



# PAS-yhteistyöryhmä

## 27.9.2019



# Kokouksen avaus

- Todetaan läsnäolijat
- Esityslistan hyväksyminen
- Edellisen kokouksen muistio

**Pyritään päättämään kokous viimeistään 10:45**

# Kulttuuriperintö-PAS-palvelusopimukset

## Tapahtunut edellisen PAS-yhteistyöryhmän jälkeen

- OKM:n ja CSC:n välinen PAS-sopimuksen uudistus saatiin valmiiksi kesäkuussa ja uusittu sopimus allekirjoitettiin 26.6.2019.
- Samalla valmistuivat myös Kulttuuriperintö-PAS- ja Fairdata PAS-palvelusopimukset
  - Julkaistu digitalpreservation.fi:ssä
- Kulttuuriperintö-PAS-palvelun osalta PAS-palvelusopimus...
  - on allekirjoitettu Celian, Kansallisarkiston, Kotuksen, Musiikkiarkiston ja Yhteiskuntatieteellisen tietoarkiston kanssa.
  - Loput sopimukset (KAVI, Kansalliskirjasto, Museovirasto ja SLS) odottavat organisaatiolta käsittelytoimien kuvaukseen liittyviä tietoja (Museovirasto omansa jo täyttänyt).
- Fairdata PAS-palvelun osalta sopimuksen täyttäminen on aloitettu GTK:n, Helsingin yliopiston, Jyväskylän yliopiston ja Oulun yliopiston kanssa.

# PAS-määrittysten päivittäminen

- Webropol kysely (24.6-30.8.2019)
  - 9 vastausta
  - Palautetta saatu myös käyttöönotoissa
- Dokumenttien päivitys (CSC)
  - Päivitetyt versiot yhteistyöryhmälle kommentoille 29.11.2019 mennessä
- Hyväksyminen
  - Yhteistyöryhmä (13.12.2019)
- Julkaisu
  - Alkuvuodesta, heti kun muutokset on toteutettu tuotantoon ja skeemoihin

# PAS-määrittysten päivittäminen

## Metatiedot



- Ehdotettu siirtymistä PREMIS versioon 3.0
  - 3.0 ei ole yhteensopiva versioiden 2.x kanssa. Samanaikainen tuki PAS-palvelussa haasteellista, koska PREMIS on varsin oleellinen osa tietopakettia.
  - Muutos 2.x => 3.0 vaikuttaa myös esim. validointiraportteihin ja siten tuotannossa olevien **hyödyntävien organisaatioiden toteutuksiin** tuomatta kuitenkaan uutta toiminnallisuutta, mutta mahdollisesti ylimääräistä työtä.
  - Ehdotamme, että PREMISin v3.0 siirytään kun meillä on ensimmäinen organisaatio, joka on ottamassa sitä tuotantoon omissa järjestelmissään
- EDTF korvataan ISO 8601-1:2019 ja ISO 8601-2:2019 –standardeilla
  - Asia sovittiin vuosi sitten, mutta päivitettyjä ISO 8601 standardeja ei silloin vielä oltu julkaistu
  - Muutos organisaatioille vähäinen, koska käytännössä aikaleimat ovat olleet vanhempien ISO 8601 mukaisia.
- Tuki Dublin Coren DC Terms laajennukselle
  - Käytännössä jo olemassa; tarkennetaan määrittystä

# PAS-määrittysten päivittäminen

## Säilytuskelpoiset tiedostomuodot

Tiedostomuotoihin saimme tänä vuonna useita kommentteja, mikä osoittaa organisaatioiden kiinnostuksen ja yhteistyön tärkeyden tässä asiassa. Tässä listattuna ja kommentteilla varustettuna saadut ehdotukset.

- DCP (elävä kuva)
  - Ei ole varsinainen tiedostomuoto, vaan joukko tiedostoja tietyssä rakenteessa
- DV
  - DV on siirtokelpoinen, ja siten "PAS kelpoinen" koska PAS-palvelu vastaanottaa myös siirtokelpoiset
  - Ei siirretä säilytuskelpoiseksi, koska tiedostomuodot määrittelyn tarkoituksena on myös ohjata mitä tiedostomuotoja kannattaa käyttää ja tämä muutos saattaisi ohjata organisaatioita väärään suuntaan
- EPUB 3.2
  - Lisätään uusi versio säilytuskelpoisiin tiedostomuotoihin
- IMF (elävä kuva)
  - Vaatii tarkemman analyysin. CSC ottaa työlisälle?

# PAS-määrittysten päivittäminen

## Säilytyskelpoiset tiedostomuodot

- IFC/A
  - Rakennetun ympäristön tietomallit
  - Seurataan tilannetta/tarvetta ja palataan asiaan myöhemmin
- JPEG2000 part 2 (.jpx)
  - Laajentaa osaa 1 mahdollistamalla esim. useiden kuvien upottamisen samaan tiedostoon.
  - Millaisessa tilanteessa tiedostomuodolle olisi todellista säilyttämisen tarvetta?
- RAW (3d)
  - Nimestä huolimatta ei liity kameroiden RAW tiedostomuotoihin
  - Vaatii tarkemman analyysin

# PAS-määrittysten päivittäminen

## Säilytyskelpoiset tiedostomuodot

- PDF/UA

- PDF/UA ei ole tiedostomuoto tai sen versio, vaan määrittelee dokumentille ominaisuuksia jotta dokumentti olisi saavutettava
  - Erityisesti: *"The version number of a file may be any value from 1.0 to 1.7, and the value shall not be used in determining whether a file is in conformance with this part of ISO 14289"*
- Mikä tahansa PDF dokumentti voi olla saavutettava vaikka ei olisi PDF/UA, tai voi olla teknisesti PDF/UA, mutta silti ei-saavutettava.
  - Saavutettava Word2016 dokumentti vs. ei-saavutettava Word2016 dokumentti?
  - HTML 4.01 dokumentti vs. HTML 4.01 dokumentti joka täyttää WCAG2.1 tasot A ja AA?
- Automaattinen validointi mahdotonta

- XHTML5

- Lisätään säilytyskelpoiseksi tiedostomuodoksi



# PAS-määrittysten päivittäminen

## Siirtokelpoiset tiedostomuodot

- DVCPRO
  - DV-perheen jäsenenä jo nyt siirtokelpoinen; täsmennetään määrittystä
- Office 2016
  - Lisätään siirtokelpoiseksi
- VideoMD
  - Metatietoformaatti elävälle kuvalle; XML muotoisena jo nyt säilytyskelpoinen
  - Epäselvää miksi VideoMD:n tulisi olla siirtokelpoinen tiedostomuoto
- .CRW (Canon) & .NEF (Nikon)
  - Kameravalmistajien omia tiedostomuotoja
  - Ei lisätä siirtokelpoisiksi, vaan näissä tapauksissa tulee käyttää esim. DNG tiedostomuotoa tai liitteen C vaihtoehtoa 1.

# PAS-määrittysten päivittäminen

## Säiliömuodot



- Matroska lisätään sallituksi säiliömuodoksi
- Selvitetään määrittelyssä olevien säiliömuotojen mahdolliset ongelmalliset yhdistelmät
  - Tällä hetkellä vain tietyt yhdistelmät (säiliömuoto + kuva/ääni) ovat määrittelyssä sallittuja, vaikka säiliömuodot tukevat laajempaa joukkoa jo määrittelyssä olevia muotoja
  - Kaikki yhdistelmät eivät kuitenkaan ole järkeviä
  - Tarkoituksena ei ole lisätä (tai poistaa) säiliömuotoja ja/tai tiedostomuotoja

# PAS-määrittysten päivittäminen

## Rajapinnat



- SWORD 3.0 rajapinta PAS-palveluun
  - SWORD 3.0 on lupaava määrittäminen, joka voisi olla hyödyksi PAS-rajapintojen kehitystyössä
  - Vaatii tarkemman analyysin ja yhteistyötä organisaatioiden kanssa
  - SWORDistä lisää seuraavilla kalvoilla
- Amazon S3 rajapinta PAS-palveluun
  - SWORDin kanssa "kilpaileva" rajapinta
  - Ei mukaan ainakaan tässä vaiheessa

# SWORD 3.0 ja PAS-rajapinnat

- SWORD 3.0 on Release Candidate –tilassa oleva rajapintamäärittely
  - Ei siis ole aivan valmis, mutta lähestulkoon
- Useita kehittäjiä, pääorganisaationa JISC
- SWORD-mallia käytetään ainakin Dspace-ohjelmistossa
  - Toteutukset olemassa aiemmille versioille 1 ja 2
- Perustuu REST-arkkitehtuurimalliin ja JSON-muotoon
  - PAS-palveluiden nykyinen rajapinta noudattelee REST+JSON-mallia (poislukien SFTP)
  - SWORD 3.0 ei määrittele kyselyiden tarkkaa syntaksia, vaan antaa toiminnoille raamit.
  - Kuitenkin, esimerkiksi kyselyille annettavat vastauskoodit ja HTTP-headerien käyttö on lueteltu, ja on lisäksi annettu muitakin ohjeita toteuttamista varten

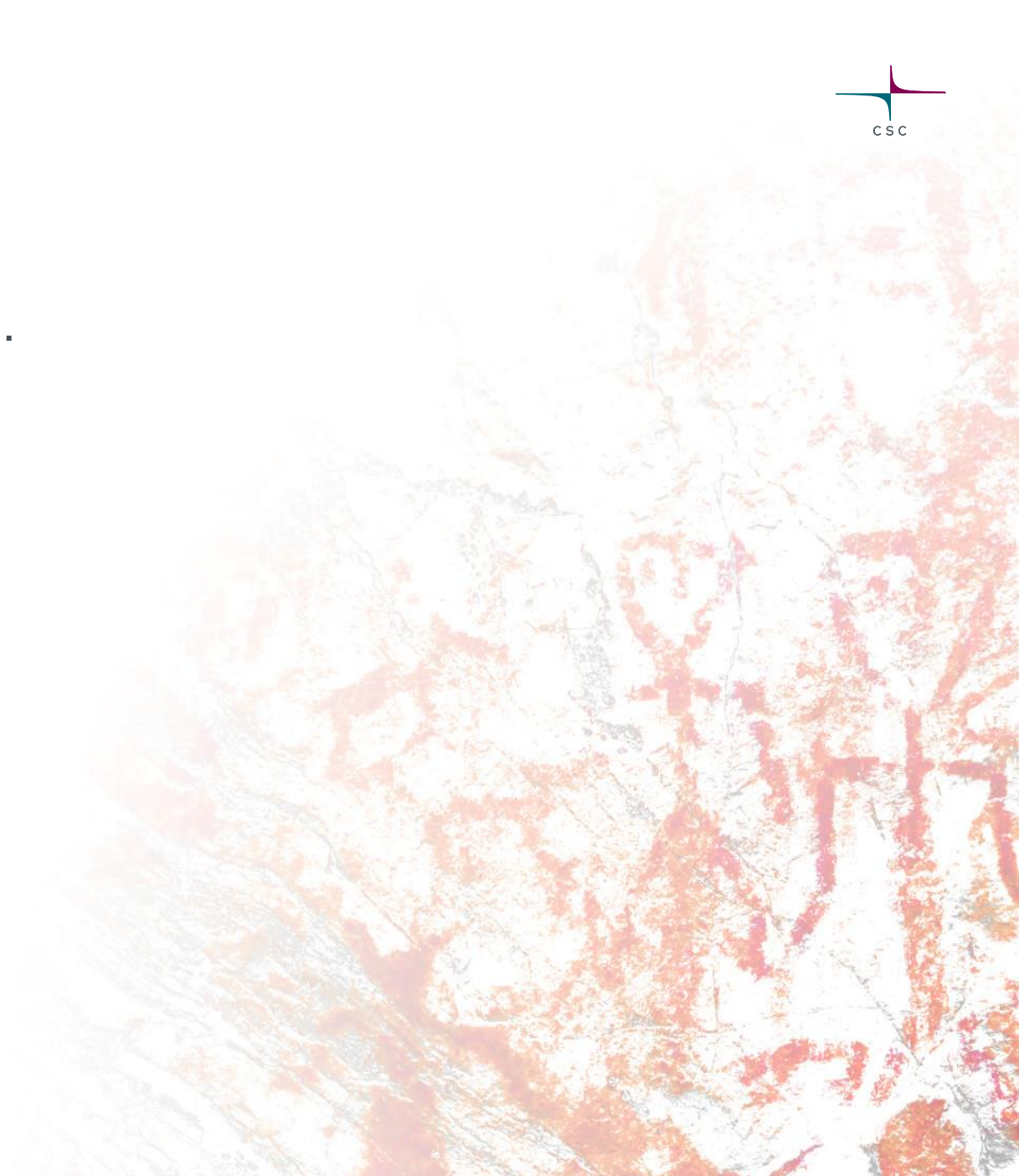
# SWORD 3.0 ja PAS-rajapinnat

- SWORD 3.0 sisältää samoja ajatuksia, kuin nykyinen PAS-palveluiden REST-rajapinta
  - Esim. 202-koodin palautus työn vastaanottamisesta, jonka valmistuminen tulee kestämään pitkään, tai validoitavien pakettien “accepted”/”rejected”-jako
- Määrittelyssä on kuitenkin joitakin sellaisia pakolliseksi luokiteltuja toimintoja, joille PAS-toiminnassa ei joko ole hallinnollista prosessia tai ei ole vielä ollut käytännön tarvetta
  - Esimerkiksi mahdollisuus aineiston poistamiseksi, tai siirtopaketin rakenteelle muutamat eri vaihtoehdot, jotka poikkeavat paketoitumäärityksestä
- Määrittelyksen uudesta versiosta olisi hyvä nähdä ensin jokin referenssitoteutus, esim. DSpace-kehitystyössä
  - Seurataan asiaa mielenkiinnolla
- SWORD 3.0 on lupaava määrittely, joka voisi olla hyödyksi PAS-rajapintojen kehitystyössä
  - Millä aikataululla organisaatiot ovat kehittämässä kokoelmanhallintajärjestelmiään sword 3.0 yhteensopiviksi?

# PAS-määrittysten päivittäminen

## Koulutukset – ehdotettuja teemoja

- Pakettien lähettäminen, validointiraportit, ...
  - Osa syksyn PAS-rajapinnat -koulutusta
- Korkean tason PAS-strategia koulutus
- Validointityökalut



# PAS-määrittysten päivittäminen

## Sekalaista



- Jakelupaketin rakenteen tarkentaminen
  - Määrittys määrittelee jo nyt jakelupaketin rakenteen; tarkennetaan tekstiä
- Digitaalisesti allekirjoitetut tiedostot?
  - Voidaan luvata vain bittitason säilytys, vaihtoehtoisesti tämä on ominaisuus, jonka migraatio olisi hyödyntävän organisaation vastuulla
- PAS-palveluiden saavutettavuus (ml. määrittymiset)?
  - Työnalla...
- Käyttöohjeiden parantaminen
  - Pyritään parantamaan jatkuvasti
- Ja joukko muita pienempiä kommentteja

# Loogisen tason säilyttäminen - etenemissuunnitelma

- Tässä kokouksessa sivuutetaan tämä asia, mutta palataan siihen kokouksien välillä
  - Prioriteetit toiminnallisuuksille organisaatioille?
  - Mitä organisaatiot tarvitsevat onnistuakseen loogisesta säilyttämisestä?





# Tapahtumatyytit - Taustaa

- Aineiston syntyhistoria kuvataan tapahtumina premis:event ja agentteina premis:agent -lohkoissa
  - Agentit suorittavat tapahtumia, ja tapahtumat liittyvät digitaalisiin objekteihin, aineistoon
  - Tapahtumat ovat tyypitetty (kontrolloidun sanaston käyttöä suositellaan)
  - Aineiston syntyhistoria on pakollinen siirtopaketin metatiedoissa
- PAS luo säilytyksen aikana tapahtumia, jotka kuvaavat aineistolle kohdistuneet toimenpiteet säilytyksen aikana
  - Tyypillisesti kyse on vastaanoton validointitapahtumat ja AIPin luonti, säännölliset eheydentarkastukset, tallennusmedioiden virkistäminen, jakelupaketin luonti
  - Tulevaisuudessa myös migraatio
- PAS-palvelun tapahtumat linkitetään aineiston metatietoihin jakelupaketin luonnin yhteydessä

# PREMIS tapahtumatyyppien määrittelyiden tarkentaminen

- Tapahtumatyyppien kontrolloidulle sanastolle on kaivattu läpikäyntiä
  - <http://digitalpreservation.fi/specifications/vocabularies>
- Listalle olisi tarve tehdä:
  - Määritellä ja kuvata PAS-palveluiden tuottamat tapahtumat (PAS-palvelu)
    - PAS-palveluiden tuottamat vastaanoton validointitapahtumat ovat kuvattu PAS-palveluiden rajapinnat-määrittelyssä
  - Käydä läpi listan muut tapahtumatyypit, kuvata ne paremmin (hyödyntävät organisaatiot)
    - Ja mahdollisesti lisätä uusia tapahtumatyyppejä
- Tavoitteet:
  - Yhdenmukaistaa hyödyntävien organisaatioiden käytännöt
  - Lista olisi hyödyllinen auttamaan hyödyntäviä organisaatioita kuvaamaan aineiston elinkaarta
- Organisaatioilta toivotaan palautetta (yhdessä tai erikseen) **25.10.2019** mennessä

# PAS-palveluiden uudet ominaisuudet Q3/2019

- Maantieteellisen hajauttamisen suunnittelu, toteutusvaihe menossa
- Arkkitehtuuri- ja ohjelmistokehitysdokumentaation sekä käytettyjen työkalujen QSOS-arviointien päivitys ja käännös englanniksi
- Python3-migraatio suurelta osin
- Digitaalisten objektien analysointikomponentin integrointi PAS-vastaanottoon
  - paketoitukomponenttiin integroitu jo aiemmin
- Em. analysointikomponentin parannus
  - metatietomallin selkeämpi eriytyminen toiminnallisuudesta
  - täydennys uusilla tiedostomuodoilla, kuten videomuodot
  - tehty korjauksia käyttöönotoissa havaittuihin puutteisiin liittyen
- Säännöllisen eheystarkastuksen tilastointikomponentti
- Staattinen koodianalyysi (tyyliopasta ja tietoturvaan koskevan automaatiotestauksen parantaminen)
- Pieniä optimointeja ja korjauksia
- SAPA-kehitystyö omassa projektiryhmässään erillisen suunnitelman mukaisesti
- Fairdata PAS:
  - Metax-ohjeistus ja muu Metax-dokumentointi
  - Metaxin PAS-katalogituki ja skeemojen päivitys
  - Käyttöönottoprosessin dokumentointi
  - Aineiston paketoinnin optimointi
  - Uuden autentikointiproxyyn käyttöönotto
  - TAR-tuki väliaikaiseen säilytyspalveluun
  - Vanhentuneiden aineistojen näyttäminen hallintaliittymässä
  - Pieniä optimointeja ja korjauksia

# PAS-palvelun säilytyksen tilanne ja käynnissä olevat käyttöönotot

## Tuotannossa olevat organisaatiot (20.9.2019)



Organisaatio	Aineistoa (Tt)	Säilytyspaketteja
Kavi	269,31	879
Kansallisarkisto / Kansallisarkiston vastaanottamat alkujaan digitaaliset valtionhallinnon asiakirjalliset aineistot	0,27	379
Kansallisarkisto / VAPA-järjestelmään siirretyt tietoaineistot	0,14	389
Kansalliskirjasto / Kulttuuriaineistolain nojalla kerätyt aineistot	142,18	1 344 168
Kansalliskirjasto / Kansalliskirjaston digitoimat kulttuuriperintöaineistot	14,73	3 888
KOTUS	9,71	403
Svenska litteratursällskapet i Finland	0,01	1
Yhteiskuntatieteellinen tietoarkisto	0,01	1 310
<b>Yhteensä</b>	<b>436,36</b>	<b>1 351 417</b>

# PAS-palvelun säilytyksen tilanne ja käynnissä olevat käyttöönotot



## Käynnissä olevat käyttöönotot

- Celia: Valmiina siirtymään tuotantoon.
- Museovirasto / Kansallisgalleria: Nämä kaksi museota yhdistäneet voimansa, koska tulevat jatkossa käyttämään samaa kokoelmanhallintajärjestelmää. Toivoneet yhdessä neuvottelua CSC:n kanssa 27.9.2019 pidettävän PAS-yhteistyöryhmän yhteydessä. Tavoite edelleen, että Kansallisgalleria pystyisi aloittamaan tuotannon tämän vuoden puolella. Aineisto on Kiasman kokoelmiin kuuluvia mediataiteen teoksia.
- Musiikkiarkisto: Uuden AV-aineiston tuen (paketointi+vastaanotto) lisääminen käynnissä .

# Ilmoitusasiat

- Syksyn koulutukset
  - PAS-palvelun rajapinnat: 29.10
  - Paketointikomponentti: 26.11
  - Tarkemmat sisällöt & ilmoittautuminen myöhemmin
- iPres2019 kuulumiset



# Kokouksen päättäminen

- Muut asiat
- Seuraava kokous
  - 13.12.2019 klo 9:30 (CSC)
- Kokouksen päättäminen

