CSC

ICT Solutions for Brilliant Minds

PAS-paketointimäärittelyt

PAS-seminaari 25.4.2023 / Juha Lehtonen

Yhteistyö mahdollistaa aineistojen pitkäkestoisen hyödyntämisen

Hyödyntäviä organisaatioita ohjaavat PAS-määrittelyt

- Yksi PAS-palvelun näkyvimpiä osia hyödyntäville organisaatioille
- PAS-määrittelyt on tehty tiiviissä yhteistyössä hyödyntävien organisaatioiden kanssa

Määrittelyt päivitetään vuosittain yhteistyössä

Palautteen kerääminen hyödyntäviltä organisaatioilta

- Epäviralliset palautteet ja virallisempi palautekysely

Päätökset muutoksista

- Ovatko muutokset välttämättömiä juuri nyt?
- Perustelut

Määrittelyiden päivittäminen yhteistyössä hyödyntävien organisaatioiden kanssa

- Hyväksyntä yhteistyöryhmässä

Uusien versioiden julkaiseminen

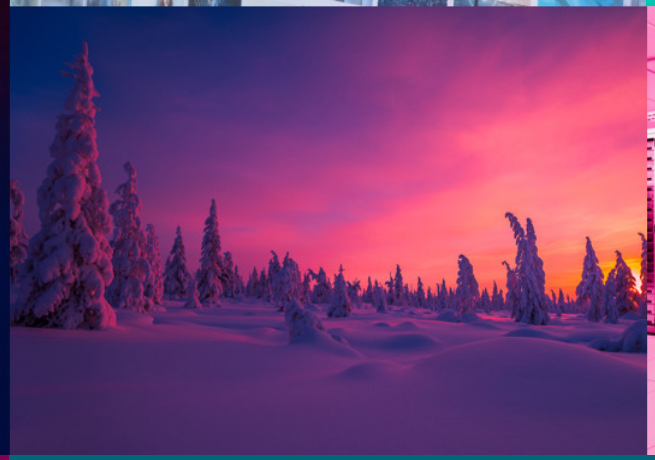
- XML-skeemat
- Muu dokumentaatio

Määrittelyt saatavilla

Myös englanniksi

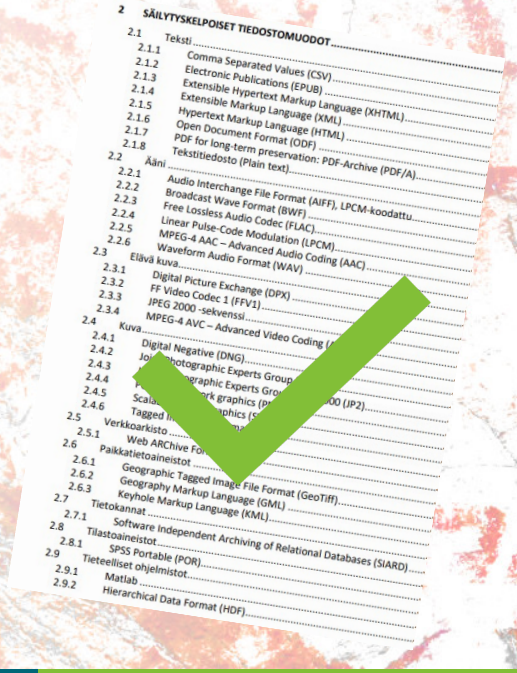
- http://digitalpreservation.fi/specifications
  - Aineistojen ja niiden metatietojen paketointi pitkäaikaissäilytykseen
    - URN:NBN:fi-fe2020100578093
  - Säilytys- ja siirtokelpoiset tiedostomuodot
    - URN:NBN:fi-fe2020100578095
  - PAS-palveluiden rajapinnat
    - URN:NBN:fi-fe2020100578097
- Vuosittaiset laaturaportit ja muita julkaisuja
- Työkalut @GitHub
  - https://github.com/Digital-Preservation-Finland/

Säilytys- ja siirtokelpoiset tiedostomuodot



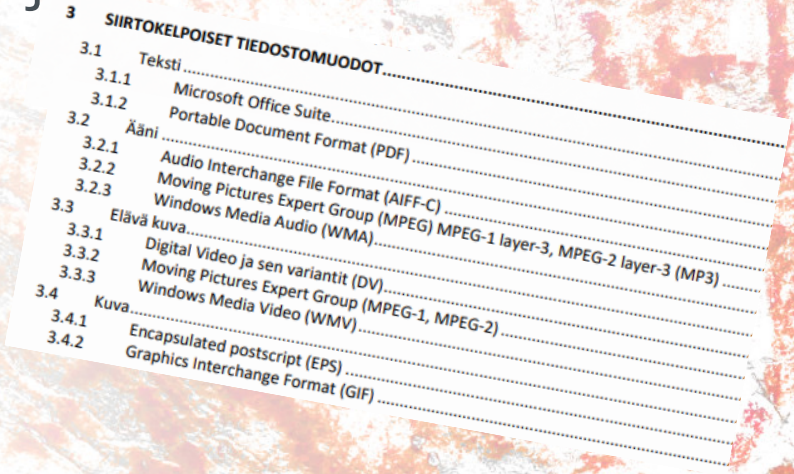
Säilytyskelpoiset tiedostomuodot

- Säilytyskelpoinen tiedostomuoto on sellainen, jonka tietosisällön säilyminen ja ymmärrettävyys voidaan taata pidemmällä aikavälillä
- PAS-palvelu vastaanottaa säilytyskelpoisia tiedostomuotoja siirtopaketeissa
- PAS-palvelu lupaa voivansa migroida hallitusti ja suunnitellusti
- Tiedostomuodot on jaettu aineistotyypittäin, esimerkiksi
  - Kuva: JPEG, PNG, SVG, TIFF …
  - Teksti: plain text, CSV, ODF, PDF …
  - Tietokannat: SIARD, POR
  - Ääni: BWF, FLAC …
  - Elävä kuva: FFV1, AVC …



Siirtokelpoiset tiedostomuodot

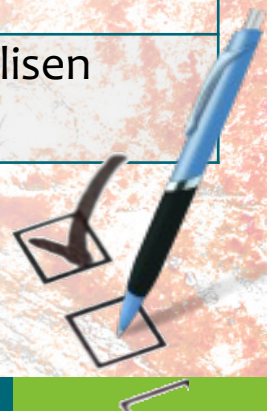
- Siirtokelpoiset tiedostomuodot ovat sellaisia, joiden tietosisällön säilymistä tai ymmärrettävyys ei voida taata pidemmällä aikavälillä
  - Mutta joita on kuitenkin huomattavasti organisaatioilla
  - Siirtokelpoisia tiedostomuotoja ei pitäisi enää tuottaa
- Käytännössä siirtokelpoinen on pakko migroida (joskus) ensiksi säilytyskelpoiseen, eli se on askeleen säilytyskelpoisen jäljessä
- Käytössä vastaava jako aineistotyypittäin kuin säilytyskelpoisissa muodoissa
  - Kaikilla aineistotyyppiryhmillä ei kuitenkaan ole siirtokelpoisia muotoja



Tiedostomuotojen arviointikriteeristö

Avoimuus	Kuinka helppoa tiedostomuodosta on saada tietoja?
Käyttö PAS-standardina	Missä määrin tiedostomuoto on muodollisesti hyväksytty pitkäaikaissäilytyksen välineeksi?
Vakaus/ yhteensopivuus	(a) Missä määrin tiedostomuoto on eteen- ja taaksepäin yhteensopiva? (b) Missä määrin tiedostomuoto on suojattu tiedoston korruptoitumista vastaan? (c) Kuinka usein tiedostomuodosta julkaistaan korvaavia versioita?
Riippuvuudet / yhteentoimivuus	Missä määrin tiedostomuoto on sidottu esimerkiksi tiettyyn laitteistoon tai ohjelmistoon?
Standardisuus	Missä määrin tiedostomuoto on käynyt läpi perusteellisen standardointiprosessin?

Arviointi: A, B tai C

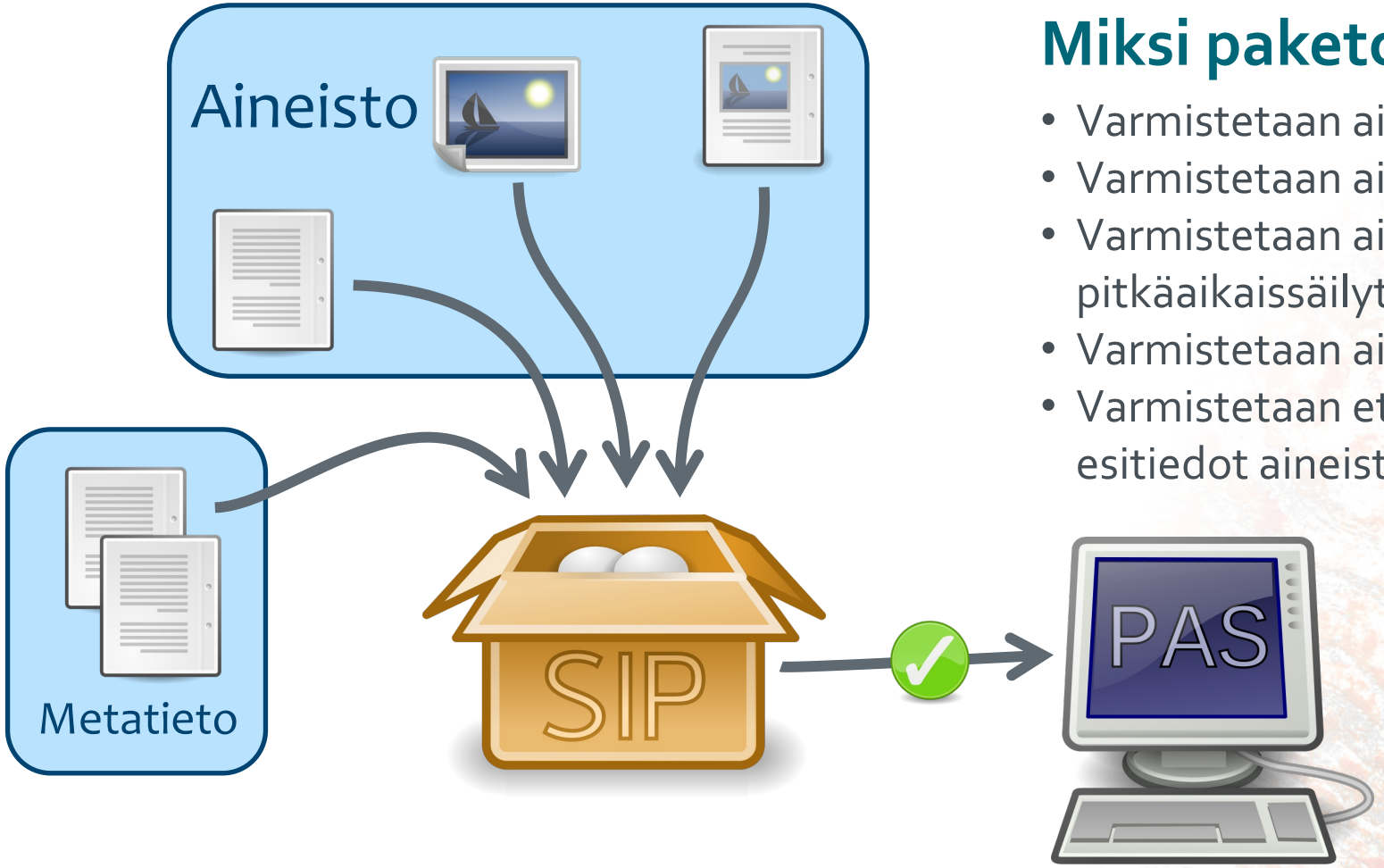


Aineistojen ja niiden metatietojen paketointi pitkäaikaissäilytykseen

Aineistojen ja niiden metatietojen paketointi pitkäaikaissäilytykseen

- Määrittelee pakolliset ja ehdollisesti pakolliset metatiedot
  - Hallinnolliset metatiedot
  - Rakenteelliset metatiedot
  - Tuettujen kuvailevan metatietomuotojen lista
- Määrittelee siirtopaketin (SIP) ja jakelupaketin (DIP) teknisen rakenteen
- Perustuu METS ja PREMIS XML-standardeihin
- Vaatimukset mahdollistavat aineistojen automaattisen validoinnin ja vastaanoton

Aineistojen paketointi pitkäaikaissäilytykseen



Miksi paketoidaan?

- Varmistetaan aineiston eheys
- Varmistetaan aineiston aitous
- Varmistetaan aineiston sopivuus pitkäaikaissäilytykseen
- Varmistetaan aineiston ymmärrettävyys
- Varmistetaan että PAS-palvelulla on riittävät esitiedot aineiston säilyttämiseksi

Aineistoja ei kannata paketoida etukäteen “varastoon”.

Paketoinnin vaihtoehdot

PAKETOINTIPALVELUN HYÖDYNTÄMINEN

Vain tutkimusaineistoille

A. IDA-säilytyspalvelu

✓Aineistot IDA:ssa

+ Paketointi automaattista
+ Ei vaadi omaa koodausosaamista
+ Ei vaadi XML-tuntemusta
– Joustamaton

B. Tiedostojen koontipalvelu

✓Aineistoa maltillisesti; yksinkertainen rakenne

+ Paketointi automaattista
+ Ei vaadi omaa koodausosaamista
+ Ei vaadi XML-tuntemusta
– Joustamaton

ORGANISAATIO PAKETOI ITSE

C. Paketointityökalun hyödyntäminen

✓Paljon aineistoa

+ Melko joustava
+ Ylläpidettävyys määrittelyiden muuttuessa
+ Vähentää tarvetta XML formaattien osaamiselle
– Jonkin verran omaa koodausosaamista
– Määrittelyiden ymmärtäminen

D. Organisaation räätälöimä paketointi

✓Paljon aineistoa; monimutkaisia tarpeita

+ Joustava ja skaalautuva
– (Runsaasti) omaa koodausosaamista
– Määrittelyiden syvällinen ymmärtäminen
– Ylläpidettävyys määrittelyiden muuttuessa

METS

- Metadata Encoding and Transmission Standard
- Kääremuoto
  - XML-standardi tiedostojen ja metatietojen niputtamiseksi yhteen
  - Kuvaileva, hallinnollinen, rakenteellinen metatieto
- METS XML -rakenteen määrittelee skeematiedosto, jota vastaan XML-dokumentin voi validoida
- Laajalti pitkäaikaissäilytyspiireissä käytössä
- METS Editorial Boardin ylläpitämä
- https://www.loc.gov/standards/mets/

PREMIS

- PREMIS Preservation Metadata
- Metatietomalli säilytysmetatiedolle
  - Määrittelee tietokentät (sanastot, sallitut arvot, kardinaliteetit) ja rakenteen
  - PREMIS XML-skeema on virallinen tuettu implementaatio tietomallista
- Laajalti pitkäaikaissäilytyspiireissä käytössä
- PREMIS Editorial Committeen ylläpitämä
- https://www.loc.gov/standards/premis/

METS-dokumentin rakenne

METS DOCUMENT

METS HEADER

DESCRIPTIVE METADATA

Descriptive metadata standards approved

ADMINSTRATIVE METADATA

- **Provenance information**
- **Technical metadata**
- **Access rights**
- **Source metadata**
- **Preservation plan**

STRUCTURAL METADATA

Logical perspective

Title page
Chapters
Sections
….

Physical perspective

Pages
Columns
Page sections
…

Linkage of objects in different sections

FILE REFERENCES

Group 1

e.g TIFF files (originals)

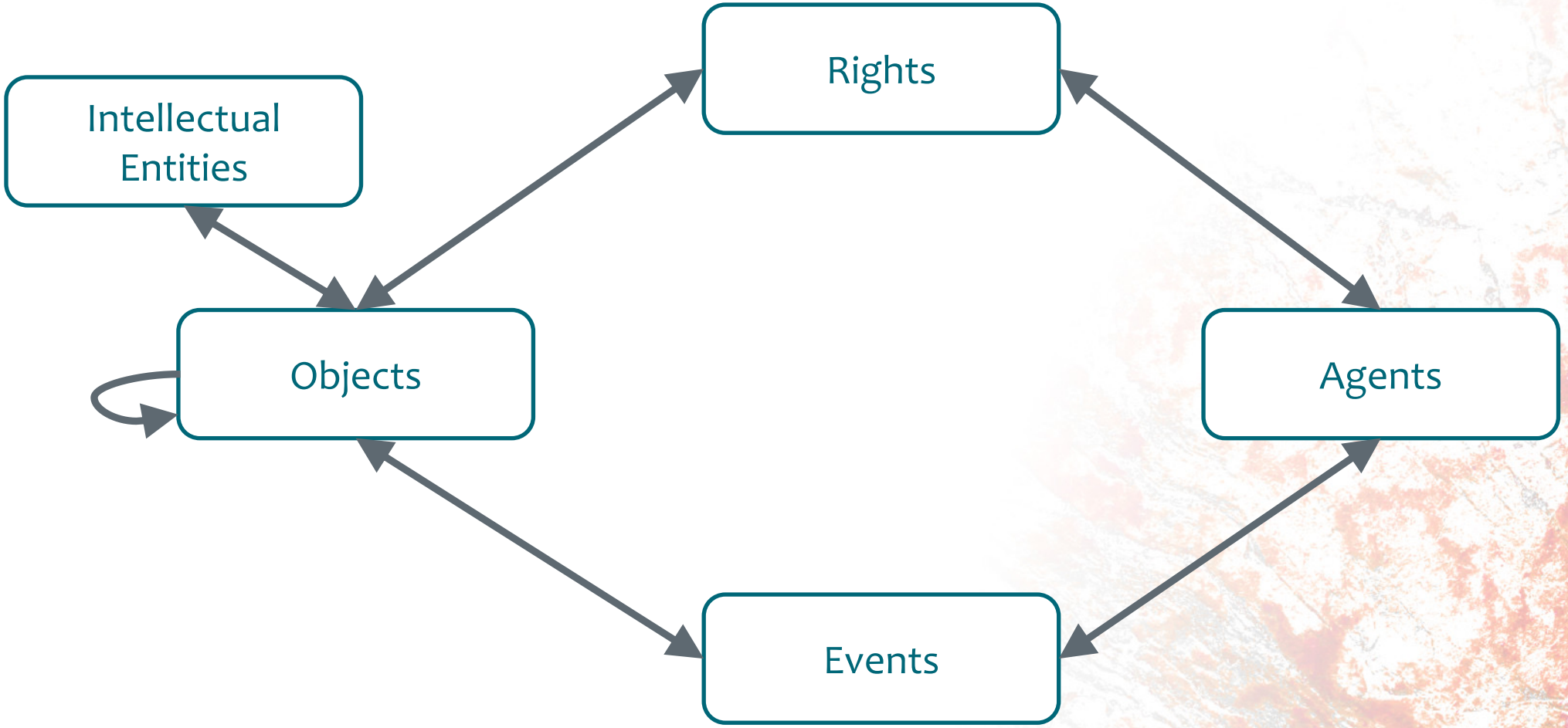
Group 2

e.g. JPEG files (thumbnails)

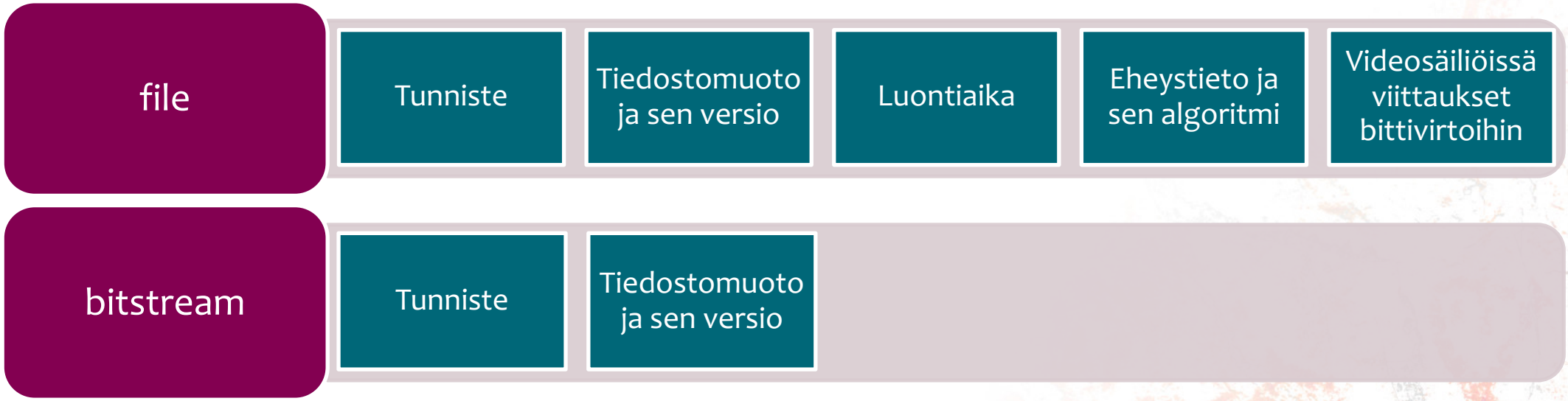
Group 3

…

PREMIS tietomalli



Pakolliset tekniset metatiedot – PREMIS Object



- Muut ehdollisesti pakolliset tekniset metatietomuodot:
  - MIX, VideoMD, AudioMD, ADDML
- Paketointityökalu osaa tuottaa digitalisten objektien pakolliset tekniset metatiedot automaattisesti

Yhteistyötä organisaatioiden kanssa nyt ja tulevaisuudessa

(myös muiden kuin asiakkaiden)

- PAS-palveluiden sähköpostilista
  - https://www.digitalpreservation.fi/2021-liity-pas-palveluiden-sahkopostilistalle
- Kuukausittaiset #PASKaffet
  - Joka kuukauden ensimmäinen perjantai klo 14:00-14:45
- #PASKlinikat
  - Säännöllisen epäsäännöllisesti
- Koulutustilaisuuksia enemmän tai vähemmän säännöllisesti
- twitter.com/dpres_FI
- Virallinen tukiosoite: pas-support@csc.fi
- https://www.digitalpreservation.fi
- https://www.fairdata.fi