



KANSALLISARKISTO

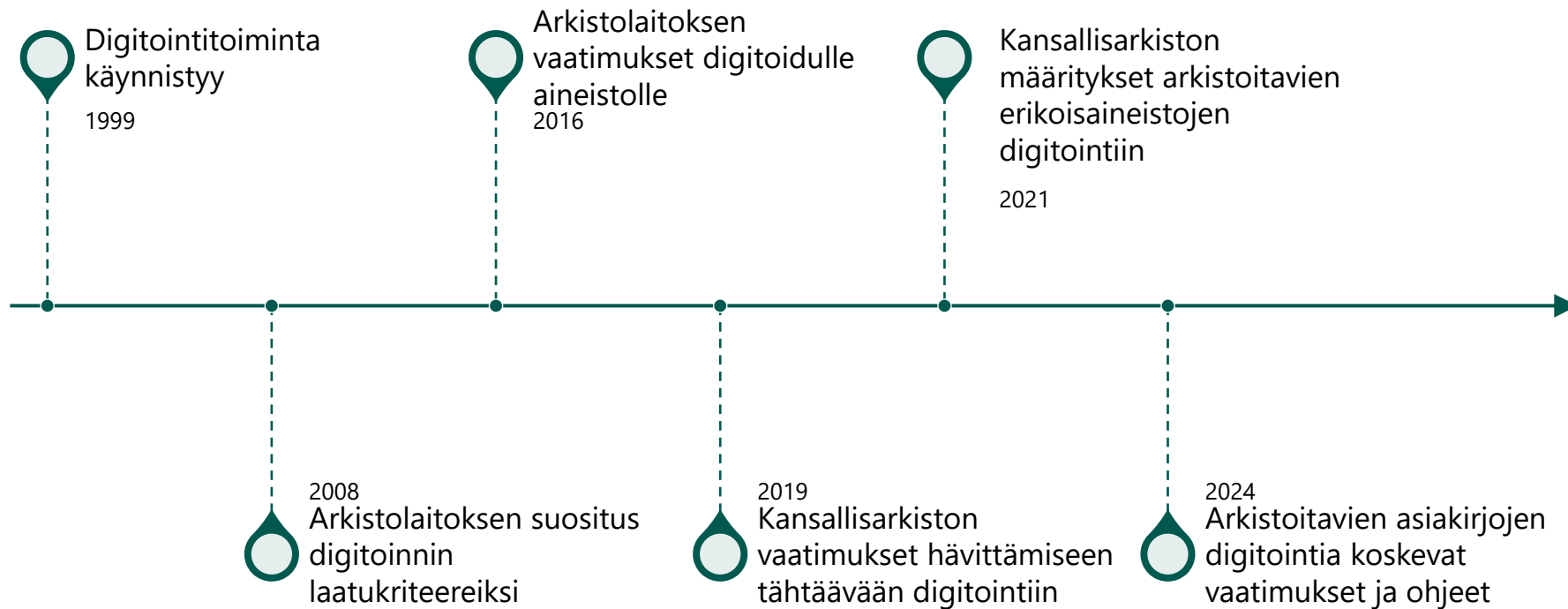
Kansallisarkiston digitointivaatimusten tausta ja tavoitteet

Reko Etelävuori, verkostopäällikkö

6.5.2025



Historiaa digitoinnista



Arkistolaitoksen suositus digitoinnin laatuksiteereiksi - 2008

- **Suositus** asiakirjallisen kulttuuriperinnön digitoinnin laatuksiteereiksi
- Kohderyhmä: *Digitointia toteuttavat muistiorganisaatiot ja yhteisöt*
- Suositus painottui asiakirjallisen aineiston, karttojen ja mikrofilmien digitointiin
- Sähköisessä muodossa syntynyt aineisto rajattiin ulos ja viitattiin SÄHKE-määrittelykseen
- Digitaalikalameralla tehty "**kuvaaminen**" rajattiin ulos
- Määriteltä "tallekappaleen" tiedostoformaattiksi TIFF. Resoluutioksi 300 ppi, bittisyvydeksi 24bit
- *Arkistolaitoksen Digitaalialkistossa olevan digitoidun aineiston hallinnassa noudatetaan arkistolaitoksen digitointistrategian linjauksia. Pitkääikaissäilytyksen ratkaisut tulevat lähivuosina täsmentymään muistiorganisaatioiden yhteistyön Kansallinen Digitaalinen Kirjasto -hankkeessa.*
 - Elimme aikaa ennen KDK-hanketta. Digitaalialkisto oli kokonaisuudessaan Kansallisarkiston ylläpidossa ja hallinnassa. Pitkääikaissäilytyspalveluita ei ollut.

Arkistolaitoksen vaatimukset digitoidulle aineistolle (2016)

- **Ohje:** Arkistolaitoksen vaatimukset digitoidulle aineistolle
- Kohderyhmä: *Digitoinnin vaatimusmäärittelyt on tarkoitettu arkistolaitokselle ja digitointia arkistolaitokselle toteuttaville muistiorganisaatioille, yhteisöille ja yrityksille*
 - Vaatimukset täydentävät arkistolaitoksen digitoinnin laatukriteerejä (AL/11130/07.01.02.04.02/2008).
- Ohjeessa pyrittiin jo dokumentoimaan hieman teknisiä metatietoja, joita ohjeistettiin tuottamaan TIFF-tiedostoihin XMP tai TIFF properties skeemoissa.
- Otettiin kantaa siirtorakenteisiin, mikäli digitoitua tietoaineistoa haluttiin siirtää Kansallisarkistoon
- Otettiin tarkemmin kantaa erilaisiin aineistotyyppeihin ja siihen, miten niissä digitointia voidaan suorittaa sekä miten tulisi digitoinnissa huomioida sekä käyttö että pitkäaikaissäilytys
- Digitaaliarkiston ylläpitoa oli jo ulkoistettu

Kansallisarkiston vaatimukset hävittämiseen tähtäävään digitointiin

- **Säädösperusta:** 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.
- **Suuri muutos:** Digitointia ei enää käsitetty niin puhtaasti laajemman käytön mahdollistajana, vaan "**surrogaattina**". Tarkoitus oli lähteä määrittelemään tapoja, miten analoginen tietoaaineisto voidaan *digitoinnin suorittamisen* näkökulmasta hävittää, mikäli muut kulttuurihistorialliseen arvoon liittyvät kriteerit toteutuvat
- Työ kesti suhteellisen kauan, koska pyrimme huomioimaan useita eri näkökulmia
- Mukaan tuli vaatimukset teknisistä metatiedoista merkittävästi laajemmin ja MIX-metatietoskeema, joka tallennetaan erilleen itse tiedostosta
- Toisaalta määriteltiin myös se, mitä tiedoston sisään tulee tallettaa metatietoina ja mihin rakenteeseen, jotta myös yksittäinen tiedosto tukisi sen autenttisuuden todentamista itsenäisesti, kun MIX-metatietoja ei olisi saatavilla.

Kansallisarkiston määritykset 2021

- **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.
 - Tuotiin mukaan erikoisaineistot (analogiset), eli ne, jotka oli aiemmin rajattu 2019 määritysten ulkopuolelle.
 - Nyt kaksi viimeisintä on yhdistetty asiakirjaan *”Arkistoitavien asiakirjojen digitointia koskevat vaatimukset ja ohjeet”*. Mukaan on tuotu myös uusia aineistotyyppejä.
 - Kahden viimeisen määrityksen yhteydessä lähdettiin jo rakentamaan **Sapa-tietojärjestelmää** ja siirryttiin **KP-PAS pitkäaikaissäilytykseen** (KDK-hanke). Samoihin aikoihin kaikki Digitaaliarkiston data migroitiin uuteen ympäristöön.
 - Migraation mahdollisti hyvin pitkälti se, että aiempia määrityksiä oli noudatettu suhteellisen hyvin

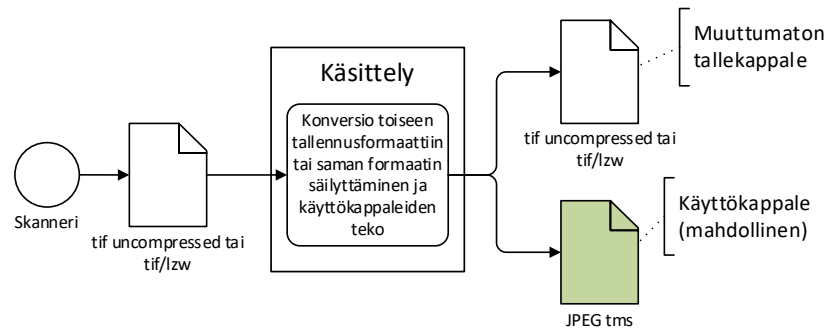
Digitoinnin jälkeisen säilyttämisen kokonaisuus

- Digitointiin liittyvien kriteerien päivitystyön käynnistyessä 2018 etenkin metatietojen rooli korostui, kun lähtökohtana oli *mahdollistaa* analogisen ilmentymä hävittäminen
 - Pelkästään määrityksiä noudattamalla tämä ei toteudu -> arvonmääritys, kulttuurihistoriallinen arvo jne
- **KP-PAS –määritykset:** määrittää *pitkäaikaissäilytyksen* vinkkelistä formaatteja ja metatietoja -> Kansallisarkiston tehtävä määritellä 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.



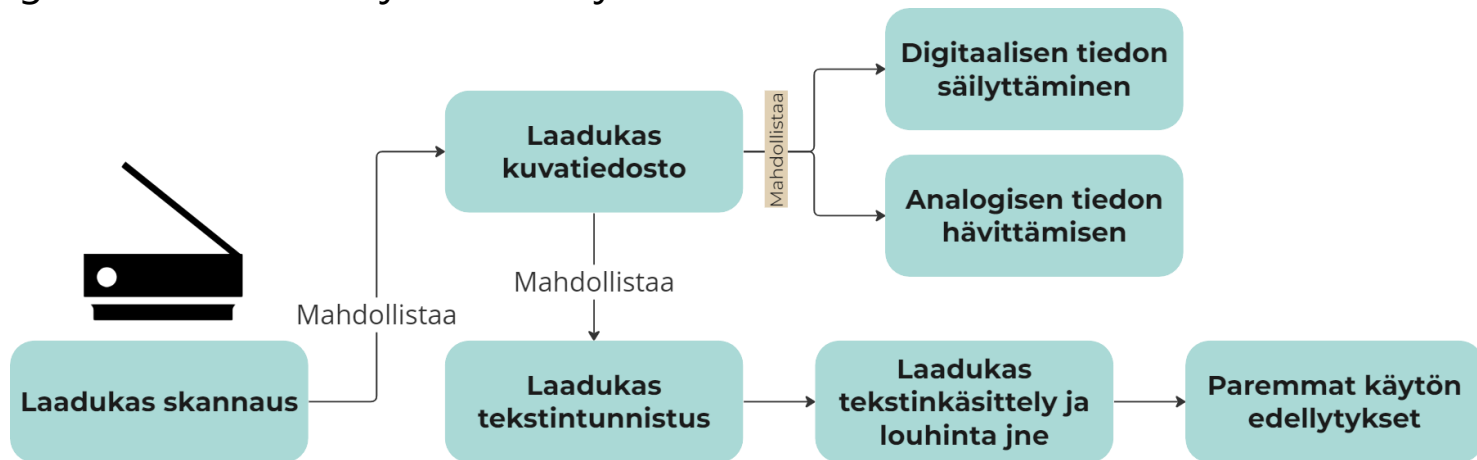
Mitkä asiat vaikuttavat –tallennusformaatti ja laatu

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

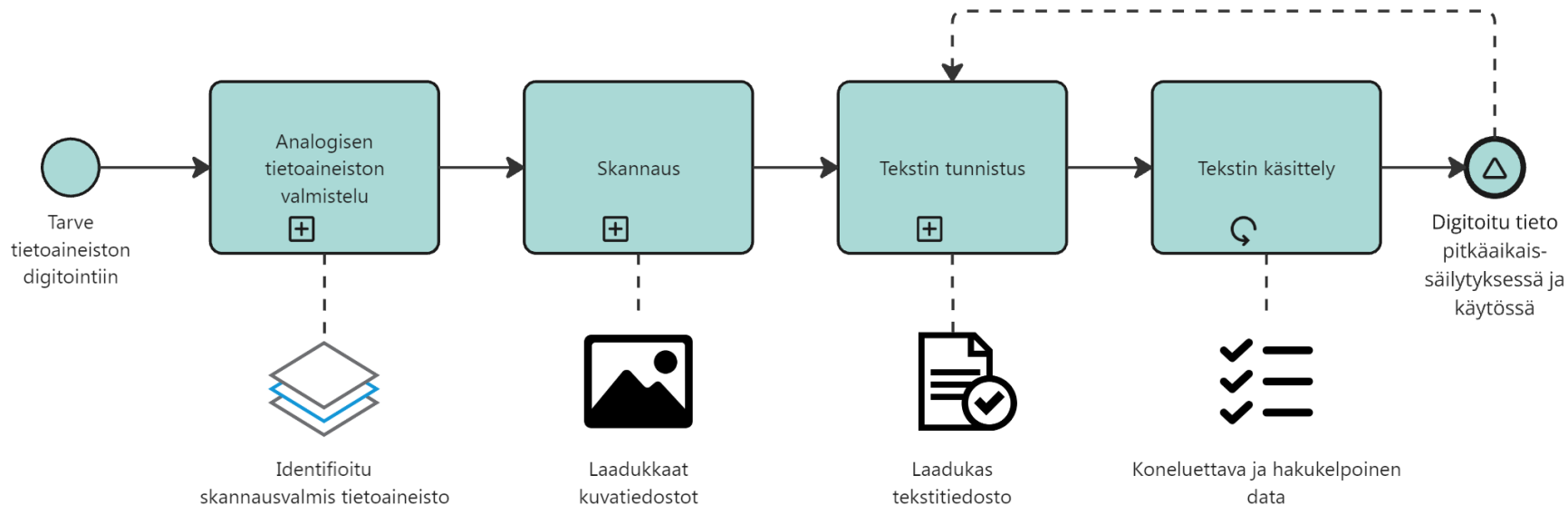


Mitä laadukas digitointi on mahdollistanut?

- Kansallisarkisto otti osaa READ-hankkeeseen (2016-2019). Hankkeessa tutkittiin mahdollisuutta tunnistaa käsinkirjoitettua tekstiä digitoinnin yhteydessä tuotetusta bittikarttakuvasta.
 - Kansallisarkisto oli mukana myös pitkälti sen takia, että digitoitu tietoa-aineistomme oli laadullisesti sellaista, että siitä oli järkevää lähteä kokeilemaan *tekoälyn* hyödyntämistä!
- Sittemmin muun muassa jalostaminen ja tekoälyn hyödyntäminen on ottanut aivan uudenlaisia askeleita ja sitä tehdään myös Kansallisarkistossa
- Kaiken pohjalla on kuitenkin digitoinnin suhteen se laadukas kuvatiedosto, jota on mahdollista hyödyntää!
- Garbage in -> Garbage out!
- Esimerkki: <https://tuomiokirjat.kansallisarkisto.fi>
 - "POC", jota on tarkoitus laajentaa muihin aineistoihin ja tietojärjestelmiin

Mitä "digitointi" on jatkossa?

- Voisiko olla näin? Tieto jalostuu koko sen olemassaolon ajan.





KANSALLISARKISTO

www.kansallisarkisto.fi



@kansallisarkisto



@kansallisarkist