

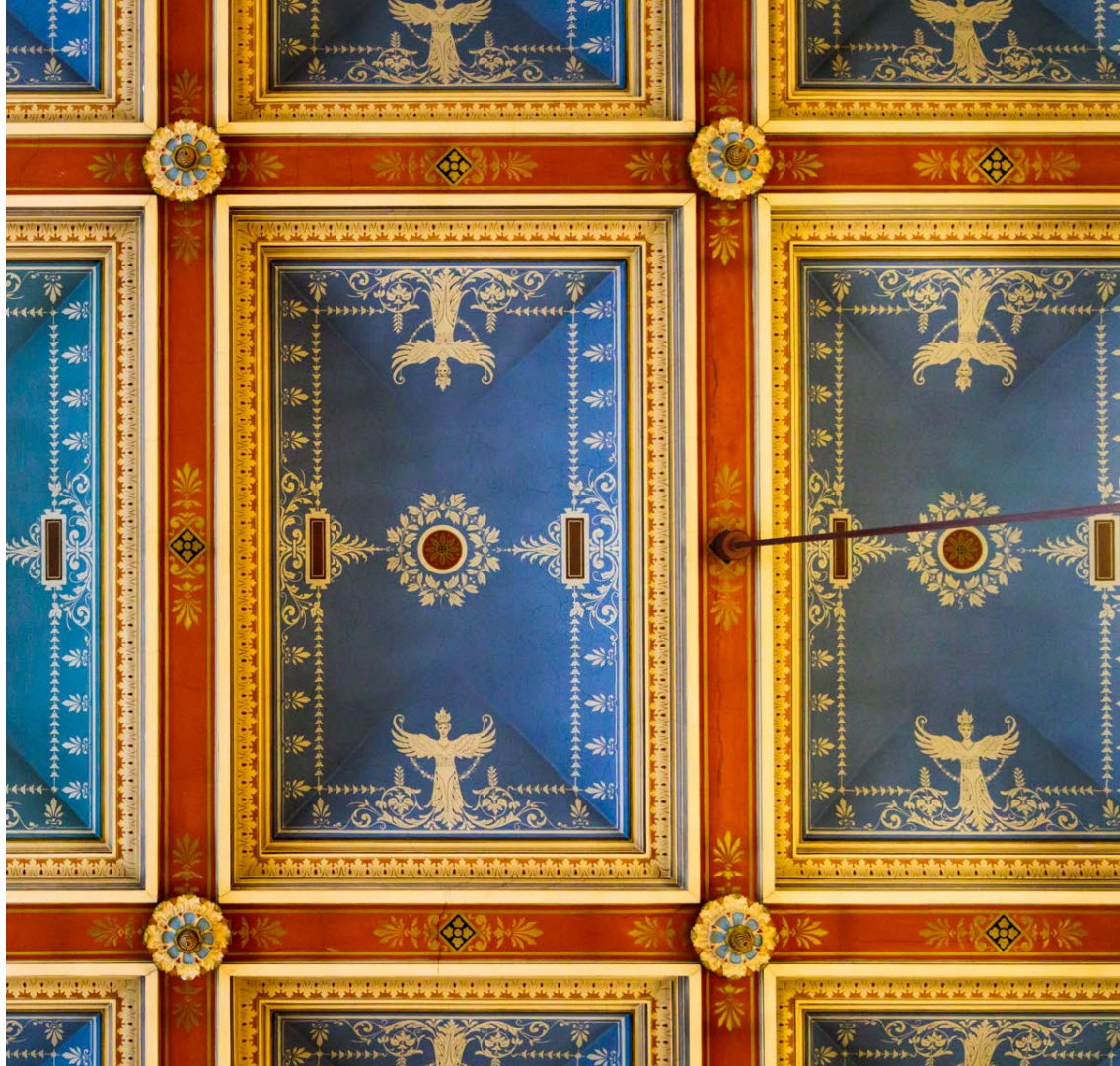


KANSALLISARKISTO

Laadun käsite massadigitoinnissa

Tiina Hölttä

Ylitarkastaja, tiiminvetäjä



Massadigitointi: kaksi toimipaikkaa

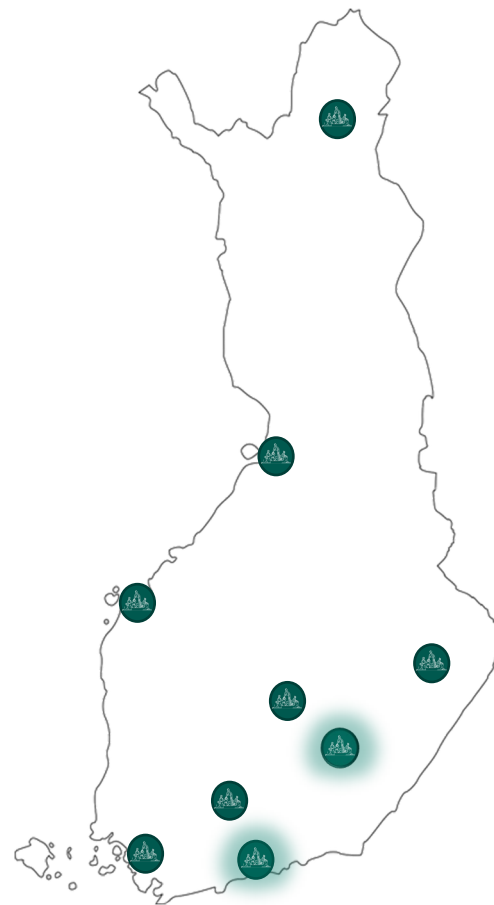


- **Helsinki**
- Hämeenlinna
- Inari (Saamelaisarkisto)
- Joensuu
- Jyväskylä
- **Mikkeli**
- Oulu
- Turku
- Vaasa



Helsingissä Digitoinnin yksikönpäällikkö 1, suunnittelijat 2, digitoijat 5

Mikkelissä Ylitarkastaja, tiiminvetäjä MD 1, suunnittelijat 2, digitoijat n. 15





Massadigitointi, mitä se on?

- Kansallisarkiston palvelu:
 - Digitointipalvelujen tuottaminen valtion viranomaisille
 - Valtionhallinnon pysyvästi säilytettävät analogiset tietoaineistot muutetaan keskitetysti digitaaliseen muotoon hävittämiseen tähtäävän digitoinnin vaatimusten mukaisesti
 - Digitoinnin jälkeen vastaava analoginen aineisto hävitetään
 - MD tunnusmerkitö: Hävittäminen, keskitetty toiminto, aineiston mittava määrä
 - Laajuus ja tavoitteet kansainvälisesti ainutlaatuisia



23.03.2024

Tiina Hölttä

MD- tuotannon prosessi

Viranomaisohjaus

- Tiekartta 135 hkm
- Digitointisuunnitelma
- Sopimukset: viranomaiset ja kuljetus

Siirtokuljetus

- Kuljetuksen suunnittelu ja kuljetus
- Vastaanoton valvonta
- Vastaanottotarkistus
- Siirron lopputoimet metatietojärjestelmässä

MD logistiikka

- Aineiston tuonti tuotannon logistiikkaohjelmaan
- Eteneminen tuotanto- ja viikkosuunnitelman mukaan
- Aineistohaut ja palautukset makasiiniin

MD valmistelu

- Digitointiohje
- Skannerien esivalinta
- Niitit, liittimet tms. pois

MD skannaus

- Erillisskannaus (S): dokumenttiskannerit (vain tarvittaessa)
- Erikoisskannaus (yli A3, postimerkkikoot, sidokset): suurteho, masto ja mankeli

MD valmistelu + skannaus

- Yhtäaikainen valmistelu ja skannaus (V + S) yleisin ja nopein tapa digitoida dokumenttiskannereilla

MD Validointi

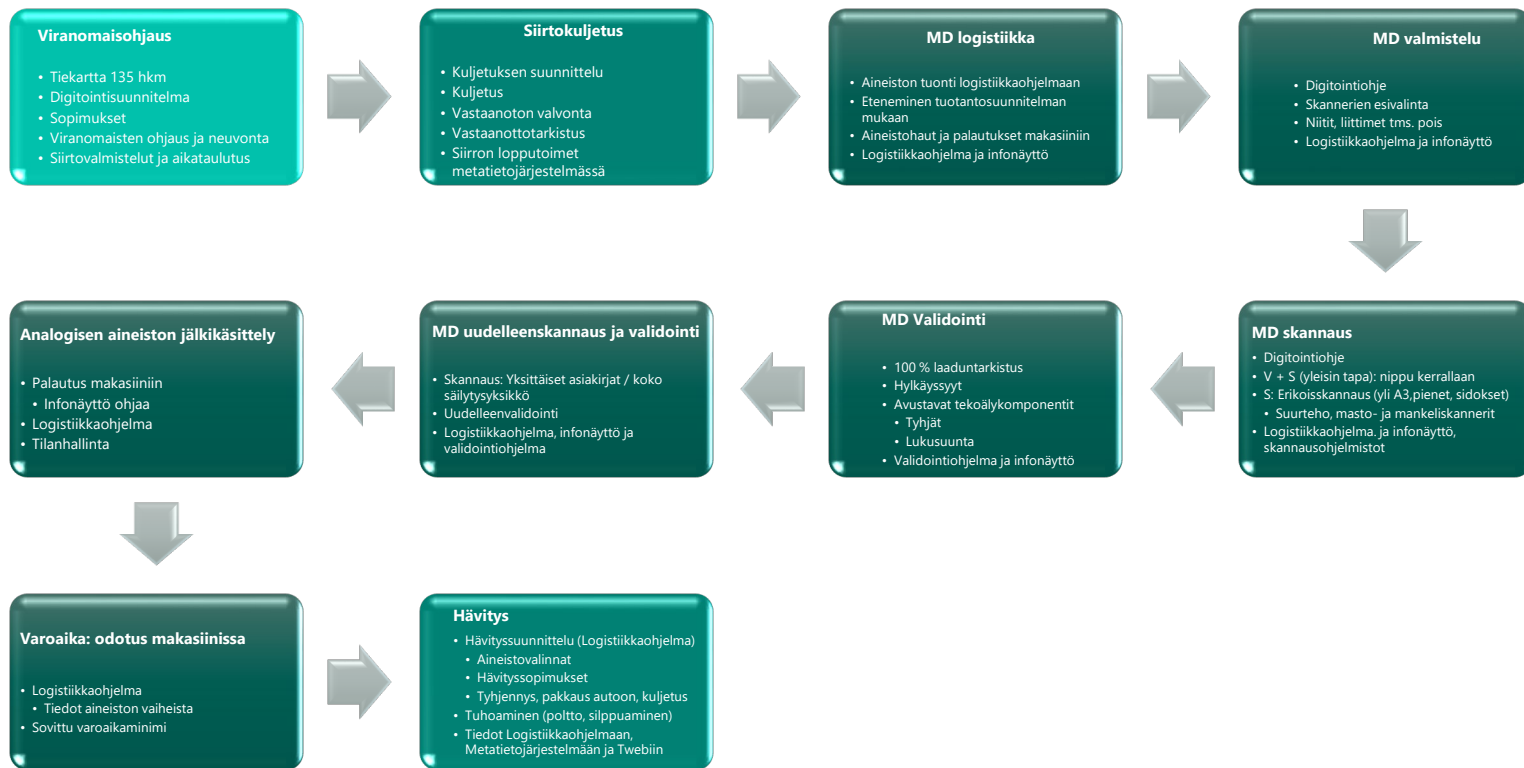
- 100 % laaduntarkistus
- Vakiodut hylkäysyyt
- Tekoälykomponentit avustavat (koneoppivat API:t)

MD uudelleenskannaus ja validointi

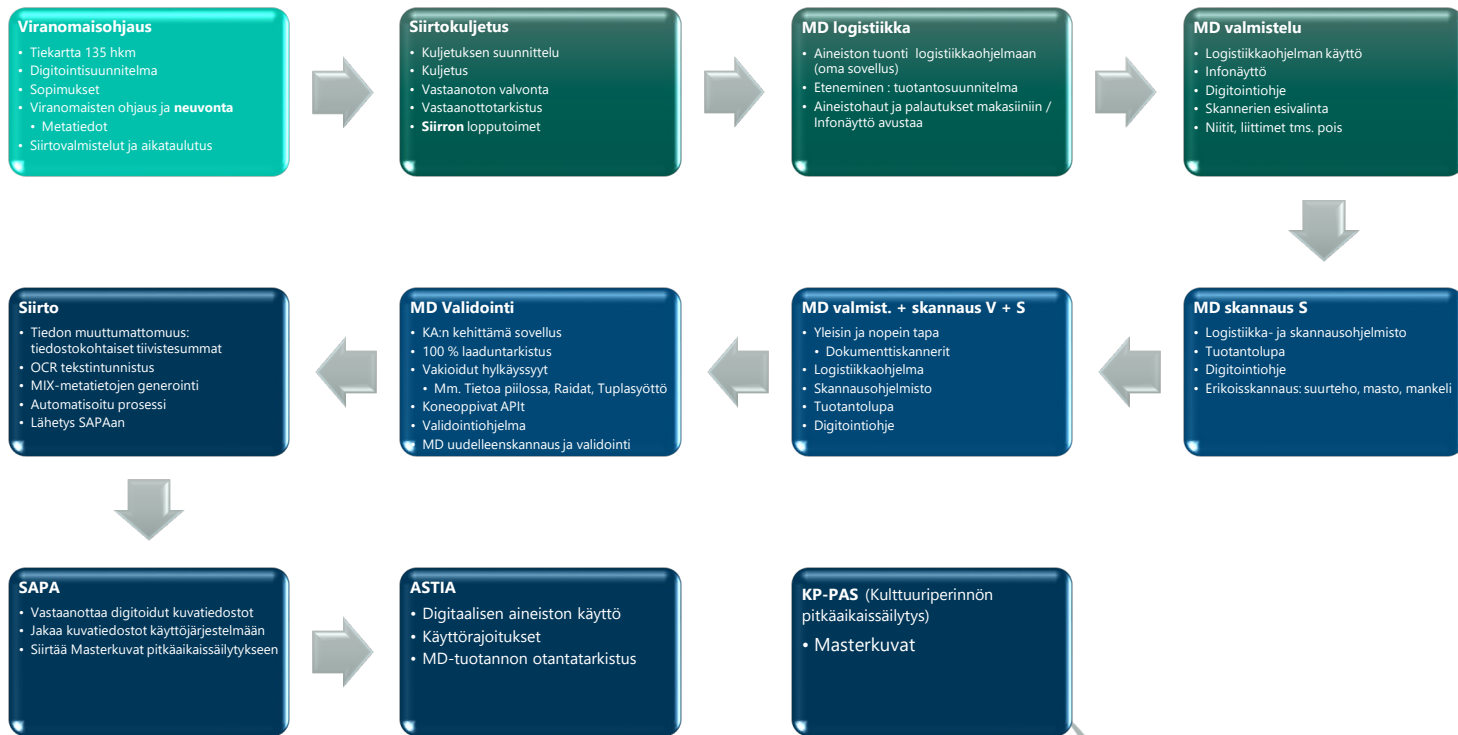
- Skannataan yksittäiset asiakirjat / koko säilytysyksikkö
- Uudelleenvalidointi

Analogisen ja sähköisen aineiston tied erkanevat

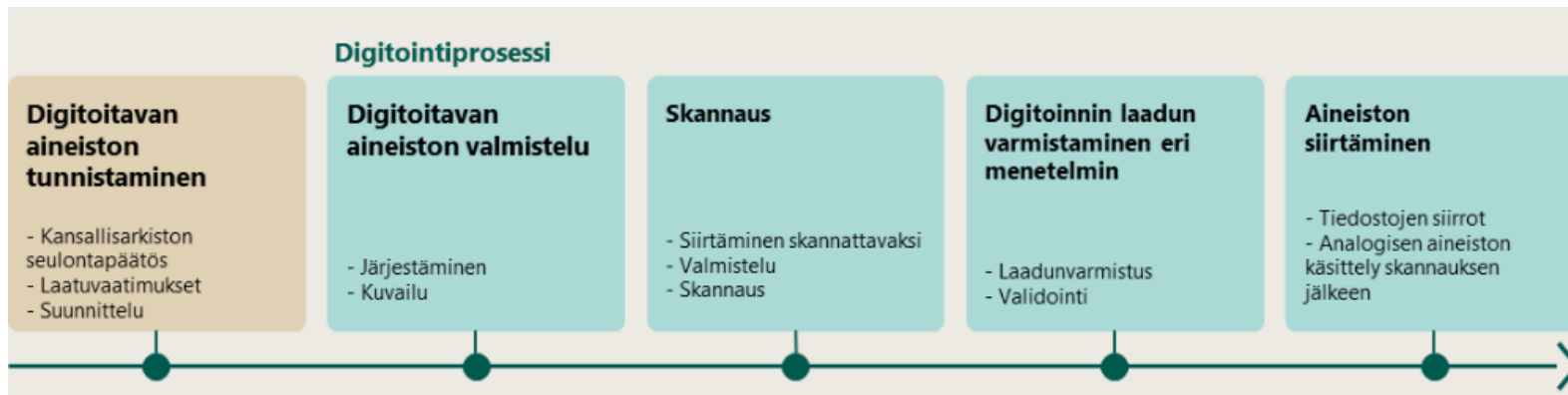
MD-tuotannon prosessi: Analoginen aineisto



MD-tuotannon prosessi: Digitaalinen aineisto



Digitointiprosessi



- Digitoinnin tavoite on tuottaa autenttisia ja todistusvoimaisia digitaalisia ilmentymiä alun perin analogisessa muodossa syntyneestä tiedosta.
- Analogisen asiakirjan tietosisältö pitää siirtää uuteen muotoon eheästi – kokonaisuudessaan ja virheettömästi.
- Tiedon säilymistä eheänä on voitava myös seurata tiedon elinkaaren kaikissa vaiheissa.



Laadun eri ulottuvuudet: Digitoinnin perusta

- Tekniset suositukset ja digitoinnin laadun mittarit
- Parhaat käytännöt: Dokumentoidut työkulut ja prosessit ohjeistuksineen, perehdytys
- Digitoinnin laadunarviointi: johdonmukaisuus ja eheys => Targettien ja digitaalisen kuvan laadun testaus ja analysointiohjelmat
- Ammattitaitoinen henkilöstö: monipuolinen osaaminen kaikissa digitointiprosessin vaiheissa
 - Tärkein voimavara!
- Lähde: FADGI-suositus, Museovirasto, Digitoinnin laadunhallinta 2023

Laadun eri ulottuvuudet massadigitoinnissa



Kansallisarkiston
hävittämiseen
tähtäävän digitoinnin
vaatimukset

Viranomaisvalmistelu
ja ennako-ohjaus

Ennakkotiedot
digitoitavasta
aineistosta

Perehdytys ja
ammattitaitoinen
henkilökunta

Digitointiohje:
yhteisen tiedon
luominen

Kuvanlaadun
määrittäykset, UTT

Toimivat laitteet ja
toimintaan soveltuvat
tilat

Tekninen ympäristö,
infra ja käytettävät
sovellukset

Työsuunnittelu,
ohjaus ja johtaminen

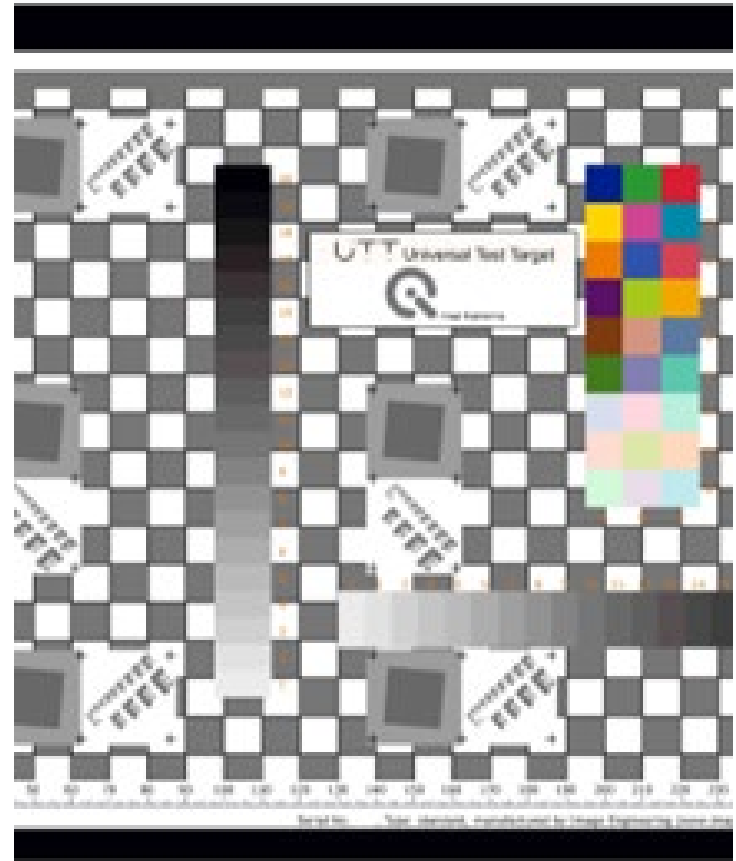
Kansallisarkiston kehittämät sovellukset ja palvelut /MD

- Kansallisarkiston kehittämät sovellukset ja niiden tehtävät
 - **Logistiikkaohjelma**
 - Fyysisten säilytysyksiköiden tuotantovaiheiden hallinta ja seuranta.
 - Ohjausarkit yhdistetyille skannaustöille.
 - **Infonäyttö**
 - Tuotannon reaaliaikainen seuranta eri tuotantovaiheiden välillä.Ka
 - Tuotantolupien hallinnointi
 - **Validointi**
 - Skannaustöiden kuvanlaadun tarkastus
 - **Siirtopalvelu**
 - Tuotantolinjastojen hallinta, metatietojen tarkastukset, siirto SAPA-järjestelmään.
 - **Prosessointipalvelut**
 - Tekstintunnistus (OCR)
 - Koneoppivat algoritmit
 - Tyhjien tunnistus
 - Kuvasuunnan tunnistus



Kuvanlaadun tarkastaminen ja tuotantolupa

- Kuvanlaatua tarkkaillaan ja mitataan päivittäin Kodak läpisyöttöskannereilla ja viikoittain kirja- ja suurokokoskannereilla.
- Kuvanlaatua mitataan kaupallisella iQ-Analyzer ohjelmistolla
- Jokaisella laitteella skannataan erillinen UTT-target (TE262) mittaussarkki ja skannattu digitaalinen kuva analysoidaan iQ-Analyzer ohjelmalla
- Skannerille annetaan tuotantolupa, mikäli laatukriteerien analyysi menee onnistuneesti läpi.
 - Epäonnistuessaan laite puhdistetaan ja analyysi tehdään uudestaan.



Validointi massadigitoinnissa

- Viimeinen tarkistuspiste ennen siirtoa SAPAan ja Astiaan sekä KP-PASSiin
- Validointivaiheessa seurataan arkistoyksiköiden kuvatiedostojen laatua ja skannausjälkeä ja tarkistetaan, että digitoiduista asiakirjoista ei puutu tietosisältöä
- Jokainen kuva tarkastetaan
- Kaupunkikohtaiset validointijonot Helsinki ja Mikkeli
- Jono validoidaan järjestyksessä
- Mahdollisuus nostaa jonon kärkeen aineistoja
- Validoitavia kuvia tarkastellaan arkistoyksikkö kerrallaan näyttönäkymittäin.
- Yksittäisille kuville valitaan tarpeen mukaan sopivat hylkäyisyty
- Aineisto hylätään tai hyväksytään. Hylätty arkistoyksikkö uudelleenskannataan, hyväksytty siirtyy SAPAan



Validoinnin hylkäyssyyt: MD

- Pyritään aina löytämään sopiva hylkäyssyy jo olemassa olevista hylkäyssyistä
- Yleisimmät hylkäyssyyt:
 - Tietoa piilossa
 - esim. taittunut kulma, jonka alle jää tietosisältöä
 - Arkki/sivu puuttuu (esim. asiakirjaa ei ole skannattu myös ilman post-it -lappua)
 - Raidallinen kuva
- Hylkäyssyitä voi tarkentaa tekstiselitteellä

1. Tuplasyöttö
2. Tietoa piilossa
3. Kuva väärinpäin
4. Huono kuvanlaatu
5. Rikkoutunut asiakirja
6. Väärä sisältö
7. Arkki/sivu puuttuu
8. Raidallinen kuva
9. Raita alkaa
10. Raita päättyy
11. Väärä järjestys
12. Muu syy + seliteteksti

Tekoälyavusteinen validointi

- Kaksi komponenttia:
 - Tyhjien tunnistus
 - Lukusuunnan tunnistus
 - Tavoitteena sujuvoittaa validointiprosessia
 - Tehokkaampi tuotanto ja vähemmän työaika validointiin
- Tällä hetkellä käytössä kaksivaiheinen validointi
 - Vaihe 1: Kuvat, jotka ovat tunnistuneet tyhjäksi, katsotaan validoijan toimesta, jos tyhjien kuvien joukossa on sisällöllisiä kuvia, ne merkitään syyllä kriittinen virhe
 - Vaihe 2: Sisällölliset kuvat katsotaan, mahdolliset hylkäyssyyt valitaan, tarkastetaan lukusuunnan tunnistuksien oikeellisuus
- Tulevaisuudessa mahdollisuus piilottaa tyhjät kuvat Astiassa
 - Vain tallekappaleissa tyhjät mukana, ei käyttökopioita tyhjästä: selailu nopeutuu

Hylkäyssyyt: tekoälyavusteinen validointi

- Merkintätavat tyhjille kuville ja kuvien kääntöasteille
 - Kriittinen virhe merkitään kuvaan, jossa komponentti on tunnistanut sisällöllisen kuvan virheellisesti tyhjäksi
1. Tyhjä
 2. Kuva väärinpäin 90°
 3. Kuva väärinpäin 180°
 4. Kuva väärinpäin 270°
 5. Kriittinen virhe

Massadigitoinnin hävittämisen kriteerit – edellytykset

- Kansallisarkiston ja viranomaisen välinen sopimus, lupa hävittää analogiset ilmentymät
 - Toteutus arkistolain 14 a § edellyttämin perustein
- Aineisto määrätty arkistoitavaksi yksinomaan digitaalisessa muodossa (ei kulttuurihistoriallista arvoa)
- Luovutuskokonaisuuden metatiedot on hyväksytty KA:n metatietojärjestelmässä ja aineiston hallinnoijana KA
- Luovutustodistus on toimitettu siirtäjälle, kun aineisto on kuljetuksen jälkeen tarkastettu
- Digitointiprosessissa noudatetaan KA:n hävittämiseen tähtäävän digitoinnin vaatimuksia ja sisäistä ohjeistusta, jossa määritellään digitoinnin toteutustapa ja laadunvarmistuksen menetelmät
 - Useita automatisoituja ja manuaalisia tarkastusvaiheita
- Digitointiprosessissa muodostuneet kuvatiedostot on tarkastettu ja hyväksytty järjestelmätasolla osana pitkäaikaissäilytys- ja käyttöliittymäjärjestelmiin toimittamista (SAPA, KP-PAS)
- Arkistoyksiköille on metatietojärjestelmässä automaattisesti muodostunut tieto digitaalisista ilmentymistä sekä elinkaaritapahtumiin metatiedot aineiston digitoinnista ja metatietojen ja digitaalisten ilmentymien tallekappaleiden siirtymisestä pitkäaikaissäilytykseen (PAS-AIP-ID)
- Digitaalisten ilmentymien (käyttökappaleiden) käytettävyys asiointipalvelussa (Astia) tarkastettu digitoinnin valmistumisen jälkeen
 - Varmistettu aineiston käytettävyys ja digitointi otannalla
- Analoginen aineisto on ollut edellisten kriteerien toteutumisen jälkeen säilytyksessä vähintään kuuden (6) kuukauden ajan (massadigitoinnin varoaika).
- Lisäedellytykset: Hävittäminen luovutuskokonaisuuksina, prosessi, työvaiheet ja viestintä määritelty, hävittäminen huomioitu riskienhallintadokumentaatioissa ja – menettelyissä (sähköisen säilyvyyden ja käytettävyyden varmistaminen häiriö- ja poikkeustilanteissa).

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



0295 33 7000



kirjaamo@arkisto.fi



Arkistopodi

www.kansallisarkisto.fi



@kansallisarkisto



@kansallisarkist