



ICT Solutions for  
Brilliant Minds



# PAS-palveluiden paketointikomponentikoulutus

16.4.2019



## Päivän ohjelma

- 10:00 Tervetuloa
- 10:10 PAS-palveluiden perusteet; Heikki Helin
- 10:30 Paketointikomponentin hyödyntäminen; Juha Lehtonen
- 12:00 Lounas (omakustanteinen)
- 13:00 Käytännön harjoituksia; Johan Kylander, Heikki Mikkolainen, Juho Kuisma
- 14:00 Kahvitauko
- 14:30 Käytännön harjoituksia
- 16:00 Koulutus päättyy

# Pitkääikaissäilytys

- Tarkoituksena säilyttää digitaiset aineistot käyttökelpoisina useiden vuosikymmenien tai satojen vuosien ajan
- Mahdollistaa digitaalisten aineistojen hyödyntäminen tuleville sukupolville
- Toteutuksen näkökulmasta korostaa jatkuvuussuunnittelua
  - toiminnan elinkaari ylittää kaikkien toteutuksen osakonaisuuksien käyttöajan
- Säilyttäminen on aktiivista toimintaa



**Nämä käy digitaaliselle historialle - Maailman ensimmäinen webbisivu on kadonnut**

Maailman ensimmäinen webbisivu on kadonnut. **Tim Berners-Lee** julkaisi tutkimuskeskus Cernissä ensimmäisen www-sivun vuonna 1990. Hän säilytti yhtä alkuaikojen sivun versiota levykkeellä, jotta sitä voisi tarvittaessa helposti demota. Jossain vaiheessa Berners-Lee sanoo kuitenkin luopuneensa levykkeen kopioinnista. Www-maailmaa ei myöskään ole suunniteltu tallentumaan - webbisivun keskimääräinen ikä muuttumattomana on 100 päivää.

Ilmoitetaan, että alkuperäistä sivua ei ole toistaiseksi löytynyt, mutta sitä yritetään palauttaa. Luontevaisti se on ollut eteenpäin ilvyky, joko Cernin tietokoneelta tai vanhoilta tallennuslaitteilta.

# Säilyttämisen palvelut ja PAS-ratkaisu

- CSC tuottaa OKM:lle keskitettyjä PAS-palveluita
  - kulttuuriperinnön säilyttämiseen (tuotannossa vuodesta 2015)
  - Tutkimusaineistojen säilyttämiseen (tavoitteena aloittaa säilytys vuoden 2019 aikana)
- Pohjautuvat yhteiseen bittitason säilytyksen ratkaisuun
- Taustalla OKM:n KDK- ja ATT-hankkeet, joissa laaditut arkkitehtuurit tunnistavat säilyttämisen tarpeet ja asemoivat PAS-palvelut
  - Kulttuuriperinnön säilyttämiseen on monilla organisaatioilla lakisääteinen velvollisuus
- Tavoitteena on mahdollistaa digitaalisten aineistojen käyttökelpoisuus, autenttisuus ja eheys kymmenien tai jopa satojen vuosien ajan
- CSC on ottanut säilytysvastuuun noin yli 300 teratavusta ja yli 1,3 miljoonasta tietopaketista (tilanne 15.4.2019)
- PAS-palvelut sisältyvät CSC:n ISO/IEC 27001:2013-auditoinnissa tarkasteltuihin palveluihin



# CSC auttaa organisaatioita säilyttämisen kaikissa vaiheissa

- Tämä tuki on merkittävä osa organisaatioille tarjottavaa palvelua
- Kokemus osoittanut, että organisaatiot tarvitsevat usein tukea käyttöönnotoissa
- Emme jätä organisaatioita yksin missään PAS-palveluihin liittyvässä asiassa

[pas-support@csc.fi](mailto:pas-support@csc.fi)



# Hyödyntäviä organisaatioita ohjaavat PAS-määritykset



- Näkyvin osa PAS-palvelusta hyödyntäville organisaatioille
- PAS-määritykset on tehty tiiviissä yhteistyössä hyödyntävien organisaatioiden kanssa
  - Ja näin myös jatketaan tulevaisuudessa

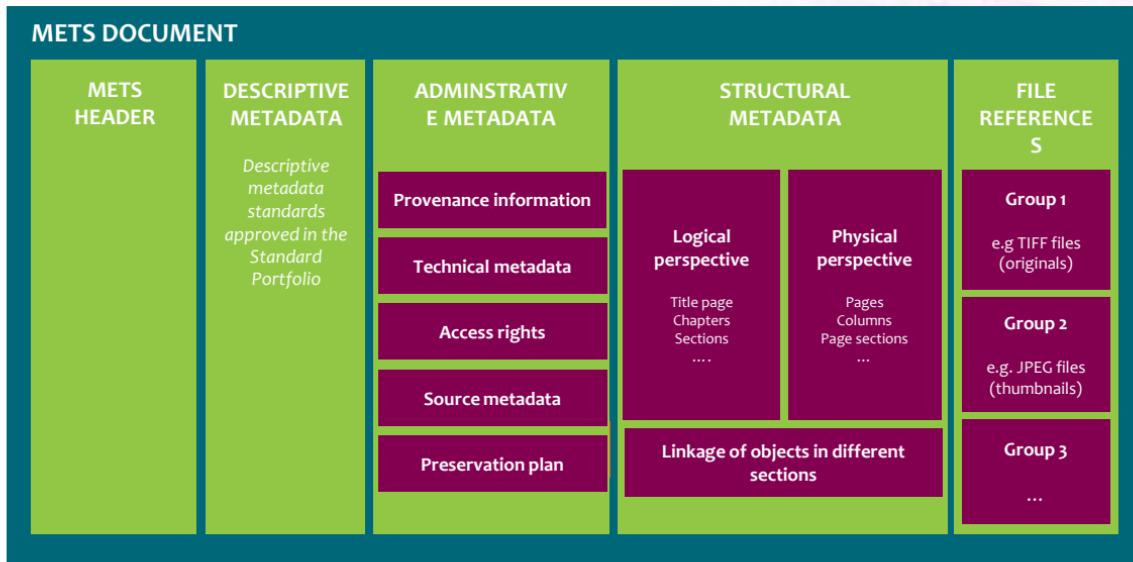
# Määritykset

Aineistojen ja niiden metatietojen paketointi pitkäaikaissäilytykseen

- Määrittelee
  - Pakolliset ja ehdollisesti pakolliset metatiedot
    - Kuvalevat, hallinnolliset ja rakenteelliset metatiedot
  - Siirtopaketin (SIP) ja jakelupaketin (DIP) teknisen rakenteen
- Perustuu METS ja PREMIS standardeihin
- Mahdollistaa täysin automaattisen työvuon PAS-palvelun vastaanotossa

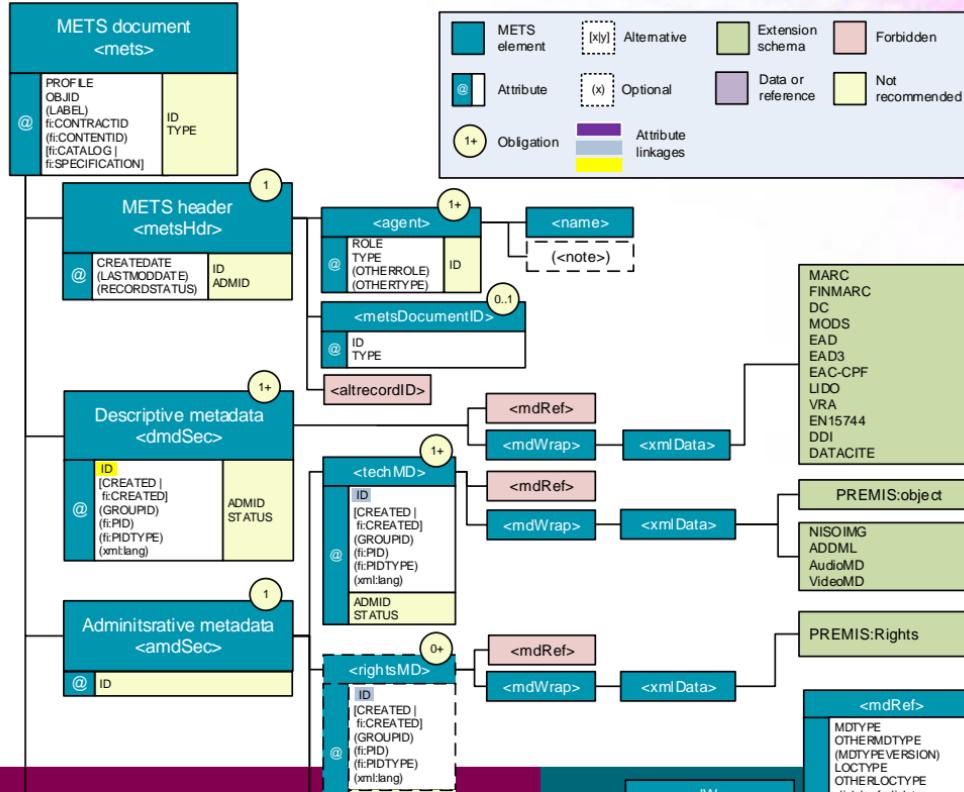
# Määritykset

## Aineistojen ja niiden metatietojen paketointi pitkäaikaissäilytykseen



# Määritykset

## Aineistojen ja niiden metatietojen paketointi pitkäaikaissäilytykseen



# Määritykset

## Säilytys- ja siirtokelvoiset tiedostomuodot

- Määrittelee
  - Yleiset vaatimukset tiedostomuodoille ja kriteeristön tiedostomuotojen arvioimiseen
  - Säilytyskelvoiset tiedostomuodot
  - Siirtokelvoiset tiedostomuodot
  - Kontrolloidun sanaston tiedostomuotojen ilmaisemiseen
  - Vaatimukset teknisille metatiedoille
- Säilytyskelvoiset tiedostomuodot
  - Tiedostomuodot, joiden arvioidaan olevan käytökkelvoisia vielä pitkään
- Siirtokelvoiset tiedostomuodot
  - Tiedostomuodot, joita arvioidaan olevan hyödyntävillä organisaatiolla runsaasti, mutta jotka eivät sovellu pitkäaikaiseen säilyttämiseen

# Määritykset

## PAS-palveluiden rajapinnat

- Määrittelee
  - Aineistojen siirtämisen PAS-palvelun vastaanottoon
  - Aineistojen haun ja hallinnan
  - Siirtopaketin tarkastusraportin
- Perustuu SFTP Ja REST -rajapointoihin

# Määritykset saatavilla

Myös englanniksi

- <http://digitalpreservation.fi/specifications>

- Aineistojen ja niiden metatietojen paketointi pitkääikaissäilytykseen (v.1.7.1)
- Säilytys- ja siirtokelvoiset tiedostomuodot (v.1.7.0)
- PAS-palveluiden rajapinnat (v.2.1.0)



- Ohjelmistioikoodia, skeemat & schematron säännöt saatavilla @GitHub

- <https://github.com/Digital-Preservation-Finland/>



# Paketointikomponentti



## Päivän ohjelmasta

- PAS-palveluiden kokonaisuus sisältää paljon erilaisia ohjelmistoja/komponentteja
  - Erityisesti Fairdata kokonaisuudessa: Metax, Qvain, IDA, ...
  - Näitä ei tänään käsitellä; ainoastaan paketointikomponenttia
- Paketointikomponentti != Paketointipalvelu



ICT Solutions for  
Brilliant Minds



## Paketoinkomponentin hyödyntäminen



# Sisällys

- Mikä on paketointikomponentti?
- Paketointikomponentin asennus
- Käyttö
  - Aineiston tuominen digitaalisiksi objekteiksi
  - Tiedostomuotokohtaisten teknisten metatietojen keruu
  - Tapahtumahistorian luonti
  - Kuvailavan metatiedon tuonti
  - Keruu tiedostoluetteloksi ja rakennekartaksi
  - Luodun ja kerätyn metatiedon koostaminen siirtopaketiksi

# Mikä on paketointikomponentti?



## Paketointikomponentti (Pre-Ingest Tool)

- Työkalu kansallisten PAS-määritysten mukaisten siirtopakettien muodostamiseksi
- Tavoitteena on helpottaa hyödyntävien organisaatioiden tuotantojärjestelmien integraatiota PAS-palveluun
- Haaste: Hyödyntävien organisaatioiden järjestelmät ovat kuitenkin hyvin erilaisia
- Mahdollistetaan myös suora hyödyntäminen, jos tuotantojärjestelmään integroiminen on tarpeeseen nähdyn kohtuutonta

# Paketointikomponentti (Pre-Ingest Tool)

- Joustava, modulaarinen ja muokattava
  - Jaetaan paketointi selkeisiin osakokonaisuuksiin, skripteihin
  - Pidetään skriptit mahdollisimman itsenäisinä ja annetaan tiedot argumentteina
  - Mahdollistaa esim. sen, että hyödynnetään esim. vain tiettyjä osakokonaisuuksia paketointikomponentista, ja toiset osat toteutetaan itse
  - Haluamiaan osakokonaisuuksia voi halutessaan täydentää tekemällä omia toteutuksia
- Saatavilla GitHubista:
  - <https://github.com/Digital-Preservation-Finland>
  - Lisenssi: GPLv3

# Paketoinnin vaiheet



# Paketoinnin vaiheet



## Tekninen metatieto

Perusmetatiedot, myös tiedostomuotovalidointi



## Kuvaileva metatieto

Valmiin XML-muotoisen kuvailumetatiedon tuonti



## Tiedostomuotokohtainen tekninen metatieto

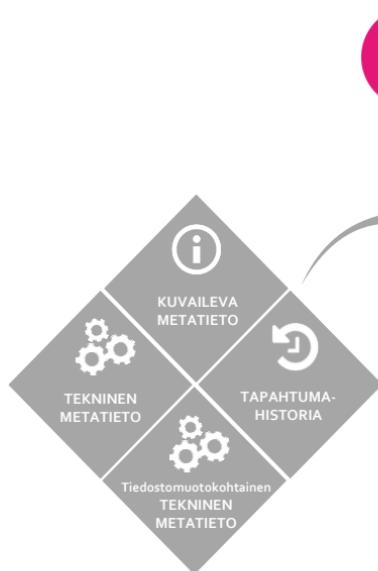
Kuva, ääni, video, rakenteellinen teksti



## Tapahtumahistoria

Aineiston syntyhistoria ja muut aineistoon kohdistuneet tapahtumat

# Paketoinnin vaiheet



## Tiedosto-osio ja rakennekartta

Tiedostoluettelo, rakenne, sekä metatiedon ja tiedostojen linkitykset



## METS-dokumentti

Koostetaan aiempien askelien pohjalta



## Allekirjoitus & paketointi

Allekirjoitetaan ja paketoidaan lähetettäväksi PAS-palveluun

## Paketointikomponentin hyödyntäminen

- Paketointikomponenttia on hyödynnetty useissa PAS-käyttöönnotoissa
- Esim. Fairdata PAS-palvelu hyödyntää paketointikomponenttia
- Paketointikomponentti ei ole “valmis”:
  - Kaikkia määritysten mukaisia tiedostomuotoja ja nk. natiivimuotoja ei vielä tueta
  - Määritykset muuttuvat vuosikellon mukaisella sykllä, ja paketointikomponentti mukana
  - Kehitystyötä tehdään PAS-käyttöönnottojen edetessä hyödyntävien organisaatioiden tarpeiden mukaisesti
- Määritysten mukaisia paketteja ei kannata alkaa tehdä, jos PAS-käyttöönotto ei ole ajankohtainen
  - Suositeltavaa kuitenkin on, että metatiedot ovat saatavilla (mielellään koneellisesti helppolukuisessa muodossa), ja että tiedostot ovat PAS-määritysten mukaisissa muodoissa.
  - Haasteet voidaan PAS-käyttöönnoton käynnistyessä ratkaista yhdessä.

# Paketointikomponentin asennus



## Paketointikomponentin asennus

- Paketointikomponenttia hyödynnetään aktiivisesti CentOS-ympäristössä.
  - Paketointikomponentin saanee samaan tapaan toimimaan myös muissa Linux-käyttöjärjestelmissä, esim. Ubuntu.
  - Paketointikomponentti ei sovella käytettäväksi Microsoft Windows -ympäristössä, vaan sen päälle tarvitaan vähintään Linux-virtuaaliympäristö.
- Asennusohjeissa CentOS-ympäristöön asennetaan virtuaaliympäristö virtualenv-työkalulla, jotta ympäristö saadaan vakioiduksi ympäristöksi.
  - Tällöin muut käyttöjärjestelmään asennetut paketit eivät vaikuta haitallisesti toimintaan.
  - Asennusta voi toki yrittää ilman virtualenv-ympäristöäkin.

## Paketointikomponentin asennus

- Ennen asennusta kannattaa varmistua, että Linux-ympäristöstä löytyy openssl-devel ja gcc-kirjastot, muussa tapauksessa digitaalisen allekirjoituksen luonti ei toimi.
- Asennus suoritetaan seuraavilla komennolla:
  - *sudo yum install python-virtualenv* (asennetaan virtuaaliympäristö-työkalu)
  - *virtualenv venv* (luodaan uusi ympäristö)
  - *source venv/bin/activate* (aktivoidaan luotu ympäristö)
  - *pip install --upgrade pip* (päivitetään asennussovellus)
  - *pip install -r requirements\_github.txt* (asennetaan luotuun ympäristöön)
  - *pip install .*
- Luodun virtuaaliympäristön voi sulkea komennolla “*deactivate*” ja aktivoida taas uudelleen edellä olevalla *source*-komennolla.

## Paketointikomponentin asennus

- Lisäksi tarvitaan seuraavat, jos EI validoida tiedostomuotojen oikeellisuutta:
  - Kuvatiedostoille: ImageMagick
  - Ääni- ja videotiedostoille: Medialinfo
  - Office-tiedostoille: Linuxin file-komennosta vähintään versio 5.30
- Tiedostomuodon oikeellisuuden validointi tarvitsee kaikkien edellisten lisäksi:
  - Tekstimuotoisille tiedostoille: JHove, v.Nu, iso-schematron-xslt1
  - Ääni- ja videotiedostoille (muu kuin WAV): FFMpeg
  - WAV-tiedostoille: JHove
  - Kuvatiedostoille: JHove
  - PDF-tiedostoille: JHove, veraPDF, GhostScript
  - Office-tiedostoille: LibreOffice
  - SPSS-tiedostoille: pspp
  - WARC-tiedostoille: warc-tools

# Siirtopaketin muodostaminen paketointikomponentilla



A hand is shown writing mathematical formulas on a transparent surface, likely a whiteboard or glass. The formulas include:
$$S(x) = \sum_{t=2}^n y_t = \frac{n-2}{n-1} y_{x-1}$$
$$y_1 = \frac{t=2}{n-1}; y_2 = \frac{t=2}{n-1}; \dots$$
$$\varepsilon_{ex} = \frac{dQ_{ex}}{de} \cdot \frac{e}{Q_{ex}}; \varepsilon_{im} = \frac{dQ_{im}}{de} \cdot \frac{e}{Q_{im}} \cdot \sqrt{\frac{g-3}{8/5}}$$
$$NE(e) = Q_{ex}(e) - eQ_{im}(e),$$
$$\Delta NE = \frac{dQ_{ex}}{de} \Delta e - e \frac{dQ_{im}}{de} \Delta e - eQ_{im}, (4)$$
$$B(a, b) = \int_0^1 (1-x)^{b-1} x^{a-1} dx = \Gamma(a) \Gamma(b) / \Gamma(a+b)$$
$$\beta_{yx} = r \frac{1}{56} \sqrt{7(-5 + \sqrt{1 + 48x})}$$
Surrounding the formulas are various mathematical diagrams, including a graph of a function  $S(x)$ , a bar chart, a pie chart, a gear diagram, and a graph of a function  $f(x)$ .



# Ohjeet

- Ohjelmistossa on README.rst, jossa on yleisiä käyttöohjeita.
- Jokaisessa skriptissä on --help-argumentti, jota käyttämällä saa lisätietoa kyseisestä skriptistä.

```
[(venv) [vagrant@dpres dpres-siptools]$ create-mix --help
Usage: create-mix [OPTIONS] FILENAME

      Write MIX metadata for an image file.

  FILENAME: Relative path to the file from current directory or from
            --base_path.

Options:
  --workspace <WORKSPACE PATH>    Workspace directory for the metadata files.
                                    Defaults to ./workspace/
  --base_path <BASE PATH>          Source base path of digital objects. If used,
                                    give path to the file in relation to this base
                                    path.
  --help                           Show this message and exit.
```

# Aineiston tuominen digitaalisiksi objekteiksi



## Aineiston tuominen digitaalisiksi objekteiksi

- Tuotavalle aineistolle luodaan perusmetatiedot. Jokaiselle tiedostolle joko annetaan parametrina tai muussa tapauksessa luodaan automaattisesti seuraavat metatiedot:
  - Tiedoston tunniste
  - Tiedostomuoto ja sen versio
  - Tarkistussumma (tämä on syytä tallentaa mahdollisimman aikaisessa vaiheessa)
  - Syntyaika
  - Tekstitiedostoille merkistökoodaus tiedostomuodon yhteyteen
- Parametrina voi antaa edellisten lisäksi soveltuvasti:
  - Jonkin formaattirekisterin (esim. PRONOM) koodin
  - Tiedoston järjestysnumeron.
- Tässä vaiheessa myös validoidaan tiedostomuodon oikeellisuus, ellei sitä ohiteta

# Tiedostojen tuominen digitaalisiksi objekteiksi

*import-object [OPTIONS] [FILEPATHS] ...*

*FILEPATHS*: Tiedosto/hakemisto, optioista riippuen voi antaa yhden tai useita

*--workspace <polku>*: Työtilan hakemisto

*--base\_path <polku>*: Aineiston juurihakemisto

*--skip\_wellformed\_check*: Tiedostomuotojen oikeellisuuden validoinnin ohitus

*--file\_format <mimetyyppi> <versio>*: Tiedoston MIME-tyyppi ja muodon versio

*--format\_registry <nimi> <koodi>*: Formaattirekisterin nimi ja formaatin koodi

# Tiedostojen tuominen digitaalisiksi objekteiksi

--*identifier* <*tyyppi*> <*tunniste*>: Tunnisten tyyppi (esim. urn) ja arvo

--*checksum* <*algoritmi*> <*summa*>: Tarkistussumman algoritmi (esim. md5) ja arvo

--*date\_created* <*aika*>: Tarkka tai epätarkka tiedoston syntymäaika

--*charset* <*merkistö*>: Tekstitiedostojen merkistökoodaus

--*order* <*numero*>: Tiedoston järjestysnumero

# Tiedostojen tuominen digitaalisiksi objekteiksi

*import-object tests/data/test\_import.pdf --workspace ./workspace/*

```
(.venv) [vagrant@dpres dpres-siptools]$ ls -l workspace/
total 12
-rw-rw-r-- 1 vagrant vagrant 1842 Apr 14 04:11 1fa65dfafdbf04b907d99e22f361602b-PREMIS%3AOBJECT-amd.xml
-rw-rw-r-- 1 vagrant vagrant 126 Apr 14 04:11 1fa65dfafdbf04b907d99e22f361602b-scraper.pkl
-rw-rw-r-- 1 vagrant vagrant 171 Apr 14 04:11 amd-references.xml
```

```
(.venv) [vagrant@dpres dpres-siptools]$ cat workspace/amd-references.xml
<?xml version='1.0' encoding='UTF-8'?>
<amdReferences>
  <amdReference file="tests/data/test_import.pdf">_1fa65dfafdbf04b907d99e22f361602b</amdReference>
</amdReferences>
```

# Varsinainen tulos

```
(.venv) [vagrant@dpres dpres-siptools]$ cat workspace/1fa65dfafdbf04b907d99e22f361602b-PREMIS%3AOBJECT-amd.xml
<?xml version='1.0' encoding='UTF-8'?>
<mets:mets xmlns:mets="http://www.loc.gov/METS/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.loc.gov/METS/ http://www.loc.gov/standards/mets/mets.xsd" PROFILE="local" OBJID="db0ef2f2-9b81-4446-afa3-4d083724d364">
  <mets:amdSec>
    <mets:techMD ID="_1fa65dfafdbf04b907d99e22f361602b" CREATED="2019-04-14T01:11:46.213424">
      <mets:mdWrap MDTYPE="PREMIS:OBJECT" MDTYPEVERSION="2.3">
        <mets:xmlData>
          <premis:object xmlns:premis="info:lc/xmlns/premis-v2" xsi:type="premis:file">
            <premis:objectIdentifier>
              <premis:objectIdentifierType>UUID</premis:objectIdentifierType>
              <premis:objectIdentifierValue>c6078630-5a17-4f56-8010-e948107ea8f6</premis:objectIdentifierValue>
            </premis:objectIdentifier>
            <premis:objectCharacteristics>
              <premis:compositionLevel>0</premis:compositionLevel>
              <premis:fixity>
                <premis:messageDigestAlgorithm>MD5</premis:messageDigestAlgorithm>
                <premis:messageDigest>87bbc5ea526a46f7db43adbdfee71e3e</premis:messageDigest>
              </premis:fixity>
              <premis:format>
                <premis:formatDesignation>
                  <premis:formatName>application/pdf</premis:formatName>
                  <premis:formatVersion>1.4</premis:formatVersion>
                </premis:formatDesignation>
              </premis:format>
              <premis:creatingApplication>
                <premis:dateCreatedByApplication>2019-04-11T15:10:26.015218</premis:dateCreatedByApplication>
              </premis:creatingApplication>
            </premis:objectCharacteristics>
          </premis:object>
        </mets:xmlData>
      </mets:mdWrap>
    </mets:techMD>
  </mets:amdSec>
</mets:mets>
```

# Videosäiliöiden tuominen digitaalisiksi objekteiksi

- Videosäiliöt voivat sisältää useita video- ja audioraitoja
- Tällainen tiedosto annetaan *import-object*-skriptille normaalisti, jolloin se luo koko säiliölle ja sen sisältämille raidoille tekniset perusmetatiedot tarvittavine keskinäisine linkityksineen kerralla.
- Tällä hetkellä tuetaan:
  - MPEG-perhe, versiot 1, 2, ja 4 sekä näiden kombinaatiot
  - Quicktime-tiedostot, joissa DV-video- ja PCM-audioraitoja
- Tukea eri yhdistelmiille lisätään tarpeen mukaan PAS-käyttöönottojen edetessä

# Tiedostomuotokohtaisten teknisten metatietojen keruu



# Tiedostomuotokohtaisten teknisten metatietojen keruu

- Metatietoihin tarvitaan erilaisia tiedostomuotokohtaisia teknisiä metatietoja
  - Ks. tiedostomuotomääritys: <https://digitalpreservation.fi/specifications>
  - Esimerkiksi kuvatiedostoista tarvitaan kuvan leveys, korkeus, värikoordinaatisto jne.
- Paketointikomponentti osaa analysoida tiedostot ja kerätä pakolliset metatiedot
- Eri tyyppisille aineistoille on omat metatietoformaattinsa:
  - Kuvatiedostoille (MIX-metatieto)
  - Äänitiedostoille (AudioMD-metatieto)
  - Videotiedostoille (VideoMD-metatieto)
  - CSV-tiedostoille (ADDML-metatieto)

## Kuvatiedostojen teknisten metatietojen keruu

*create-mix [OPTIONS] FILENAME*

*FILENAME*: Kuvatiedosto

--workspace <polku>: Työtilan hakemisto

--base\_path <polku>: Aineiston juurihakemisto

Esimerkki:

*create-mix images/image.tif --workspace ./workspace*

## Äänitiedostojen teknisten metatietojen keruu

*create-audiomd [OPTIONS] FILENAME*

*FILENAME*: Äänitiedosto

--workspace <polku>: Työtilan hakemisto

--base\_path <polku>: Aineiston juurihakemisto

Esimerkki:

*create-audiomd audio/sound.wav --workspace ./workspace*

## Videotiedostojen teknisten metatietojen keruu

*create-videomd [OPTIONS] FILENAME*

*FILENAME*: Videotiedosto

--workspace <polku>: Työtilan hakemisto

--base\_path <polku>: Aineiston juurihakemisto

Esimerkki:

*create-videomd video/video.m1v --workspace ./workspace*

## Videosäiliöiden teknisten metatietojen keruu

Videosäiliölle video- ja audioraidat tulee kerätä erikseen:

*create-videomd video.mp4 --workspace ./workspace*

- Kerää kaikkien videoraitojen metatiedot video.mp4-tiedostosta
- Jokaisesta raidasta tulee oma VideoMD-metatietolahkonsa

*create-audiomd video.mp4 --workspace ./workspace*

- Kerää kaikkien audioraitojen metatiedot video.mp4-tiedostosta
- Jokaisesta raidasta tulee oma AudioMD-metatietolahkonsa

# Rakenteellisen tekstin (CSV) teknisten metatietojen keruu

*create-addml [OPTIONS] FILENAME*

*FILENAME*: CSV-tiedosto

--workspace <polku>: Työtilan hakemisto

--base\_path <polku>: Aineiston juurihakemisto

--header: Käytetään, jos CSV-tiedostossa on otsikkorivi

--charset <merkistö>: CSV-tiedoston merkistökoodaus (pakollinen)

--delim <merkki>: Kenttien eroinmerkki CSV-tiedostossa (pakollinen)

--sep <merkki>: Tietueiden eroinmerkki CSV-tiedostossa (pakollinen)

--quot <merkki>: Lainausmerkit CSV-tiedostossa (pakollinen)

# Tapahtumahistorian luonti



## Tapahtumahistorian luonti

- Metatiedoissa on pakollista kuvata aineiston tapahtumahistoria.
  - Vähintään aineiston syntyhistoria on kuvattava.
- Tapahtumahistoria lisää todistusvoimaisuutta, kun kirjattujen tapahtumien avulla voidaan esimerkiksi selvittää, minkälaisia aineistoja muuttavia toimenpiteitä aineistolle on tehty.
- Digitoinnissa on tärkeää tietää tekijän ja ajan lisäksi esimerkiksi skannerin tai kameran merkki, malli, asetukset sekä kaikki ne vaiheet, jotka digitoinnissa on tehty. Tällaiset tapahtumat on kannattavaa tallentaa mahdollisimman aikaisessa vaiheessa, koska niiden selvittäminen voi myöhemmin olla varsin vaikeaa.

## Tapahtumahistorian luonti

*premis-event [OPTIONS] EVENT\_TYPE EVENT\_DATETIME*

*EVENT\_TYPE*: Tapahtumatyyppi

*EVENT\_DATETIME*: Tapahtuman aikaleima

--workspace <polku>: Työtilan hakemisto

--event\_target <polku>: Aineiston hakemisto/tiedosto, johon tapahtuma kohdistuu

--event\_detail <kuvaus>: Lyhyt kuvaus tapahtumasta (pakollinen)

## Tapahtumahistorian luonti

--event\_outcome <tulos>: Tapahtuman lopputulos (pakollinen)

--event\_outcome\_detail <tuloksen kuvaus>: Tarkempi tuloksen kuvaus

--agent\_name <nimi>: Agentin nimi

--agent\_type <tyyppi>: Agentin tyyppi

Esimerkki:

```
premis-event creation '2016-10-13T12:30:55' --workspace ./workspace \
--event_target './data' --event_detail 'Creation of data' \
--event_outcome 'success' --event_outcome_detail 'Outcome detail' \
--agent_name 'Demo Application' --agent_type 'software'
```

# Kuvailevan metatiedon tuonti



## Kuvailevan metatiedon tuonti

- Kuvaileva metatieto kuvilee mm. mitä aineisto sisältää ja siten asettaa sen kontekstiin.
- Kuvailevaa metatietoa ei voida yleisesti tuottaa automaattisesti, ja siten paketointikomponentti ei osaa sitä luoda.
- Siirtopakettiin tarvitaan vähintään yksi kuvaileva metatietokokonaisuus jossakin standardisalkussa esitetyssä muodossa.
- Paketointikomponentilla olemassa olevan XML-muotoisen metatiedon voi liittää paketoitavan aineiston osaksi.

## Kuvailevan metatiedon tuonti

*import-description [OPTIONS] DMDLOCATION*

*DMDLOCATION*: Polku XML-muotoiseen kuvailevaan metatietoon

*--workspace <polku>*: Työtilan hakemisto

*--dmdsec\_target <polku>*: aineistohakemisto, johon kuvaileva metatieto kohdistuu

*--remove\_root*: Poistaa juurielementin tuodusta kuvailevasta metatiedosta

- Tämä voi olla tarpeen esim. Dublin Core -muotoisessa metatiedossa, jossa ei varsinaisesti ole juurielementtiä, vaan metatieto on yleensä standardiin kuulumattoman elementin sisällä.

- Esimerkki:

```
import-description 'metadata/dc_description.xml' --workspace ./workspace  
--dmdsec_target 'tests/data/' --remove_root
```

# Keruu tiedostoluetteloksi ja rakennekartaksi



## Keruu tiedostoluetteloksi ja rakennekartaksi

- Määritysissä on pakollisina metatietoina tiedosto-osio ja rakennekartta. Nämä linkittävät tiedostot ja edellä luodut metatieto-osiot yhteen.
- Paketointikomponentissa tämä toteutuu automaattisesti omalla skriptillä
  - Komponentti ei kuitenkaan osaa luoda monimutkaisia rakennekartoja
  - Tälle osa-alueelle tarvittaisiin käyttötapauksia esim. PAS-käyttöönottojen kautta
- Rakennekartoja tarjotaan paketointikomponentissa tällä hetkellä kahdenlaisia:
  - Hakemistorakenteen mukainen fyysinen rakenne
  - EAD<sub>3</sub>-metatiedon mukainen looginen rakenne
- Muunlaisia rakennekartoja punnitaan ja lisätään PAS-käyttöönottojen edetessä

# Keruu tiedostoluetteloksi ja rakennekartaksi

*compile-structmap [OPTIONS]*

--workspace <polku>: Työtilan hakemisto

--structmap\_type <tyyppi>: Rakennekartan tyyppi

--root\_type <tyyppi>: Juurielementin tyyppi

--dmdsec\_loc <polku>: Kuvalevan metatiedon sijainti

- Käytetään, jos rakennekartta luodaan kuvalevan metatiedon perusteella.

- Tällä hetkellä tässä toiminnossa tuetaan vain EAD3-metatietoa.

- Esimerkki: *compile-structmap --workspace ./workspace*

# Luodun ja kerätyn metatiedon koostaminen siirtopaketiksi



A hand is shown writing various mathematical formulas and graphs on a transparent surface. The formulas include:
$$S(x) = \sum_{t=1}^n y_t = \frac{n-2}{n} \sum_{t=1}^n y_{t-1}$$
$$\bar{y}_1 = \frac{\sum_{t=2}^n y_t}{n-1}; \quad \bar{y}_2 = \frac{\sum_{t=2}^{n-1} y_{t-1}}{n-1}; \quad \dots$$
$$\varepsilon_{ex} = \frac{dQ_{ex}}{de} \cdot \frac{e}{Q_{ex}}; \quad \varepsilon_{in} = \frac{dQ_{in}}{de} \cdot \frac{e}{Q_{in}} \cdot \sqrt{8/5}$$
$$NE(e) = Q_{ex}(e) - eQ_{in}(e),$$
$$\Delta NE = \frac{dQ_{ex}}{de} \Delta e - e \frac{dQ_{in}}{de} \Delta e - eQ_{in}, \quad (4)$$
$$B(a, b) = \int_0^1 (1-x)^{b-1} x^{a-1} dx = \Gamma(a) \Gamma(b) / \Gamma(a+b)$$
$$\beta_{yx} = r \frac{1}{56} \sqrt{7(-5 + \sqrt{1 + 4x})}$$

Integrate  $f(t)/(\lambda^t)$

$\frac{8}{105} (t + \sqrt{t})$

Graphs include a sine wave, a bell curve, a bar chart, and a scatter plot with a regression line.



## Luodun ja kerätyn metatiedon koostaminen siirtopaketiksi

- Edellä olevilla askelilla on nyt luotu tarvittavat metatiedot ja lueteltu aineiston tiedostot, mutta tieto on vielä hajallaan. Tässä vaiheessa aineisto kootaan yhteen.
- Tämä käsittää seuraavat vaiheet:
  - METS-dokumentin luonti aiemmin luodun ja kerätyn metatiedon perusteella
  - Siirtopaketin sähköinen allekirjoitus
  - Siirtopaketin kääriminen TAR-paketiksi.

# Metatietojen keruu METS-dokumentiksi

*compile-mets [OPTIONS] METS\_PROFILE ORGANIZATION\_NAME CONTRACTID*

*METS\_PROFILE*: METS-profiili

*ORGANIZATION\_NAME*: Siirtopaketin luoneen organisaation nimi

*CONTRACTID*: PAS-palvelun antama sopimustunniste

--*workspace* <*polku*>: Työtilan hakemisto

--*base\_path* <*polku*>: Aineiston juurihakemisto

--*objid* <*tunniste*>: Siirtopaketin tunniste

# Metatietojen keruu METS-dokumentiksi

- label <nimi>: Siirtopaketin lyhyt kuvaus
- contentid <tunniste>: Aineistokokonaisuuden tunniste
- create\_date <aikaleima>: Siirtopaketin syntyaika
- last\_moddate <aikaleima>: Siirtopaketin viimeisin muokkausaika
- clean: Poistetaan väliaikaistiedostot
- copy\_files: Kopioidaan paketoitava aineisto työtilaan (workspace)

## Siirtopaketin allekirjoitus

*sign-mets [OPTIONS] SIGN\_KEY*

*SIGN\_KEY*: Privaattiavain, jolla aineisto allekirjoitetaan

--workspace <polku>: Työtilan hakemisto

- Avaimien luontiin annetaan erilliset ohjeet PAS-palvelun käyttöönnotossa:
  - <https://digitalpreservation.fi/specifications/ohjeita-kayttoonotosta>
- Lisätietoja allekirjoituksesta mm. paketoinnin määritysdokumentissa:
  - <https://digitalpreservation.fi/specifications>

## Siirtopaketin kääriminen TAR-tiedostoksi

*compress [OPTIONS] DIR\_TO\_TAR*

*DIR\_TO\_TAR*: Hakemisto, joka käärítäään.

*--tar\_filename <tiedosto>*: Luotavan TAR-tiedoston nimi.

- TAR-käärimisen voi toki tehdä myös esim. Linuxin tar-komennolla tämä sijasta, tai sen voi pakata ZIP-tiedostoon.
- TAR-tiedosto (tai ZIP-tiedosto) siirretään siirtopakettina PAS-palveluun.

## Koostaminen: Esimerkki

- Koostetaan METS-dokumentti:

```
compile-mets ch 'CSC' '47265f3e-f423-4926-9f27-7bab08508732' \|  
--workspace ./workspace --copy_files --clean
```

- Allekirjoitetaan siirtopaketti:

```
sign-mets rsa-key.crt --workspace ./workspace
```

- Kääritään TAR-tiedostoon:

```
compress ./workspace --tar_filename sip.tar
```

# Kiitos!

digitalpreservation.fi  
pas-support@csc.fi

