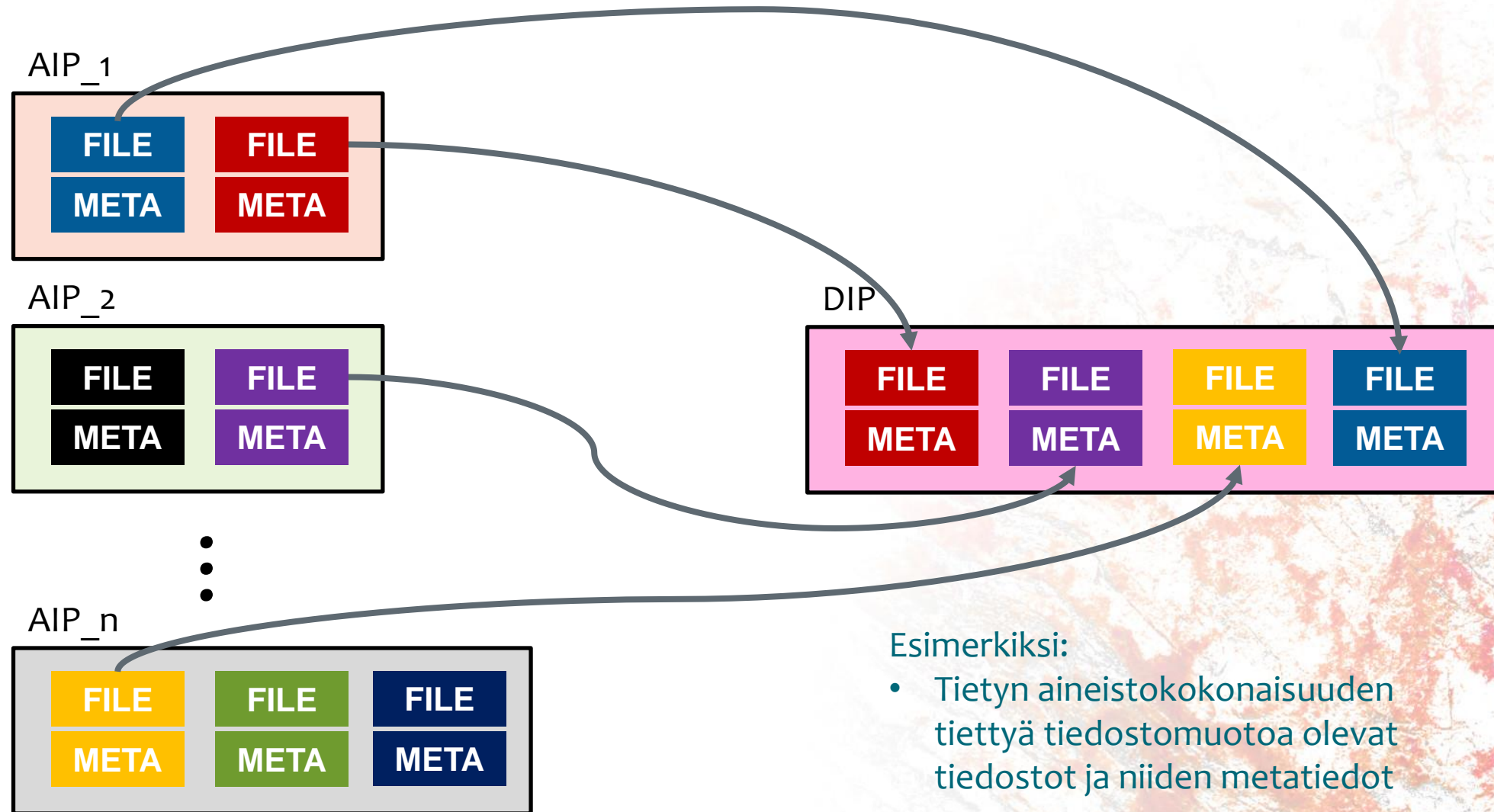
Tavoitetila – vain osa isosta säilytyspaketista



Esimerkiksi:

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

Tavoitetila – tietopakettien yhdistäminen



Esimerkiksi:

- Tietyn aineistokokonaisuuden tiettyä tiedostomuotoa olevat tiedostot ja niiden metatiedot

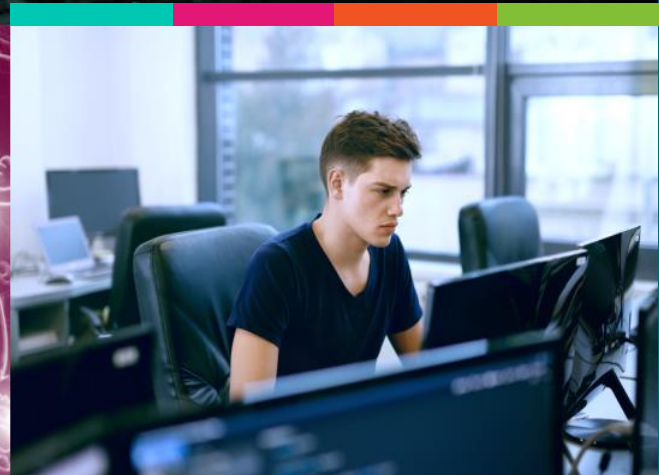
Aineiston siirtäminen PAS-palveluun (REST)

- Nykytila: Aineisto siirretään PAS-palveluun SFTP-rajapintaa hyödyntäen
 - SFTP:llä aikanaan päästiin nopeasti alkuun, mutta se on kankea ja vaikeasti hallittava tapa.
- Toiminto voidaan toteuttaa REST:iin
 - Mahdollistaa helpommin aidon automaattisen kommunikoinnin järjestelmien välillä
 - Jokainen siirto on yksilöllisesti seurattavissa omalla siirtotunnisteellaan
 - Vastaanoton tarkastusraporttia ei tarvitse etsiä kotihakemistoista, vaan sen voi saada REST-keskustelun tuloksena.
 - Ei tarvita enää kahta erilaista rajapintaa (SFTP+REST)
 - Valmiit työkalut kommunikointiin olemassa
 - Helpompi ja kevyt käyttää, mahdollistaa ketterämmin eri tyyppisten ratkaisujen toteuttamisen
 - HTTPS/REST-automatiikka yksinkertaistaa arkkitehtuuria verrattuna SFTP:hen

Aineiston siirtäminen PAS-palveluun (REST)

- Hylättyä aineistoa ei tarjottaisi REST-rajapinnasta
 - Toiminto on ollut SFTP-rajapinnassa, mutta käytännössä sitä ei ole tarvittu mihinkään
 - Hylätyllä aineistolla ei ole säilytysvastuuta, jolloin aineisto on joka tapauksessa oltava hyödyntävällä organisaatiolla hyväksymiseen asti
- SFTP-rajapinnan deprekointi pitkällä aikavälillä?

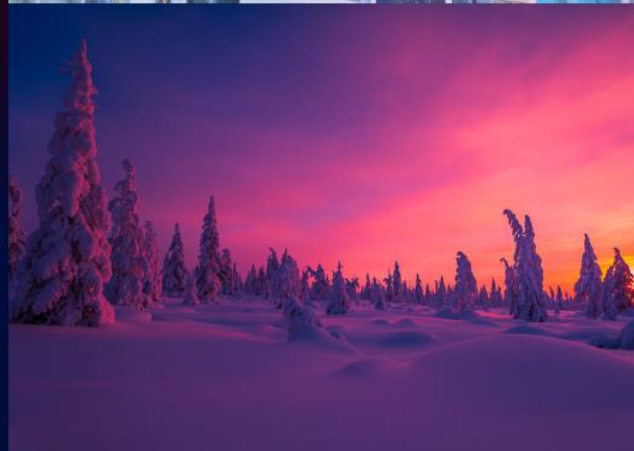
Digitaaliset allekirjoitukset säilytettävissä aineistoissa



Digitaaliset allekirjoitukset säilytettävissä aineistoissa

- Nykyinen rajoitus *"Tiedostoja ei saa allekirjoittaa digitaalisesti"* on liian tiukka
- Ehdotus: Muutetaan rajoitus muotoon *"Tiedostossa oleva digitaalinen allekirjoitus ei saa estää tiedoston käsittelyä"*
- On kuitenkin huomattava, että digitaalinen allekirjoitus kuitenkin rikkoontuu (viimeistään) migraatioissa
 - Eli digitaalista allekirjoitusta ei luvata säilyttää
- Muutos tiedostomuodot määrittelyn seuraavaan versioon

PAS-palveluiden uudet ominaisuudet Q1/2021



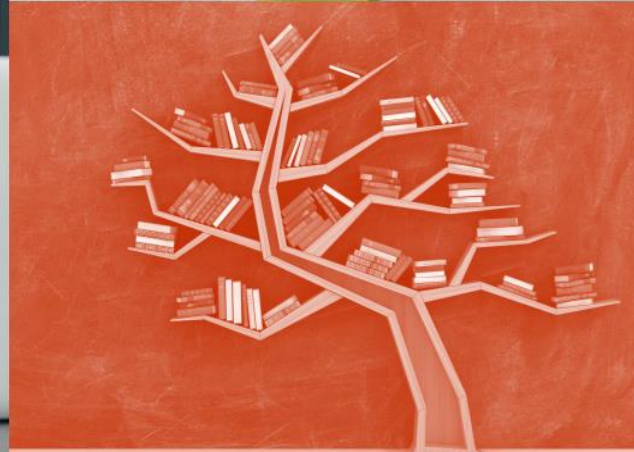
PAS-palveluiden uudet ominaisuudet Q1/2021

- Maantieteellinen hajautus on valmis ja tuotannossa
 - Säilytyksessä oleva aineisto on jo kopioitu toiseen sijaintiin
- Määritysmuutokset
 - Toistaiseksi vastaanoton tarkastusraporttiin palautetaan uuden "accession"-tapahtuman rinnalla myös vanha "preservation responsibility change"-tapahtuma
- Tiedostomuotovalidoinnin korjauksia, mm.
 - Kontrollimerkkien käsittely tekstitiedostoissa
 - Unicode-merkistöjen käsittely HTML5-tiedostoissa
- SAPA-kehitystyö omassa projektiryhmässään erillisen suunnitelman mukaisesti
- Fairdata PAS:
 - Paketointipalvelun siirto uuteen laitealustaan
 - Siirtyminen kohti Python3-kehitystä

Hieman CentOS7:sta ja Python3:sta

- Python2 on jo vanhentunut
 - Pikkuhiljaa siirrytään Python3:een...
- Python3:n myötä paketoitukomponentti uudistuu
 - Nykyinen on jo kaikkine ominaisuuksineen aika monimutkainen
 - Nykyisessä suorituskyky on aika huono suurilla aineistomäärillä
 - Uudistamisen suunnittelu on vasta alkamassa
- CentOS7:ssa elinkaari 2024 asti
 - CentOS8:ssa elinkaari kuitenkin vain vuoden 2021 loppuun asti
 - Asiaa selvitetään, toistaiseksi jatketaan CentOS7:ssa

PAS-palveluiden säilytyksen tilanne ja käynnissä olevat käyttöönotot



Sopimukset – KP-PAS (1/2)

(22.3.2021)



Organisaatio	Kapasiteetti (Tt)	Aineistoa (Tt)	Täyttöaste	Säilytyspaketteja
Celia	80	63,51	79,39 %	23 066
Kavi	1200	910,59	75,88 %	2 447
Kansallisarkisto Kansallisarkiston vastaanottamat alkujaan digitaaliset valtionhallinnon asiakirjalliset aineistot	41	0,28	0,68 %	408
Kansallisarkisto VAPA-järjestelmään siirretyt tietoaineistot	1	0,14	14,00 %	389
Kansallisarkisto Kansallisarkiston massadigitointi-hankkeen aineistot	114	24,18	21,21 %	21 578
Kansallisarkisto Kansallisarkiston digitaaliarkistosta siirrettävät aineistot ja takautuvan digitoinnin aineistot	805	34,58	4,30 %	36 885
Kansallisarkisto Kansallisarkiston yksinomaan digitaalisessa muodossa olevat yksityisarkistoaineistot	27	0	0,00 %	0

Sopimukset – KP-PAS (2/2)

(22.3.2021)



Organisaatio	Kapasiteetti (Tt)	Aineistoa (Tt)	Täyttöaste	Säilytyspaketteja
Kansallisgalleria	20	0,75	3,75 %	237
Kansalliskirjasto Kulttuuriaineistolain nojalla kerätyt aineistot	165	145,01	87,88 %	1 374 935
Kansalliskirjasto Kansalliskirjaston digitoimat kulttuuriperintöaineistot	175	14,73	8,42 %	3 889
KOTUS	60	9,71	16,18 %	403
Museovirasto	1	0,23	23,00 %	20 719
Musiikkiarkisto	70	0	0,00 %	0
Svenska litteratursällskapet i Finland	50	1,14	2,28 %	166
Yhteiskuntatieteellinen tietoarkisto	1	0,01	1,00 %	1 593
Yhteensä	2 810,00	1 204,86	42,88 %	1 486 307

Säilytyspäätökset – FD-PAS

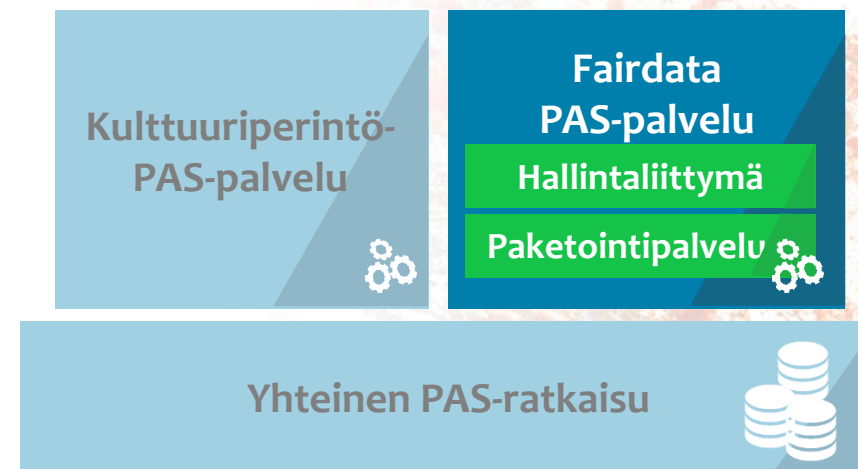
(22.3.2021)



Organisaatio	Kapasiteetti (Tt)	Aineistoa (Tt)	Täyttöaste	Säilytyspaketteja
Geologian Tutkimuskeskus GTK:n tomografialaitteen tuottamat tietoaineistot	12	2,53	21,08 %	60
Helsingin yliopisto Helsingin yliopiston SMEAR-aineistojen valikoima meteorologisia - ja ilmanlaatumittauksia	2	0,01	0,55 %	13
Helsingin yliopisto M. cinxia and C. melitaeorum in the Åland metapopulation system	2	0,00	0,05 %	1
Jyväskylän yliopiston kiihdytinlaboratorio 250-Nobeliumin hajoamisspektroskopia	1	0	0 %	0
Oulun yliopisto, Sodankylän geofysikaalinen observatorio Havaintoaineistot	30	0	0 %	0
Turun yliopisto Historian, kulttuurin ja taiteiden tutkimuksen arkiston aineistot (HKT-arkisto)	20	0	0 %	0
Yhteensä	67	2,54	3,79 %	74

Ilmoitusasiat

- Määritykset päivitetty ja julkaistu
 - Sekä suomeksi että englanniksi
- ”Digitaalisten aineistojen tapahtumat” koulutus 22.4.2021
- Digime-standardisalkku 1.0. on julkaistu
- Fairdata-verkosto edistämään yhteistyötä ja tiedonkulkua
 - Fairdata PAS-palvelu osana verkostoa



Kokouksen päättäminen

- Muut asiat
- Seuraava kokous
 - 17.6.2021 (sovittu doodlella aiemmin)
- Kokouksen päättäminen

