



# PAS-yhteistyöryhmä

## 21.9.2021

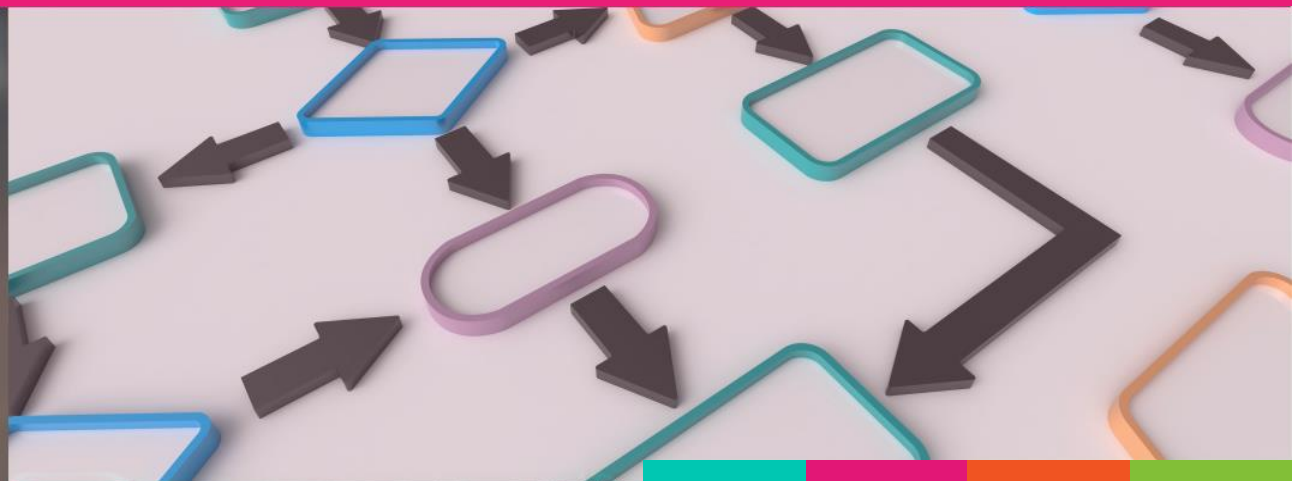


# Kokouksen avaus

- Todetaan läsnäolijat
- Esityslistan hyväksyminen
- Edellisen kokouksen muistio



# Määrittelyiden päivittäminen 2021



# PAS-määrittelyiden päivittäminen

- Webropol kysely (31.5-20.8.2021)
  - Kolme vastausta
  - Tuttuun tapaan, palautetta saatu myös päivittäisessä PAS-työssä
- Dokumenttien päivitys (CSC)
  - Päivitetyt versiot yhteistyöryhmälle kommenteille 26.11.2021 mennessä
- Hyväksyminen
  - Yhteistyöryhmä (14.12.2021)
- Julkaisu
  - Heti kun muutokset on toteutettu tuotantoon ja skeemoihin
  - Tavoite: Helmikuun aikana (muutokset tuotantoon maaliskuun huoltokatkossa)

# PAS-määrittelyiden päivittäminen

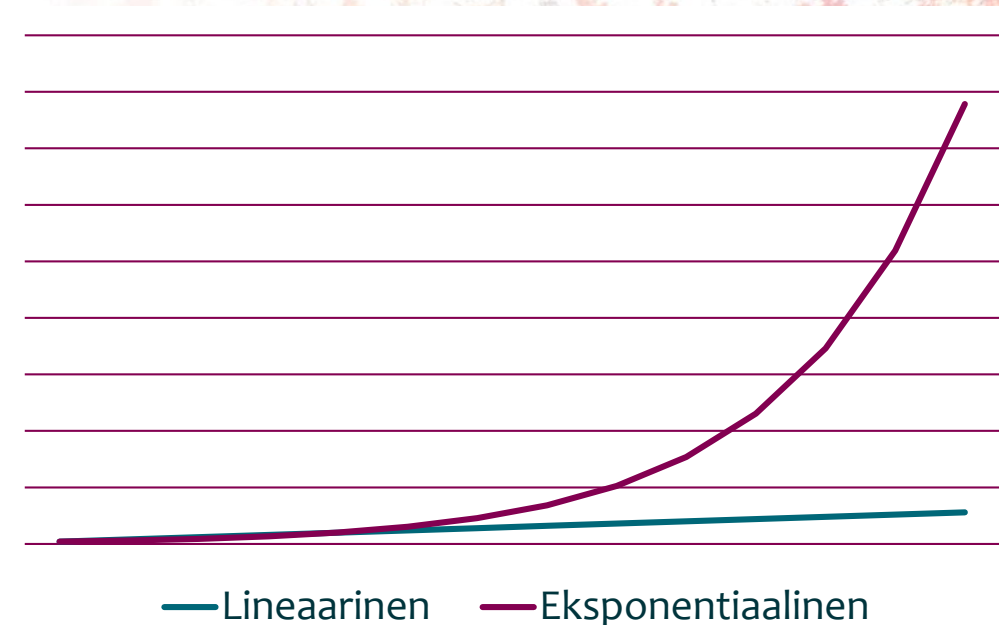
## Metatiedot/paketointi

- Kuvailevat metatiedot
  - DataCite v.4.4
    - Ok
  - LIDO v.1.1 beta
    - Ei lisätä vielä, koska beta
    - Palautetta toivotaan LIDOn käyttäviltä organisaatioilta
  - Inspire
    - Ehdollisesti ok. Vaatii tarkempaa tutkimista
    - Koska Inspire on kuitenkin XML pohjainen, niin käyttäminen on toki mahdollista jo nyt
- PREMIS 3.0
  - Ei edelleenkään vielä
- EN15744 poistaminen määrittelystä?

# PAS-määrittelyiden päivittäminen

## Metatiedot/paketointi

- Yhteen siirtopakettiin sisällytettävien tiedostojen määrän rajoittaminen
  - Käytännössä pakko tehdä, koska validoinnin aikavaativuus ei ole lineaarinen suhteessa tiedostojen määrään
  - Haasteena erityisesti XML-tiedostojen koon kasvu (METS, vastaanottoraportti), ei niinkään digitaalisten objektien koko tai niiden validoiminen (toki niilläkin on vaikutusta)
    - Pelkästään koodia optimoimalla haasteesta ei päästä eroon
- Päätösehdotus
  - Suositeltava max. määrä 20 000 tiedostoa/paketti
  - "Kova raja" 50 000 tiedostoa/paketti, jolloin paketti hylätään ilman, että validointia edes aloitetaan
    - Rajoituksilla ei pitäisi olla vaikutusta nykyisille hyödyntäville organisaatioille



# PAS-määrittelyiden päivittäminen

## Tiedostomuodot



- Uusia tiedostomuotoja
  - Geopackage
    - Ehdollisesti ok, mutta vaatii lisäselvittelyä
  - Matroska (mkv)
    - Tämä on jo mukana...
- Uusia versioita
  - PDF/A-4
    - Haasteena ehkäpä validaattorin puute, mutta selvitellään
  - FLAC uudet "versiot"
    - Nämä uudet versiot eivät edelleenkään ole uusia versioita tiedostomuotoon
  - SIARD 2.1.1
    - Ok, ja samalla poistetaan 2.1
  - GML uusi versio 3.2.2.
    - Ok

# PAS-määrittelyiden päivittäminen

## Tiedostomuodot

- Poistetaan siirtokelpoisista PDF1.2 ja 1.3
  - Näitä todennäköisesti on hyvin vähän jos yhtään?
  - Helpompi normalisoida ennen siirtoa jos näitä jostain löytyy?
- Päivitetään digitaalisten allekirjoituksen rajoitus
  - Sovittu jo aiemmin
- Edelleen toivottu että siirtokelpoisten sallittaisiin uusia versioita...
  - Vaikka niitä ei enää pitäisi tuottaa
  - Ristiriita?
- Joitakin korjauksia määrittelyn taulukoihin
  - Käytännössä typoja, eli taulukko ei vastaa leipätekstiä



# PAS-määrittelyiden päivittäminen

## Rajapinnat

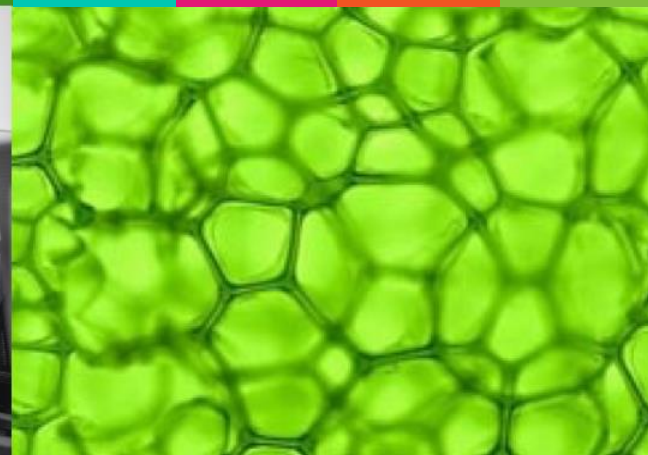
- Ei varsinaisia kommentteja
- Erilaisia tilastotietoja toivottu enemmän ja näitä on tulossa.
- Erillinen työpaja tulossa, mutta sen tulokset eivät välttämättä päädy vielä seuraavaan määrittelyn versioon

# PAS-määrittelyiden päivittäminen

## Sekalaista palautetta

- Yhteinen suljettu työskentelyalusta
  - Onko tarvetta?
  - Riittääkö tilapäiseen tarpeeseen esim. googledocs? Jota olemme käyttäneet lähiaikoina esim. loogisen säilyttämisen vaatimusmäärittelyyn ja NDSA matriisin kääntämiseen.
- Yhteinen avoin "työskentelyalusta"
  - Esim. linkkejä tiedonlähteisiin
- Erilaisia koulutuksia pyydetty paljonkin
  - Kyllä näitä tullaan järjestämään mahdollisuuksien mukaan

# Valittujen kansainvälisten määrittelyiden suhde kansallisiin PAS-palveluihin: Case Sword3



# Valittujen kansainvälisten määrittelyiden suhde kansallisiin PAS-palveluihin



- PAS-palveluille on aika ajoin ehdotettu erilaisiin kansainvälisiin määrittelyihin tutustumista suhteessa PAS-palveluihin
  - Tarkoituksena on, että PAS-palvelu tutustuu valittuihin määrittelyihin, tekee vertailua kansallisiin PAS-määrittelyihin sekä esittelee tästä nousevia asioita PAS-yhteistyöryhmässä keskustelua varten
  - Tavoitteena on arvioida valittujen määrittelyiden soveltuvuutta ja tarpeellisuutta PAS-palveluihin liittyen
- Tässä kokouksessa on vuorossa Sword3-rajapintamäärittelyn esittely
- Myöhemmin tutustumme E-ARK-määrittelyihin
  - Alustavasti joulukuun PAS-yhteistyöryhmässä

# SWORD<sub>3</sub>

<https://sword.cottagelabs.com/>



- SWORD = Simple Web Service Offering **Repository Deposit**
- Määrittely valmistunut 1.9.2021
- Nykyajalle tyypillinen JSON-muotoon nojaava REST-rajapinta
  - Kuten PAS-palvelun REST-rajapinta
- Repository Deposit: Rajapintamäärittely tallentamiseen
  - Ei tue mm. jakelupaketteja. OAIS?
- Fokuksena tutkimusaineistot
- SWORDin jatkuvuus on epävarmaa

# SWORD<sub>3</sub>

- Tietopakettien muotona SWORDBagIt-muoto (mm. SHA-256, Dublin Core)
- Eri dokumenttityypit määritelty
  - Palveludokumentti, statusdokumentti, metadatatokumentti, viittausdokumentti, metadata+viittausdokumentti, segmentoidun siirron dokumentti, virhedokumentti
- Tilan seuranta työnkulussa
  - Palauttaa useille komennoille laajan oloisen JSON-muotoisen statusdokumentin, jossa tämän hetkinen tila työnkulussa
  - Tilaa seuraamalla (pollaamalla) tiedetään, milloin työ valmistuu
  - PAS-palvelun rajapinnassa on sama käytäntö

# SWORD<sub>3</sub>-määrittelyn toiminnot

- Voidaan jakaa yhteen tai useampaan "palveluun", myös sisäkkäin
  - "Palveluun" pystyy tallentamaan aineistoja
  - Kullakin "palvelulla" on yksi palveludokumentti, jossa on kuvaus palvelusta ja rajapinnan sallituista toiminnoista
  - Vrt. PAS-palvelun sopimustunnus muodostaa PAS-palvelun rajapinnassa kullekin sopimukselle oman "palvelun"
- Resurssin siirto palveluun
  - Resurssi: SWORD<sub>3</sub>-tietopaketti, tiedosto, tiedostoviittaukset tai metatietodokumentti
  - Yksittäisen tiedoston siirto palveluun myös useana samankokoisena segmenttinä
- Resurssiin lisääminen
- Resurssin vaihtaminen ja poisto
  - Vaihtaminen lähtökohtaisesti hävittää edellisen version
  - Vaihtaminen säilyttää alkuperäisen URL:in, poisto+siirto ei

# SWORD<sub>3</sub>-määrittelyn toiminnot

- Resurssin statusdokumentin nouto
  - Kertoo tietoja resurssista ja sen tämän hetkisen tilan työkulussa
  - Käytetään aineiston tilan seurantaan (pollaus) ja tiedostojen noutoon
  - Työnkulut: Siirto palveluun, lisääminen, vaihtaminen, poisto
  - Tiloina esim. *accepted, inProgress, inWorkflow, ingested, rejected, deleted*
- SWORD<sub>3</sub>-muotoisen metatiedon nouto (JSON)
- Yksittäisen tiedoston nouto **synkronisesti**
  - Myös ulkoinen (ei-Sword<sub>3</sub>) linkki saa olla mahdollista tiedostoille ja metatietodokumenteille
  - Tietopakettien statusdokumentin avulla pystyy käymään paketin tiedostot läpi
  - Synkronisuus estää asynkroniset operaatiot
    - Esim. jakelupaketin muodostus tai tarkistussumman laskenta ei onnistu
    - Tiedoston oltava heti rajapinnan saatavilla



# PAS-palvelun rajapinnan toiminnot

- Ryhmittely sopimustunnuksiin
- Siirtopaketin siirto
  - Aineiston siirto myös osissa mahdollista (CONTENTID)
  - Lisääminen myöhemmin mahdollista (CONTENTID)
- Päivityspaketin siirto
  - Vanhojen versioiden automaattinen säilyminen
- Siirto- tai päivityspaketin tarkastusraportin haku ja nouto
- Aineistojen luettelointi, suodatus halutuilla hakuehdoilla
- Tilan seuranta työnkulussa

# PAS-palvelun rajapinnan toiminnot

- Säilytyspaketin muodostus jakelupaketiksi
  - Suodatetun haun antaman luettelon avulla pystyy noutamaan halutut säilytyspaketit
- Säilytyspakettien osien muodostaminen jakelupaketiksi
- Jakelupaketin nouto ja poisto
- Metatietojen nouto
- Tilastotiedot

Ryhmittely ”palveluihin”

Yleiset

Ryhmittely sopimustunnuksiin  
Tilastotiedot

Resurssin siirto palveluun  
Lisääminen resurssiin  
Resurssin vaihtaminen  
Tiedoston siirto palveluun segmentoituna

Siirto palveluun

Siirtopaketin siirto  
Päivityspaketin siirto  
Tuki aineiston siirrolle osissa  
Tuki lisäämiselle  
Tuki vanhoille versioille  
Tarkastusraportin haku ja nouto

Resurssin tilan seuranta työkulussa  
Tiedostolista tietopaketeissa  
Resurssin poisto

Hallinta palvelussa

Tilan seuranta työkulussa  
Aineistojen luettelointi suodatuksella  
Säilytyspaketin muodostus jakelupaketiksi  
Osien muodostus jakelupaketiksi  
Jakelupaketin poisto

Tietopaketin SWORD3-muotoisen metatiedon nouto  
Yksittäisen tiedoston nouto synkronisesti  
Ulkoiset linkit (tiedostot/metatiedot)

Nouto palvelusta

Metatietojen nouto  
Jakelupaketin nouto

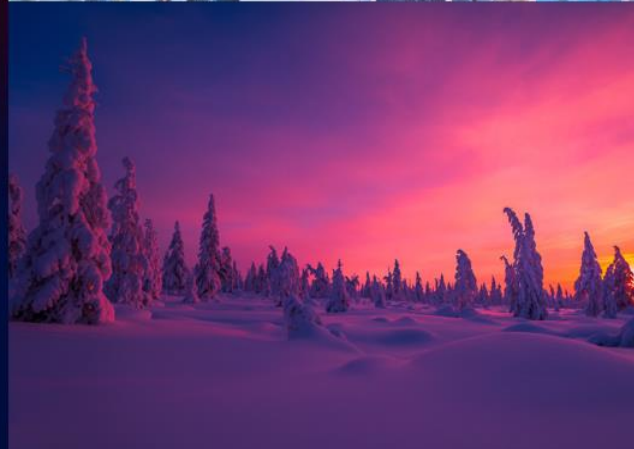
## Yhteenveto (1/2)

- SWORDin tulevaisuus epäselvä
  - SWORDin takana ei varsinaisesti mikään organisaatio
  - SWORDille etsitään uutta "kotia"
- SWORD<sub>3</sub> tuki muissa järjestelmissä
  - dSpace on kehittänyt uuden oman rajapinnan lähiaikoina, vaikkakin se vielä tukeekin SWORDin versioita v1 ja v2
  - Inveniossa on proof-of-concept toteutus; lienee ainoa maailmassa?
- Tarkkaan standardinmukaisuuteen ei PAS-palveluissa päästä
  - Omien PAS-laajennuksien toteuttaminen SWORD<sub>3</sub>:een olisi mahdollista
  - Osa SWORD<sub>3</sub> toiminnallisuuksista ei soveltune suoraan PAS-palveluun
    - => Rajapintaa ei voisi kutsua SWORD<sub>3</sub> yhteensopivaksi, eikä mahdolliset kaupan hyllyltä löytyvät ohjelmistot välttämättä toimisi

## Yhteenveto (2/2)

- Pelkästään SWORD<sub>3</sub> yhteensopivan rajapinnan toteuttaminen poistaisi nykyisiä PAS-palveluiden toiminnallisuuksia
  - Kahden rajapinnan ylläpitäminen ei liene realistista, tai ei ainakaan kustannustehokasta.
  - Kahden rajapinnan käyttäminen hyödyntävissä organisaatiossa?
- Joitakin asioita SWORD<sub>3</sub>:sta voidaan kuitenkin lainata
  - Palveludokumentti konsepti, JSON-dokumenttien kenttiä, RESTiin perustuva ingest on jo suunnitteilla ...
- Mitä lisäarvoa SWORD<sub>3</sub>-tuki PAS-palveluissa toisi organisaatiollenne?
  - Oletteko harkinneet SWORD<sub>3</sub> käyttöä joissakin palveluissanne?
  - Onko teillä tiedossa ohjelmistoja, joiden avulla aiotte hyödyntää SWORD<sub>3</sub>-rajapintaa?

# PAS-palveluiden uudet ominaisuudet Q3/2021



## PAS-palveluiden uudet ominaisuudet Q3/2021

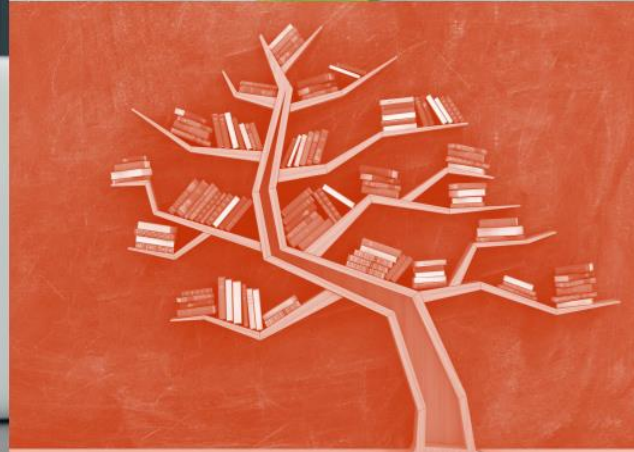
- Tilastointitoiminnon laajentaminen (valmistumassa)
- Migraatiotyökalu ARC/WARC → WARC
- Säilytyspäivämäärän lisääminen indeksiin
- Tarkastusraportin muodostumisen nopeuttaminen
- Tarkastusraportin virheilmoitusten selkeyttäminen teknisten piirteiden tarkastuksessa
- Erilaisia pieniä bugikorjauksia ja muita tehtäviä
- SAPA-kehitystyö omassa projektiryhmässään erillisen suunnitelman mukaisesti
- Fairdata PAS:
  - **Rajapintakomponentti PAS-palvelun REST-rajapinnan automaattiseen hyödyntämiseen**
  - AVI-tuki tiedostojen analysointikomponenttiin
  - Tiedostomuotojen luokittelu tiedostojen analysointikomponentissa
  - Erilaisia pieniä bugikorjauksia ja muita tehtäviä

# Rajapintakomponentti

- Työkalu PAS-palvelun REST-rajapinnan automaattiseen hyödyntämiseen
- Sisältää tällä hetkellä kaksi komentoa:
  - Säilytyspakettien listaus annetulla suodatuksella
  - Tietopakedin nouto
- Piilottaa hyödyntävältä ohjelmistolta (tai ihmiseltä):
  - PAS-palvelun rajapintaan kirjautumisen
    - Ensimmäisellä käyttökerralla tallennetaan kirjautumistiedot, jonka jälkeen työkalu hoitaa tämän jokaisella käyttökerralla automaattisesti
  - Pienimuotoisen vikasietoisuuden
    - Yrittää GET-muotoista komentoa automaattisesti enintään 5 kertaa satunnaisten yhteyskatkojen varalta
  - Jakelupaketin muodostamiseen ja noutamiseen liittyvän kommunikaation
    - Noutokomennon syöttäminen johtaa automaattiseen jakelupaketin muodostamisen aloittamiseen, muodostamisen seurantaan (pollaus) ja jakelupaketin noutoon



# PAS-palveluiden säilytyksen tilanne ja käynnissä olevat käyttöönotot



# Sopimukset – KP-PAS (1/2)

(8.9.2021)



Organisaatio	Kapasiteetti (Tt)	Aineistoa (Tt)	Täyttöaste	Säilytyspaketteja
Celia	110	67,29	61,17 %	24 449
Kavi	1 200	1 038,71	86,56 %	2 767
Kansallisarkisto Kansallisarkiston vastaanottamat alkujaan digitaaliset valtionhallinnon asiakirjalliset aineistot	41	0,28	0,68 %	408
Kansallisarkisto VAPA-järjestelmään siirretyt tietoaaineistot	1	0,14	14,00 %	389
Kansallisarkisto Kansallisarkiston massadigitointi-hankkeen aineistot	114	37,34	32,75 %	31 443
Kansallisarkisto Kansallisarkiston digitaaliarkistosta siirrettävät aineistot ja takautuvan digitoinnin aineistot	805	159,73	19,84 %	281 950
Kansallisarkisto Kansallisarkiston yksinomaan digitaalisessa muodossa olevat yksityisarkistoaineistot	27	0	0,00 %	0

# Sopimukset – KP-PAS (2/2)

(8.9.2021)



Organisaatio	Kapasiteetti (Tt)	Aineistoa (Tt)	Täyttöaste	Säilytyspaketteja
Kansallisgalleria	20	0,77	3,85 %	239
Kansalliskirjasto Kulttuuriaineistolain nojalla kerätyt aineistot	355	145,01	40,85 %	1 374 935
Kansalliskirjasto Kansalliskirjaston digitoimat kulttuuriperintöaineistot	175	14,73	8,42 %	3 889
KOTUS	60	9,71	16,18 %	403
Museovirasto	1	0,33	33,00 %	27 149
Musiikkiarkisto	70	0	0,00 %	0
Svenska litteratursällskapet i Finland	50	1,14	2,28 %	166
Yhteiskuntatieteellinen tietoarkisto	1	0,01	1,00 %	1 627
<b>Yhteensä</b>	<b>3 030,00</b>	<b>1 475,19</b>	<b>48,69 %</b>	<b>1 749 913</b>

# Säilytyspäätökset – FD-PAS

(8.9.2021)



Organisaatio	Kapasiteetti (Tt)	Aineistoa (Tt)	Täyttöaste	Säilytyspaketteja
Geologian Tutkimuskeskus GTK:n tomografialaitteen tuottamat tietoaineistot	12	3,28	27,33 %	75
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
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>70</b>	<b>3,29</b>	<b>4,70 %</b>	<b>89</b>

# Ilmoitusasiat

- Rajapinnat työpaja 26.10.2021 (alustavasti)
- #PASKaffet jatkuu normaalisti
- Seuraava #PASKlinikka (alustavasti) 15.12.2021
- [PDF Days 2021](#) 27.9-25.10.2021
- [iPRES2021](#) 19.-22.10.2021
- [Maailman digitaalisen pitkäaikaissäilytyksen päivä \(WDPD\)](#) 4.11.2021

# Kokouksen päättäminen

- Muut asiat
- Seuraavat kokoukset
  - 14.12.2021 (sovittu aiemmin doodlella)
- Kokouksen päättäminen

