



KANSALLISARKISTO

# Siirtopakettien ja tiedostojen validointi

Pitkäaikaissäilytyksen vaatimukset

Vesa-Matti Ovaska ja Pekka Rapinoja

29.5.2026



# Johdanto

- Tiedostojen validointi varmistaa tiedostojen säilymisen eheinä ja käytettävänä digitaalisen aineiston pitkäaikaissäilytyksessä
- Kansainvälisessä vertailussa tiedostojen tekninen validointi tarkkaa ja keskeisessä roolissa Kansallisarkiston digitaalisten aineistojen vastaanotossa
- Käytännön tiedonhallinnassa suuri riippuvuus sovellustoimittajien- ja laitevalmistajien ratkaisuista
- Tiedostojen tekninen validointi olisi tärkeää jo tiedostojen syntyhetkellä sekä kaikissa myöhemmissä tiedostokonversioiden vaiheissa.

# Siirtopakettien muodostaminen ja validointi

- Siirtopakettien muodostamisohjeet ja sallitut tiedostomuodot: <https://arkisto-osaaja.kansallisarkisto.fi/fi/digitaaliset-siirrot/ohjeet/siirtopakettien-muodostamisen-ohjeet>
- Siirtopaketit validoidaan SAPAssa kahdessa vaiheessa:
  1. Siirtopaketin rakenteen validointi (SAPA)
    - Onko tiedostot järjestetty ja nimetty siirtorakenteen vaatimusten mukaisesti?
  2. Siirtopaketin tiedostojen validointi (SAPA + KP-PAS)
    - Ovatko tiedostot kelvollisia pitkäaikaissäilytykseen:
      - Ovatko tiedostomuodot vastaanottokelpoisia?
      - Noudattavatko tiedostot pitkäaikaissäilytyksen yleisiä rajoituksia?
      - Ovatko tiedostot valideja eli noudattavatko ne tiedostomuodon vaatimuksia?

# Kulttuuriperinnön pitkäaikaissäilytyspalvelu KP-PAS

- CSC:n toteuttama keskitetty kansallinen palvelu kirjastoille, arkistoille ja museoille
- Takaa digitaalisten aineistojen pysymisen hyödynnettävinä satojenkin vuosien ajan
- Keskeistä tiedostomuotojen pysyvyyden arviointi. Kriteerejä: onko tiedostomuoto
  - avoin,
  - yleisesti hyväksytty pitkäaikaissäilytykseen,
  - vakaa (esim. eteen- ja taaksepäin yhteensopiva ja suojattu tiedoston korruptoitumiselta),
  - laite- ja ohjelmistoriippumaton ja
  - standardoitu.
- Tarjoaa tiedostojen validointityökalun (File Scraper), jota myös SAPA käyttää
  - Validoi tiedostosi -palvelu: <https://validointi-sapa.csc.fi/#/>

# Tiedostomuotojen vastaanottokelpoisuus

Ei-vastaanottokelpoiset  
tiedostomuodot



RTF, WordPerfect, LaTeX,  
BMP, Outlook (.msg ja  
.pst), ...

Siirtokelpoiset  
tiedostomuodot



MS Office-dokumentit  
(.doc(x), .xls(x) ym.), PDF,  
MP3, WMV, AVI, MOV, ...

Muunnetaan  
tulevaisuudessa

Säilytyskelpoiset  
tiedostomuodot



Open Document Format  
(.ods, .odt ym.), PDF/A,  
AAC, WAV, JPEG, TIFF, ...

# Tiedostojen yleiset rajoitukset

- Tiedostoissa ei saa käyttää salasanasuojauksia eikä muita salaustekniikoita
- Tiedostoissa ei saa käyttää DRM (Digital Rights Management) -tekniikoita
- Tiedostossa oleva digitaalinen allekirjoitus ei saa estää tiedoston käsittelyä
- Tiedostoja ei saa pakata
- Tiedostosta ei saa puuttua sen esittämiseen tarvittavia ulkoisia komponentteja
  - Näitä ovat esimerkiksi fontit ja muut linkitetyt aineistot, jotka vaikuttavat tiedoston esittämiseen

# Tiedostojen validointi

- KP-PASin validointityökalu File Scraper tarkistaa, onko tiedosto eheä ja hyvinmuodostettu
  - Saatavilla Githubista: <https://github.com/Digital-Preservation-Finland/file-scraper/tree/master>
  - Yhdistää useita laajalti käytettyjä validaattoreita, esim. PDF(/A)-tiedostot validoidaan VeraPDF:llä, JHOVElla ja Ghostscriptillä
- Tiedosto voi avautua ja näyttää täysin kelvolliselta mutta tulla silti hylätyksi
- Esimerkkejä validointivirheistä:
  - Värimääritykset PDF/A-tiedostossa eivät ole standardinmukaiset
  - JPEG-kuva on moniosainen (MPO-kuva, joita jotkin kamerat tuottavat)
  - XML-tiedosto ei noudata XSD-skeeman määrittämiä
- Useimpiin ongelmiin löytyy hyvät korjausmenetelmät
- Sähköisen arkistoinnin palvelu auttaa ongelmatilanteissa