



KANSALLISARKISTO

# Digitoinnin kehittäjä: standardit ja suositukset

Reko Etelävuori

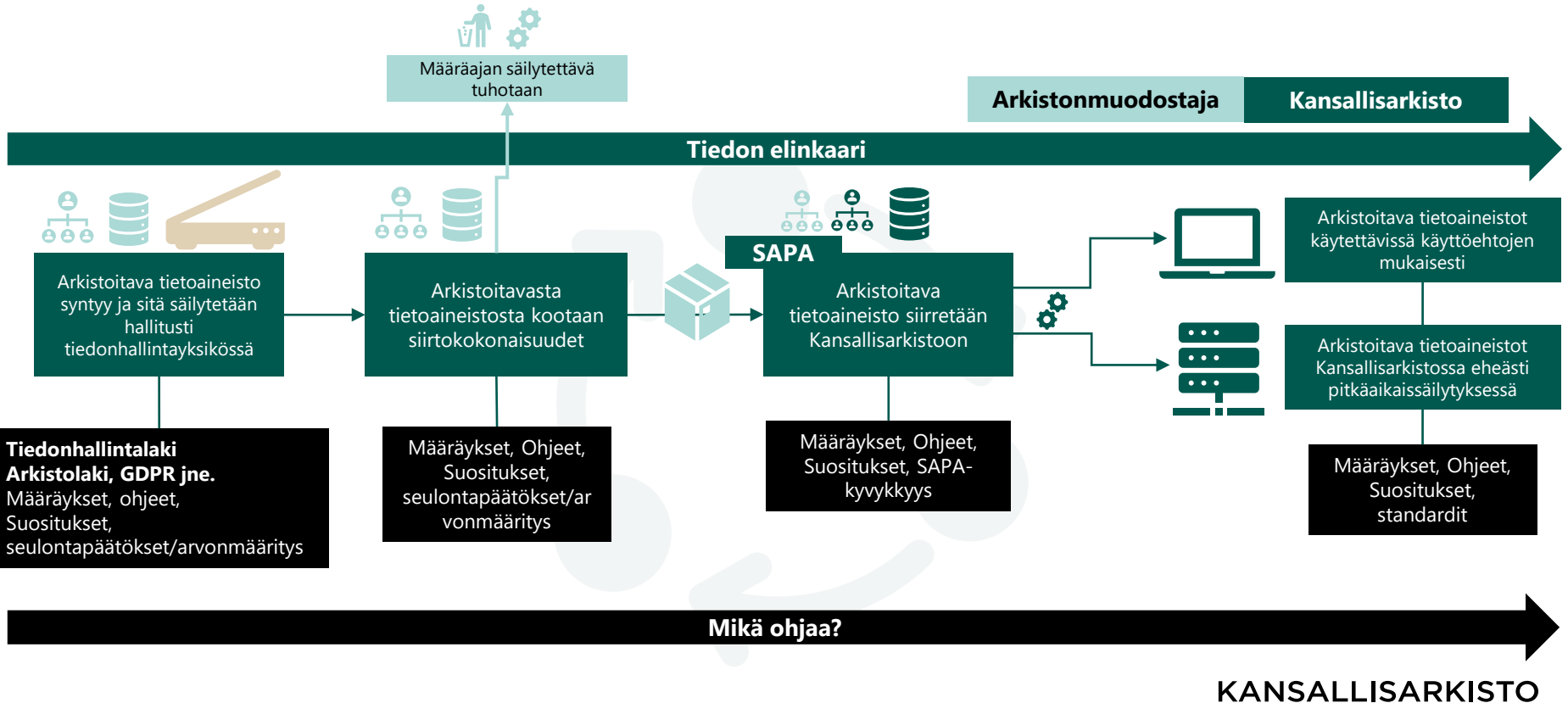
21.2.2024



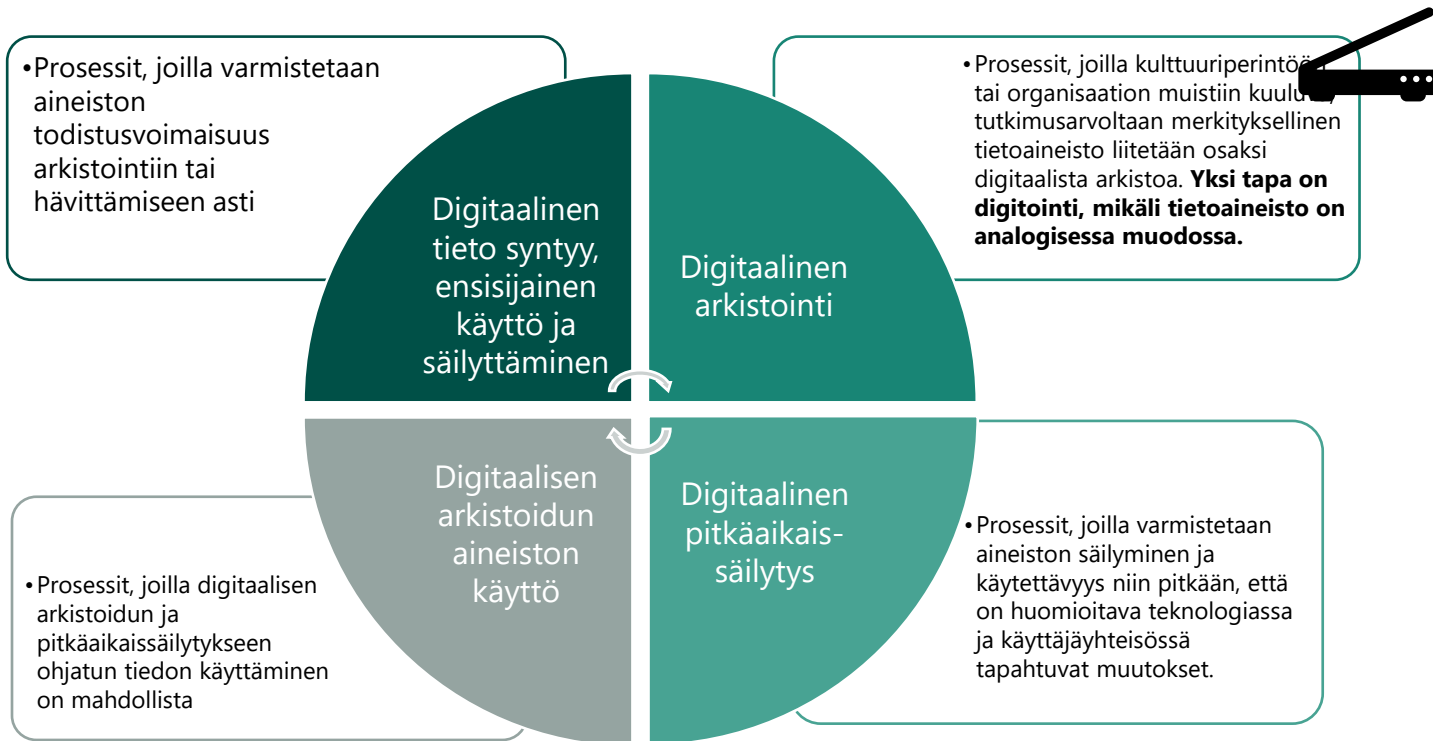
# Kansallisarkiston määritykset

- **Kansallisarkiston määritykset arkistoitavien erikoisaineistojen digitointiin**
  - Säädösperusta Laki Kansallisarkistosta 1145/2016 § 2
  - **Mitä ohjaa:** Erikoisaineistot koostuvat karttakokoelmista, piirustuskokoelmista, valokuvakokoelmista, negatiiveista, dioista, mikrofilmeistä sekä sidoksista, joissa sidosten tietosisältö esitetään aukeamittain. Ohje ei koske erikoisaineistoiksi luokiteltavia yksittäisiä asiakirjoja, jotka on arkistoitu A4- ja foliokokoisten asiakirjojen joukkoon.
  - [19.1.2021 KA/15780/07.01.02.04.02/2020 - Kansallisarkisto - Viranomaisten määräyskokoelmat - FINLEX ®](#)
- **Kansallisarkiston vaatimukset hävittämiseen tähtäävään digitointiin**
  - Arkistolaki (831/1994) 14 a §
  - **Mitä ohjaa:** Analoginen ilmentymä tarkoittaa asiakirjakokonaisuutta, joka koostuu pääsääntöisesti A4/foliokokoisista –paperiasiakirjoista (analogisista asiakirjoista), mutta se voi sisältää myös sitä suurempia tai pienempiä asiakirjoja.
  - [12.04.2019 KA/3357/07.01.02.04.02/2019 - Kansallisarkisto - Viranomaisten määräyskokoelmat - FINLEX ®](#)

# Tiedon elinkaari – teknistoiminnallinen tarkastelu (ei säädöskatsaus)



# Tiedonhallinnan kehä suhteessa arkistointiin



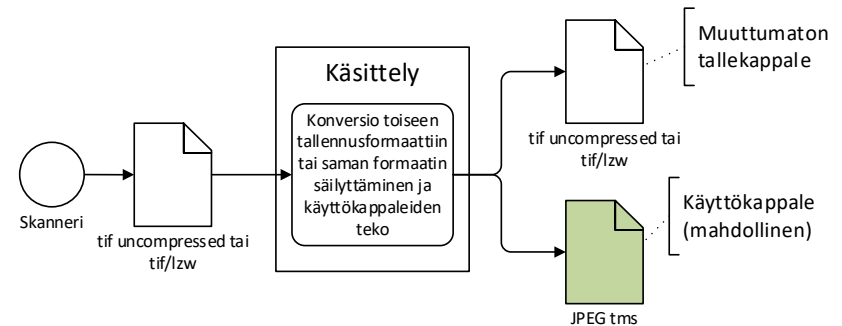
# Digitoinnin jälkeisen säilyttämisen kokonaisuus

- Digitointiin liittyvien kriteerien päivitystyön käynnistyessä 2018 etenkin metatietojen rooli korostui, kun lähtökohtana oli *mahdollistaa* analogisen ilmentymä hävittäminen
  - Pelkästään määrityksiä noudattamalla tämä ei toteudu -> arvonmääritys
- **KP-PAS –määritykset:** määrittää *pitkäaikaissäilytyksen* vinkkelistä formaatteja ja metatietoja -> Kansallisarkiston tehtävä määrittellä tähän mukaan toiminta. [Määrittelyt | Digitalpreservation.fi](#)
- **KV-kriteerit:** kriteerit on huomioitu Kansallisarkiston määrityksissä, mutta ei ole asetettu eksplisiittisiä kuvanlaatuun liittyviä arvoja, jotka empiriaan perustuen haastavaa tavoittaa -> *arvioitu tarkoituksenmukaisuutta*
  - **FADGI:** [Guidelines: Technical Guidelines for Digitizing Cultural Heritage Materials - Federal Agencies Digital Guidelines Initiative \(digitizationguidelines.gov\)](#)
  - **Metamorfoze:** [Digitization | Metamorfoze](#)
- Formaattivalinnoissa kyse **tietosisällöstä** – ei alkuperäisen asiakirjan fyysisestä koosta tms.



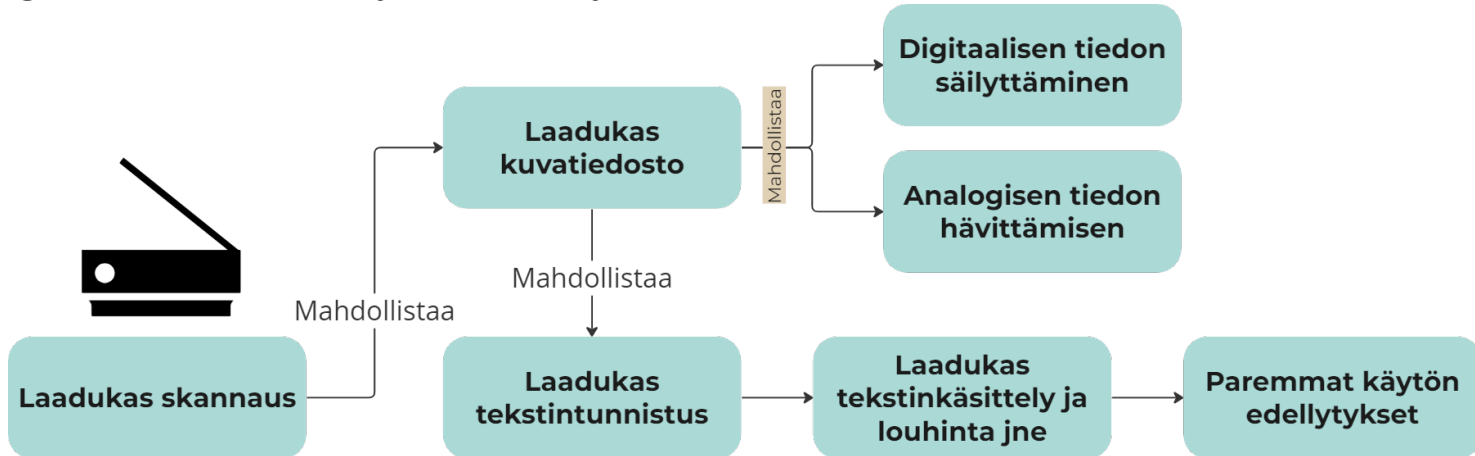
# Mitkä asiat vaikuttavat –tallennusformaatti ja laatu

- Sustainability, Cost factors, Settings and Capabilities, System Implementation Factors (Full Lifecycle)
- **Kansallisarkisto “määritysten” taustalla**
  - Kuvanlaatuanalyysi
    - Ohjelmallinen mittaus
    - visuaalinen arvio - testiryhmä
  - Kansainväliset määritykset
    - Muiden Kansallisarkistojen linjaukset
    - Aiemmin mainitut tekniset määritykset
  - Toimintaympäristö ja kyvykkyydet
  - Teknologia ja toimittajat



# Digitoinnin jälkeisen säilyttämisen kokonaisuus

- Analogista maailmaa ei ole tarkoitus *käöntää* sellaisenaan digitaaliseksi (vrt PDF-kuori = kotelon kannet)
- Digitoinnin tuloksena syntyneet tuotokset tulisi olla avoimia, "käsiteltäviä", "migroitavia", "käytettäviä" **tästä ikuisuuteen**
- Digitoinnin tuotoksia jalostetaan jatkossa



# Autenttinen digitaalinen asiakirja

- Prosessoinnin dokumentointi (eheys, laatu jne) sekä kuva että metatiedot
- Teknologia: laitteet ja järjestelmät - eheys on sekä datan eheyttä (*esimerkiksi tarkistesummin todettavaa*) että tietosisällön eheyttä (*jatkuvaa laaduntarkkailua*)
- Kuvanlaatu (digitaalinen *versio* sisältää saman tietosisällön kuin analoginen)
  - Kuvanlaatu perustuu mitattuihin tuloksiin
  - Tarve asettaa laaturaja ja kontrolloida tämän toteutumista



# Tietoturva ja käsittely

- Tietoturva huomioitava kaikissa käännteissä. Tietoturva koskee sekä analogista että digitaalista tietoa ja keskiössä nimenomaan tieto
- Yleislaeista sääntelyä muun muassa tiedonhallintalaissa, arkistolaissa, julkisuuslaissa
- Tietosuojaan liittyvää sääntelyä suoraan tietosuoja-asetuksesta
- Kasaavia vaateita ja kehikoita esimerkiksi KATAKRI:ssa ja JULKRI:ssa
- Tiedon luokittelu avainasia, joka myös määrittelee pitkälti käsittelyn ja prosessoinnin vaatimuksia
- Lähtökohtaisesti moni asia on riskiperustaista
- DPIA, TIA, Riskienhallinta, auditoinnit, arvioinnit, sertifikaatit jne

# Tekoäly ja arkistoitavat tietoineistot

- **Muutos – kuten yleensä – on pääosin toiminnallinen**
- **Prosessien tehostus**
  - Voidaan siirtää koneille tietynlaisia tehtäviä
  - Näitä voisi olla vaikkapa digitoinnin validointi (kuvien tarkastaminen), erilaisten asioiden tunnistaminen (vaikka tyhjät sivut) jne.
- **Tiedon käytettävyys**
  - Voidaan tunnistaa kuvien sisältämä teksti (OCR/HTR)
  - Tämän jälkeen voidaan tunnistaa tekstistä metatietoja/entiteettejä. Esimerkiksi nimiä, asialuokkia, organisaatioita, hetuja, asiakirjojen kieliä jne...
  - Tunnistettujen tietojen avulla voidaan esimerkiksi
    - Helpottaa löydettävyyttä
    - Yhdistellä tietoja
    - Anonymisoida tietoja julkaisua varten
    - Tarjota monikielisiä hakuja
    - Jne!

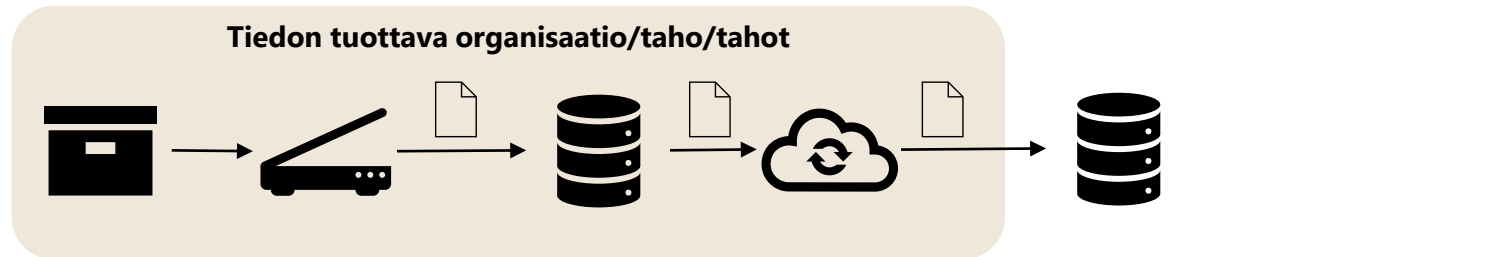


# Tekoäly ja arkistoitavat tietoineistot

- Edellyttää koko elinkaaren hallintaa
  - Mikäli ajatellaan vain yhtä transaktiota tai toimintoa, voi myöhempi käsittely olla hankalaa tai mahdotonta, joka johtaa laajaan "datan siivoamiseen"
- Tekoälyn jälkikäteinen hyödyntäminen asettaa vaatimuksia myös datanhallintaan
- Tiedon omistajuuden oltava selkeää
- Metatiedon laatu korostuu
- Tekoälyn tekemiä toimenpiteitä pitää niin ikään dokumentoida
- Voidaan ajatella, että tiedon käsittelyketjua jatketaan – elinkaari ei saa unohtua

# Siirtäminen Kansallisarkistoon tai muihin pitkäaikaissäilytysjärjestelmiin

- Digitointimääritykset **eivät ole** siirto-ohjeita
- Kun digitaaliseksi muuntaminen on tehty määritysten mukaisesti, on tiedon pitkäaikaissäilyttäminen on mahdollista ja esimerkiksi Kansallisarkiston palveluihin siirtäminen on mahdollista
- Julkishallinnossa tuotettu **tieto** on tarkoitus saada samoin teknisine ominaisuuksineen tallennettua tuleville sukupolville eheästi – ketju alkaa hetkestä, kun digitaalinen tieto syntyy
- Tiedonhallintalaki - erityisesti **15§** ja **19§**



Esitys on toteutettu osana Kaakkois-Suomen ammattikorkeakoulu Xamk:n, Ammattiopisto Samiedun ja Kansallisarkiston yhteishanketta JoDi-Joustavat koulutus ja työelämäpolut tulevaisuuden digitointiosaajille (ESR+ 2023-2025).



**Euroopan unionin  
osarahoittama**





# KANSALLISARKISTO

[www.kansallisarkisto.fi](http://www.kansallisarkisto.fi)



@kansallisarkisto



@kansallisarkist