



# PAS-yhteistyöryhmä 14.6.2022



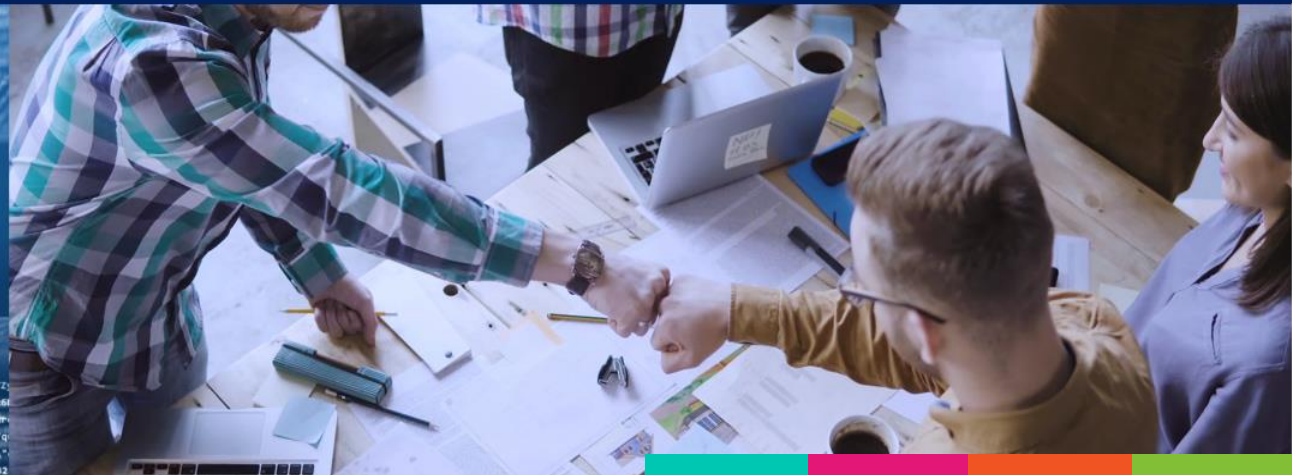
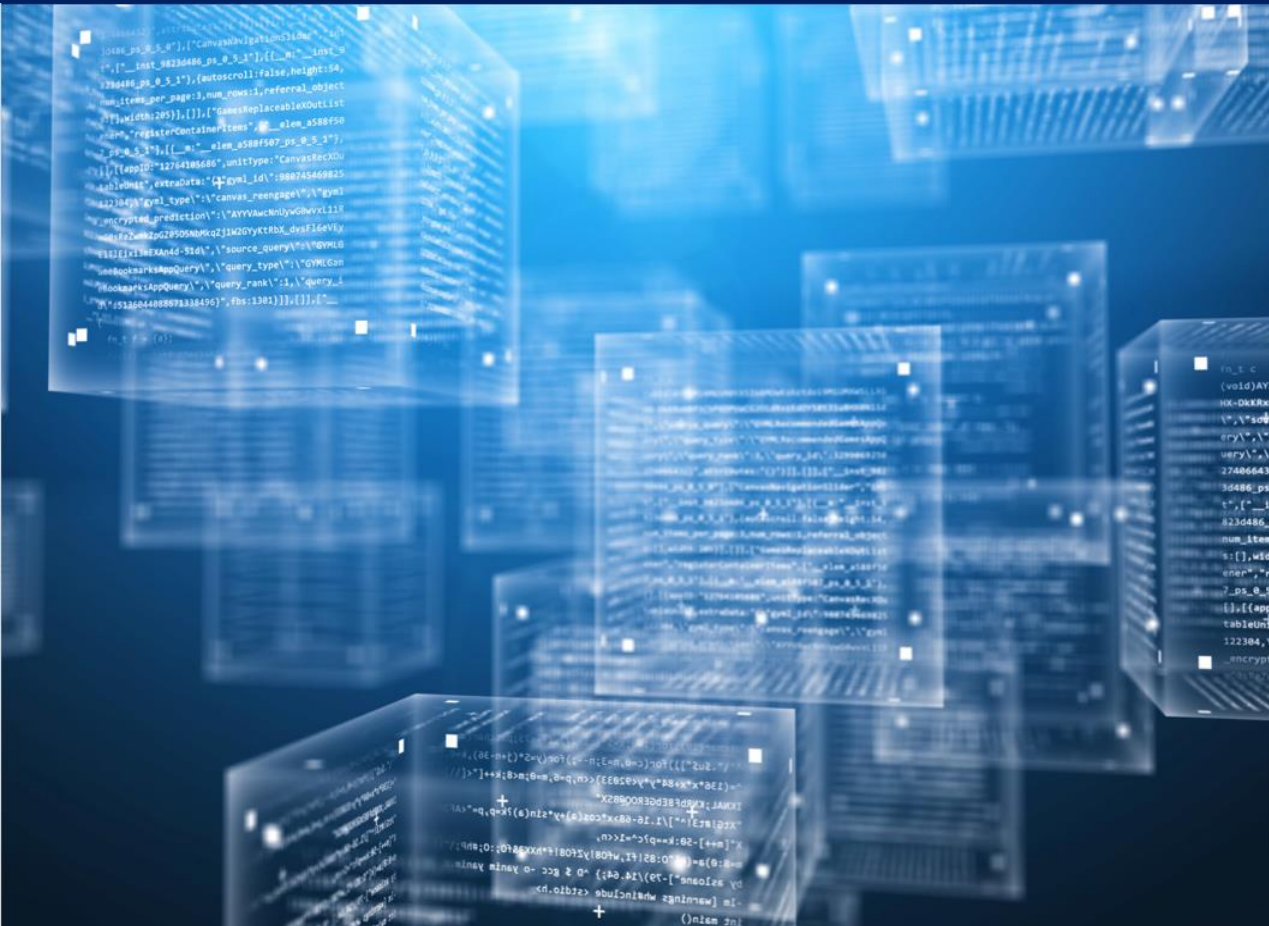
# Kokouksen avaus

- Todetaan läsnäolijat
- Esityslistan hyväksyminen
- Edellisen kokouksen muistio



# PAS-palveluiden kokonaisarkkitehtuuri

Alexi Salo, CSC



# ARC/WARC-migraatiotyö

- Migroitavaa aineistoa on noin 80 teratavua, satoja tuhansia tietopaketteja
  - ARC- ja WARC-tiedostot melko yhtenäisiä, muutamia haasteellisempia tapauksia tiedossa
  - Tehdyt valinnat vaikuttavat laajaan määrään aineistoa
- Verkkoarkisto on ollut PAS-palvelussa vuodesta 2015 alkaen
  - PAS-määrittelyt ovat monin tavoin muttuneet
    - Esim. attribuutti aineiston tunnisteelle (fi:CONTENTID) on ollut olemassa vasta myöhemmin
  - Metatietojen yhtenäistäminen sekä METS:ssä että digitaalisten objektien otsaketiedoissa
- Migraatioprosessin suunnittelu ja toteutus jatkuu
  - Työ on ollut mm. Python3-migraation myötä alkuvuodesta 2022 tauolla
  - Aiemmin on jo toteutettu mm. migraatiotyökalu
    - Työkalu osaa migroida yhden ARC/WARC-tiedoston
    - Työkalua on laajasti testattu Verkkoarkistossa
    - Saatavilla Githubista: <https://github.com/Digital-Preservation-Finland/warc-migrator>

# ARC/WARC-migraatiotyö – Erilaisia piirteitä



- Tiedostomigraatio
- Metatietojen käsittely
  - METS ja tiedostojen otsaketiedot
  - Kuvaileva ja hallinnollinen metatieto
  - Tunnisteiden käsittely
  - Tapahtumahistoria
- Uusien pakettien allekirjoittaminen
- Koemigraation toteutus
- Migraatio on aineistomäärän vuoksi laaja työ
  - Työvuon toteutus
  - Toipuminen ja käyttäytyminen virhetilanteessa
  - Prosessin rinnakkaistaminen hajautetussa ympäristössä
  - Kirjanpito

# ARC/WARC-migraatiotyö – Loppuraportti

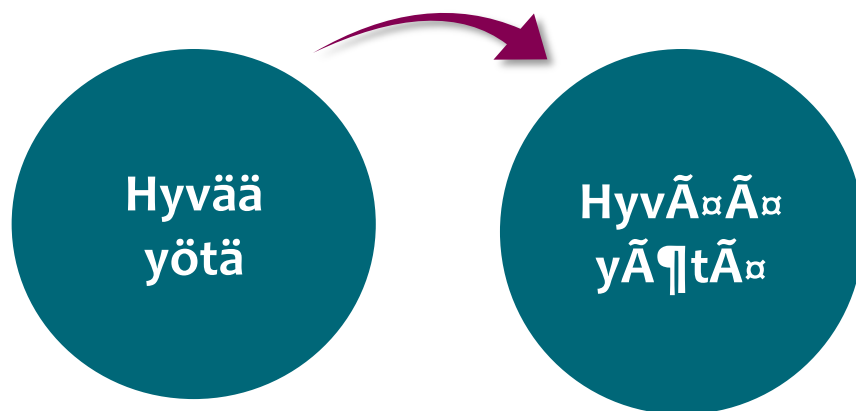
- Loppuraportti dokumentoi koko migraatioprosessin
- Loppuraportille on olemassa raporttipohja
  - Pohjautuu aiemmin tehtyyn Loogisen säilyttämisen vaatimusmäärittely –dokumenttiin
- ARC/WARC-migraatio tuo raporttipohjalle käytännön koeponnistuksen
  - Pohjaa päivitetään saatujen kokemusten mukaisesti

# PAS-palveluiden ohjelmistokerros

- PAS-palvelut ovat siirtyneet käyttämään Python 3:sta
  - Julkaistut komponentit tukevat sekä Python 2:sta että Python 3:sta
  - Asennusohjeet päivittyvät Python 3-yhdenmukaisiksi
- Python 2 tuki poistetaan 1.3.2023
- CentOS 7 elinkaari päättyy kesällä 2024
  - Tuotanto siirtymässä maksulliseen RHEL9 ympäristöön
  - Kehittäminen ja testaaminen (komponenttien virallinen tuki) eri ympäristössä kuin RHEL9
  - Ympäristö ja aikataulu tarkentuu tämän vuoden aikana

# Tiedostoihin upotettujen metatietojen merkistöt

- Tiedostot saattavat sisältää upotettuja metatietoja
  - Esimerkiksi Exif, XMP ...
- Tuntematonta merkistöä on vaikeaa tunnistaa koneellisesti
  - Migraatio rikkoo semanttisen ymmärrettävyyden jos lähtömerkistö ei ole tunnettu
- Määrittelyihin tulossa tarkennus näiden metatietojen suhteen
  - Hyväksytään vain ASCII tai UTF-8 merkistöjä





# PAS-palveluiden uudet ominaisuudet Q2/2022



- Validointikapasiteetin laajentaminen
- Python3-ohjeistus README-tiedostoihin
- Prosessi Github-muutospyyntöjen käsittelemiseksi
- EPUB-tiedostomuototuki
  - Samalla tuki uudemman JHOVE:n kanssa
- Korjauksia MPEG-4- ja Quicktime-tiedostomuotojen tunnistukseen
- Määrittelymuutokset 1.7.4
- Indeksoinnin ja tilastoinnin korjauksia
- Bugikorjauksia ja muita pienempiä kehitystöitä
- SAPA-kehitystyö omassa projektiryhmässään
- Fairdata PAS
  - Uuden paketointityökalun suunnittelu
  - Säilytyspäätösکوhtaisten autorisoinnin suunnittelu
  - Aineistotiedostojen hallintatoimintoja (valmisteluvaiheen tallennuspalveluun liittyen)
  - AVI-tiedostomuototuki
  - Bugikorjauksia ja muita pienempiä kehitystöitä (mm. merkistöongelmien korjaus)

# Sopimukset – KP-PAS (1/2)

(8.6.2022)



Organisaatio	Kapasiteetti (Tt)	Aineistoa (Tt)	Täyttöaste	Säilytyspaketteja
Celia	110	79,93	72,66 %	29 269
Kavi	2 400	1 277,05	53,21 %	3 352
Kansallisarkisto Kansallisarkiston vastaanottamat alkujaan digitaaliset valtionhallinnon asiakirjalliset aineistot	41	0,30	0,73 %	577
Kansallisarkisto VAPA-järjestelmään siirretyt tietoaaineistot	1	0,14	14,00 %	389
Kansallisarkisto Kansallisarkiston massadigitointi-hankkeen aineistot	114	61,64	54,07 %	51 913
Kansallisarkisto Kansallisarkiston digitaaliarkistosta siirrettävät aineistot ja takautuvan digitoinnin aineistot	805	272,26	33,82 %	545 827
Kansallisarkisto Kansallisarkiston yksinomaan digitaalisessa muodossa olevat yksityisarkistoaaineistot	27	0	0,00 %	0

# Sopimukset – KP-PAS (2/2)

(8.6.2022)



Organisaatio	Kapasiteetti (Tt)	Aineistoa (Tt)	Täyttöaste	Säilytyspaketteja
Kansallisgalleria	20	4,22	21,10 %	414
Kansalliskirjasto Kulttuuriaineistolain nojalla kerätyt aineistot	355	197,46	55,62 %	1 919 324
Kansalliskirjasto Kansalliskirjaston digitoimat kulttuuriperintöaineistot	175	14,73	8,42 %	3 889
KOTUS	60	9,71	16,18 %	403
Museovirasto	1	0,53	53,00 %	40 046
Musiikkiarkisto	70	0	0,00 %	0
Svenska litteratursällskapet i Finland	50	1,14	2,28 %	166
Yhteiskuntatieteellinen tietoarkisto	1	0,05	5,00 %	8 301
<b>Yhteensä</b>	<b>4 230,00</b>	<b>1 918,98</b>	<b>45,37 %</b>	<b>2 603 870</b>

# Säilytyspäätökset – FD-PAS

(8.6.2022)



Organisaatio	Kapasiteetti (Tt)	Aineistoa (Tt)	Täyttöaste	Säilytyspaketteja
Geologian Tutkimuskeskus GTK:n tomografialaitteen tuottamat tietoaineistot	12	5,10	42,50 %	102
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
Helsingin yliopisto FIRE (The Finnish Reflection Experiment)	1	0,00	0 %	0
Itä-Suomen yliopisto SENSOTRA	1	0	0 %	0
Jyväskylän yliopiston kiihdytinlaboratorio 250-Nobeliumin hajoamisspektroskopia	1	0	0 %	0
Oulun yliopisto, Sodankylän geofysikaalinen observatorio Havaintoaineistot	30	0	0 %	0
Tampereen yliopisto Yhteiskuntatieteiden tiedekunnan Kansanperinteen arkiston A-K-kokoelma	2	0	0 %	0
Turun yliopisto Historian, kulttuurin ja taiteiden tutkimuksen arkiston aineistot (HKT-arkisto)	20	0	0 %	0
<b>Yhteensä</b>	<b>71</b>	<b>5,11</b>	<b>7,20 %</b>	<b>116</b>

# Ilmoitusasiat

- CSC sai jatkoa ISO 27001-sertifikaatille vuoteen 2025 asti
- Määrittelyiden päivittäminen kysely käynnissä
- Digitaalisen kulttuuriperinnön pyöreä pöytä, kyvykkyyks-kysely käynnissä
- OKM on julkaissut ehdotuksen kulttuuriperintöstrategiaksi
- METS Editorial Board esittää iPRES-konferenssissa paperin METS2:sta
- #PASKlinikka 20-21.6

# Kokouksen päättäminen

- Muut asiat
- Seuraavat kokoukset
  - Sovittu doodlella aiemmin: 22.9 & 15.12
- Kokouksen päättäminen

