



Aineistojen paketoinnin pilotit 2012

PAS-seminaari 2013
Kuisma Lehtonen

Pilottien aikataulu

Kevät 2012

SLS

Oulun yliopiston kirjasto

Kansalliskirjasto: Doria

Syksy 2012

Mikkelin digitointi- ja konservointikeskus

KAVA

Museo 2015 / Musketti

FSD

Arkistolaitos VAPA

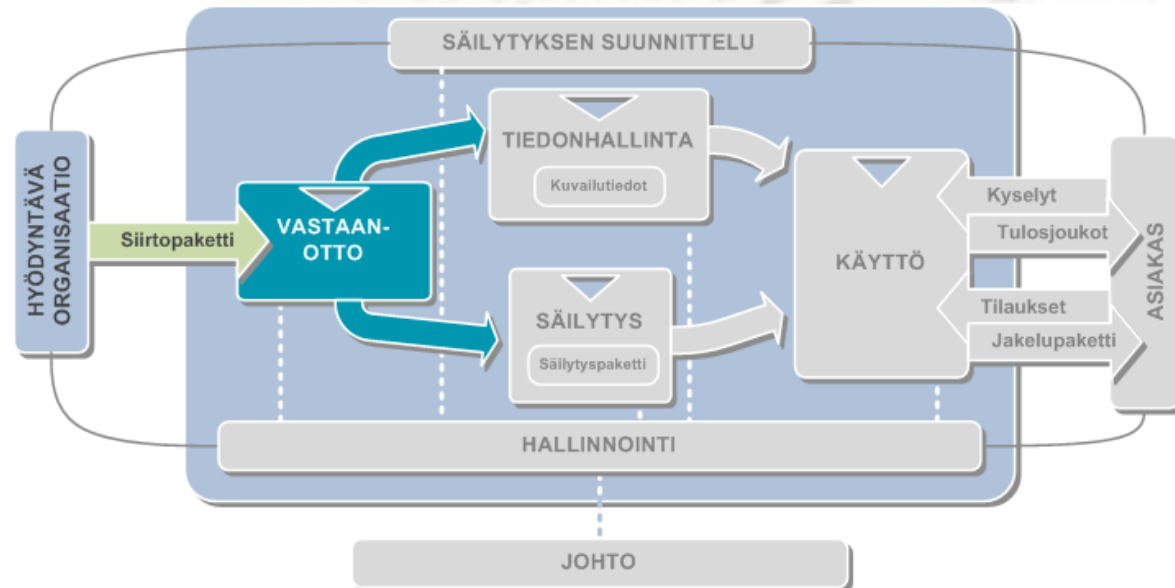
Arkistolaitos Digitaaliarkisto

Valtion taidemuseo

- Yhteensä 10 pilottia 8 organisaation kanssa
- Noin 80 aineistopakettia sisältäen yli 5000 tiedostoa

Tavoitteet

- kerätä hyödyntävien organisaatioiden kokemuksia aineistojen paketoinnista,
- tarkentaa KDK:n metatietomäärittäystä,
- konkretisoida keskitetyn paketointipalvelun suunnittelua sekä
- lisätä ja tunnistaa PAS-osaamista hyödyntävissä organisaatioissa



Pilottien painopisteet

Keväällä 2012

- Konkreettiset kokemukset KDK:ssa laaditun määrittelyn mukaisen aineiston käsittelystä
 - Määrittelyjen mukaisten siirtopakettien muodostaminen
 - Aineiston teknisen rakenteen ja metatietojen tarkastaminen laadittua määrittelyä vasten
- KDK:n pitkäaikaissäilytykseen siirrettäville aineistoille asetettavien vaatimusten täsmentäminen
- Vastaanottopalvelun prosessien tarkentaminen ja teknisten toteutusmahdollisuuksien kartoittaminen

Syksyllä 2012

- Pilottien laajentuminen aineistojen monimuotoisuuden, määrän ja koon suhteen
- Siirtopakettien tarkastamisen automatisointi
 - Tekninen tarkastaminen, vastaanottoprosessin seuranta, raportointi
- Tarkentuneiden määrittelyjen mukaan ottaminen

Siirtopakettien sisällöt

- SLS (27.2. – 30.3.2012)
 - Aineistomäärä: 3 SIP-pakettia, yhteensä alle 10 GT
 - Aineistotyypit: kuvia (TIFF, JPG) ja av-materiaalia (MP3, WAV)
- Oulun yliopiston kirjasto (26.3 - 25.5.2012)
 - Aineistomäärä: 30 SIP-pakettia, yhteensä reilut 100 MB
 - Aineistotyypit: kokotekstiaineistoa pdf –muodossa (Oulun yo:n julkaisuarkistosta)
- Kansalliskirjasto: Doria (7.5. – 8.6.2012)
 - Aineistomäärä: 30 SIP-pakettia, yhteensä reilut 100 MB
 - Aineistotyypit: Äänitteitä MP3–muodossa (nk. Equal Dreams -aineisto)
- Kansalliskirjasto: Mikkelin digitointi- ja konservointikeskus (27.8.-21.9.2012)
 - Aineistomäärä: 5 SIP-pakettia, yhteensä noin 1,5 GT
 - Aineistotyypit: JP2, JPG, PDF, TIFF, XML
 - Mets.xml-tiedoston koot: 26 – 95 Mt

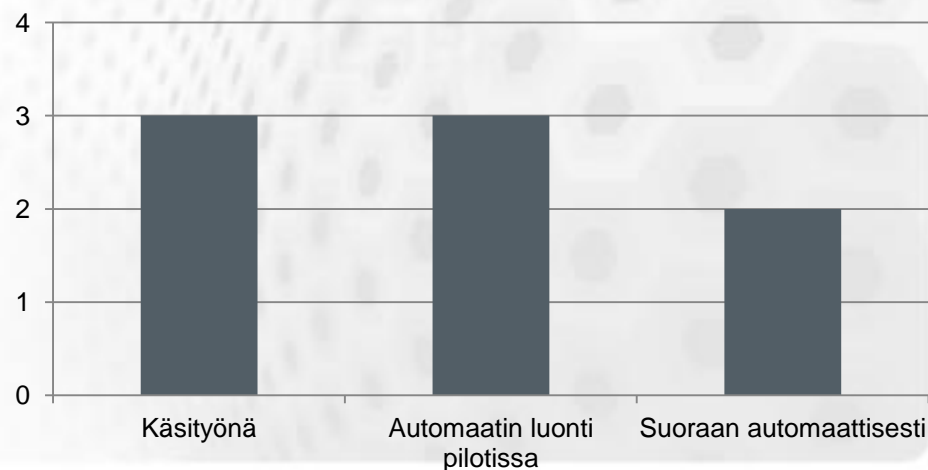
Siirtopakettien sisällöt

- KAVA (24.9.–31.10.2012)
 - Aineistomäärä: 1 SIP-paketti, alle 10 GT
 - Aineistotyytit: : 1 DCP-paketti, jossa elävä kuva ja audio MXF-muodossa, lisäksi XML-tiedostoja
 - DCP (Digital Cinema Package) on käytännössä tietyllä standardilla tavalla muodostettu hakemisto, joka voi sisältää audio-, video-, tekstitys- sekä aputiedostoja
- Museo 2015 / Musketti
 - Aineistomäärä: 2 SIP-pakettia, yhteensä alle 100 Mt
 - Aineistotyytit: kuvia (TIFF, JPG, XML)
- FSD (8.10.–23.11.2012)
 - Aineistomäärä: 8 SIP-pakettia, yhteensä alle 100 Mt
 - Aineistotyytit: Tutkimusaineistoa (PDF, XML, XHTML, JPEG)
- Arkistolaitos, VAPA (22.10.–30.11.2012) & Arkistolaitos, Digitaaliarkisto (12.11.–7.12.2012)
 - Aineistomäärä: 1 METS-dokumentti
- Valtion taidemuseo / Muusa (19.11.–14.12.2012)
 - Aineistomäärä: 1 SIP, yhteensä alle 100 Mt
 - Aineistotyytit: kuva (TIFF)

Siirtopakettien luonti

- ➊ Pakettien luontitavat eri organisaatioissa jakautuivat kolmeen ryhmään:
 - Siirtopaketin/-pakettien luonti käsityönä
 - Automaatin luonti tai jo olemassa olevan automaatin kehittäminen pilotin aikana
 - Siirtopaketin luonti suoraan (ilman kehitystyötä)

Paketin luonti [organisaatioita]



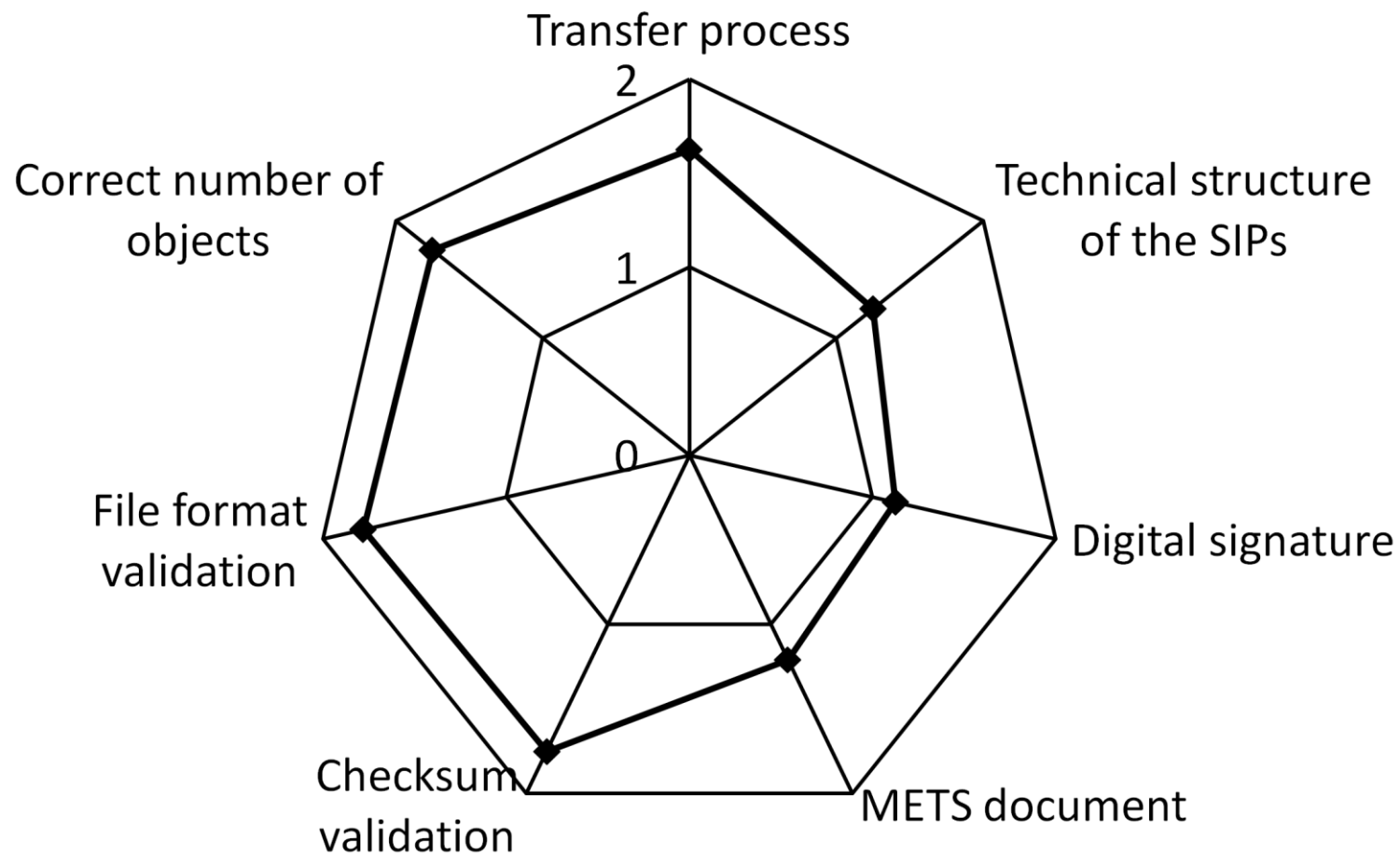
Siirtopakettien tarkastaminen

- Siirtopakettien tarkastus jaettiin eri osioihin:
 - Siirto komentorivityökalulla
 - Siirtopaketin tekninen rakenne
 - Sähköinen allekirjoitus
 - METS-dokumentin oikeellisuus
 - Tiedostojen eheystiedot
 - Tiedostoformaattien oikeellisuus
 - Tiedostojen oikea määrä

Siirtopakettien pisteytys

- ➊ Aineistojen tarkastuksen eri osioiden tulokset pisteytettiin asteikolla 0..2
 - 2 = (Lähes) virheetön
 - » Ei virheitä tai hyvin yksittäinen kokonaisuuden kannalta mitätön virhe
 - 1 = Selvästi puutteellinen
 - » Sisältää selkeitä merkittäviä puutteita tai merkittävän määrän pieniä puutteita
 - 0 = Täysin puutteellinen
 - » Puuttuu kokonaan tai ei juurikaan noudata ohjeistusta

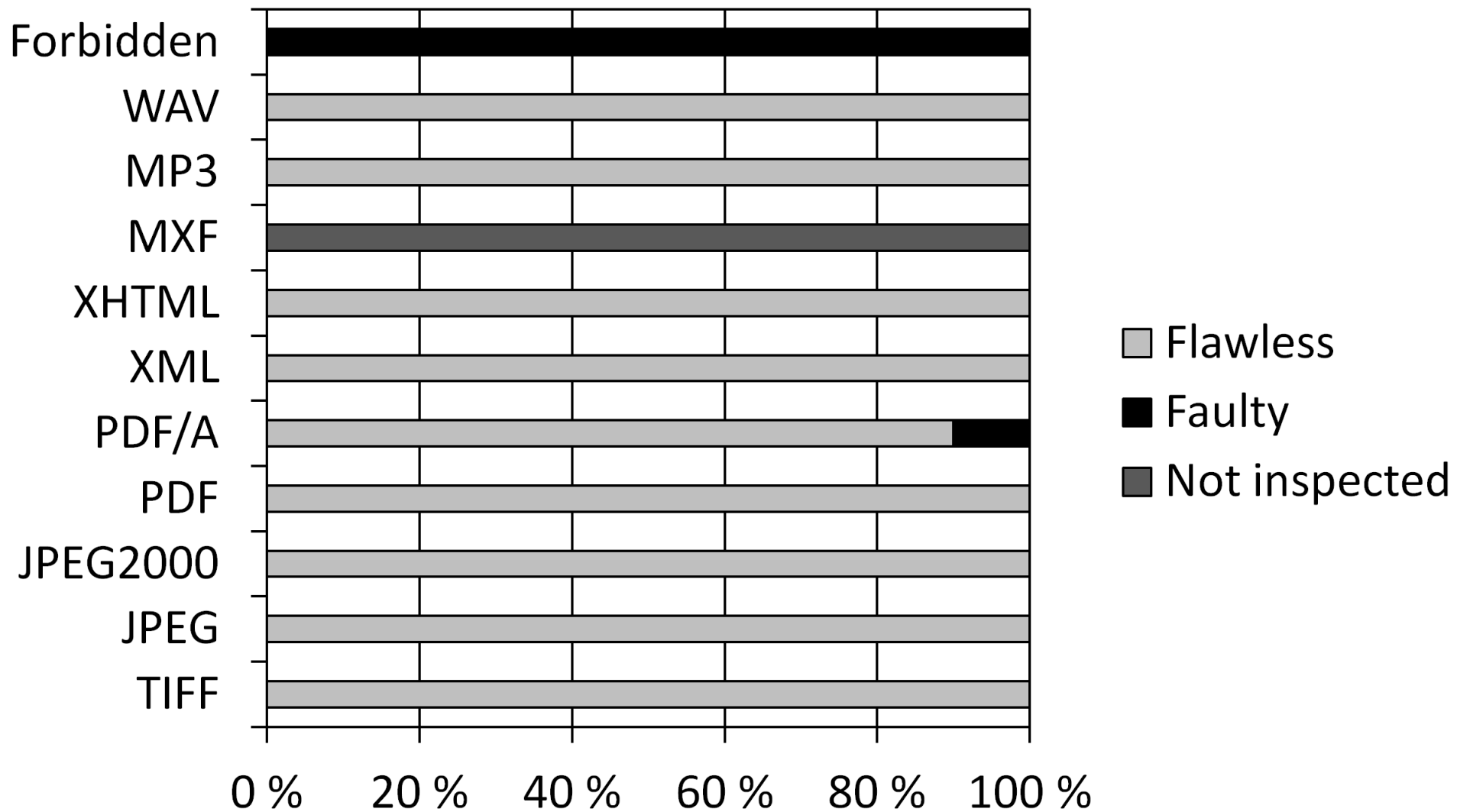
Siirtopakettien pisteytyksen keskiarvot



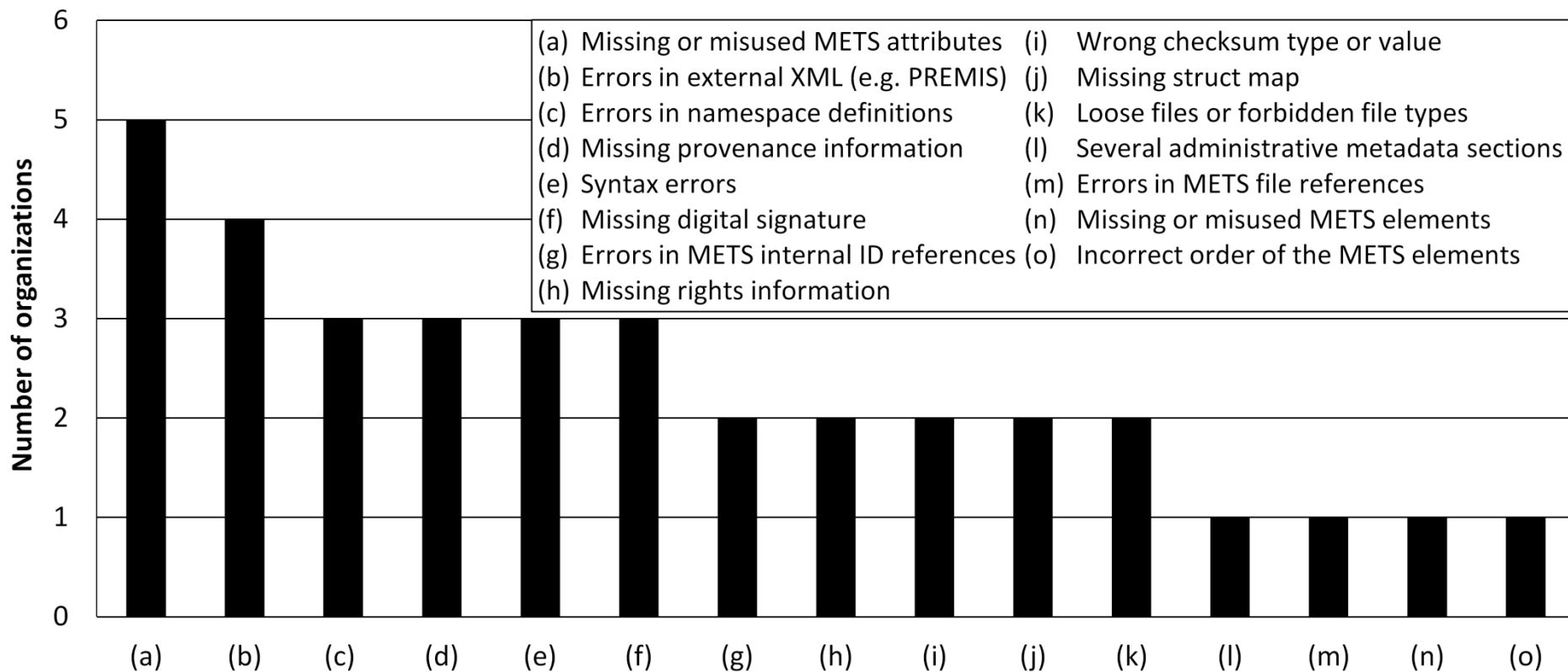
Pisteytyksen yhteenveto

- Onnistumiset:
 - Tiedostomuodot olivat paketeissa pääsääntöisesti oikeellisia,
 - Tiedostojen eheystiedot oli annettu
 - Tiedostoja oikea määrä
 - Aineistojen siirto sujui hyvin
- METS-dokumenttien luonti sujui kohtuullisesti
 - Virheet olivat pääsääntöisesti yksittäisiä kohtia
- Teknisen rakenteen ja sähköisen allekirjoituksen osalta havaittiin useita puutteita
 - Määritysdokumentin epäselvyydet ja ohjeistuksen puutteet
 - METS-dokumentin työstämiseen saattoi mennä niin kauan aikaa, että organisaation aikatauluresurssit eivät enää riittäneet sähköisen allekirjoituksen tai paketin teknisien rakenteiden tutustumiseen

Tiedostomuotojen validointi



Siirtopakettien puutteet koskien METS-dokumenttia



Pilottien tulokset

- Organisaatiot pystyvät tuottamaan vaaditun muotoisia aineistopaketteja siirrettäväksi pitkäaikaissäilytykseen
- Organisaatioiden valmiudet paketoinnin tekniseen suorittamiseen hyvin erilaiset
- (Pilottien pienellä otannalla) organisaatioiden aineistot olivat tiedostotasolla hyvässä kunnossa
- Organisaatiot tarvitsevat keskitettyä tukea aineistojen paketoinnissa
- Määrittysten ylläpitoon ja jatkuvaan kehittämiseen on syytä panostaa
- Toimintamallin kehittämisessä huomioitava paketoinnin kuormittavuus

Kiitos!

