



KANSALLISARKISTO

Massadigitoinnin suunnitteluprojekti
9.2.2018

Vipuvoimaa
EU:lta
2014–2020



Euroopan unioni
Euroopan aluekehitysrahasto

Digitointiprosessin toteuttamissuunnitelma

Kansallisarkisto
9.2.2018



Sisällys

1	Johdanto	1
1.1	Raportin rajaukset	1
1.2	Digitaaliseen muotoon siirtämisen kokonaisuus	2
1.3	Prosessit	2
1.4	Aineistotyypit	3
1.5	Käsitteistö	3
2	Viranomaisvalmistelu (Prosessi 1.)	5
2.1	Taustoitus	5
2.2	Kansallisarkiston uusi säilytyspäättös	6
2.3	Viranomaisen valmistelutoimet aineistolle	8
2.4	Käsittelysuunnitelma	10
3	Logistinen siirto (Prosessi 2.)	10
3.1	Taustoitus	11
3.2	Logistiikan prosessikuvaus	11
3.3	Logistiikan tietojärjestelmävaatimukset ja ohjelmistoarkkitehtuuri	13
4	Aineiston fyysinen valmistelu digitointiin (Prosessi 3.)	14
4.1	Taustoitus	14
4.2	Yleisohjeet kaikille aineistotyypeille	15
4.3	Tarkentavat toimenpiteet	17
4.3.1	Paperiarkit sisältäen kortistokortit	17
4.3.2	Sidokset	18
4.3.3	Kartat ja piirustukset	19
4.3.4	Valokuvat	19
5	Digitointi (Prosessi 4.)	20
5.1	Taustoitus	20
5.2	Laitteistot	21
5.3	Skannaus (Prosessi 4.1)	22
5.3.1	Taustoitus	22



5.3.2	Prosessi	22
5.4	Validointi (Prosessi 4.2)	23
5.4.1	Taustoitus	23
5.4.2	Prosessi	23
5.5	Laadunvarmistuksen käytänteitä	25
5.5.1	Skannereiden ja skannausohjelmistojen mahdollistama laadunvarmistus	26
5.6	Digitoinnissa tuotettavat tiedostot	26
5.6.1	Bittisyyden vaikutus kuviin.....	27
6	Digitaalisten ilmentymien siirto (Prosessi 5.)	28
6.1	Taustoitus	28
6.2	Prosessi	29
7	Analogisen aineiston käsittely digitoinnin jälkeen (Prosessi 6.)	29
7.1	Taustoitus	29
7.2	Prosessi	30
7.2.1	Analoginen aineisto säilytetään digitoinnin jälkeen.....	30
7.2.2	Analoginen aineisto hävitetään prosessoinnin jälkeen	30
8	Tietotekninen infrastruktuuri ja tuotannonohjausjärjestelmä	31
8.1	Taustoitus	31
8.2	Järjestelmät, prosessit ja tietovirrat.....	31
8.3	Tuotannonohjausjärjestelmä	34
9	Yhteenveto tiedostoformaateista	36
9.1	Yhteenveto kuvaformaateista	36
9.2	Datamäärät.....	37
9.3	Kuvaformaatit.....	37
9.4	Optinen tekstintunnistus (OCR).....	39
9.5	Metadata	39



1 Johdanto

Tässä raportissa on määritelty massadigitoinnin toteuttamisen koko prosessi, jonka alkupiste on aineiston viranomaisvalmistelun käynnistäminen ja päätepiste digitaalisen ilmentymän siirtäminen pitkäaikaissäilytykseen. Tarkoituksena on antaa kokonaiskuva digitointiin liittyville eri toimijoille (viranomaiset, Kansallisarkisto, digitoiva taho, logistiikka jne.) siitä, miten prosessikokonaisuus tulisi järjestymään.

Digitointi ei ole yhtä kuin skannaus ja sitä ei voida käsittää ainoastaan teknisenä toimenpiteenä. Skannaus on tässä raportissa esiteltävässä prosessikokonaisuudessa varsin pieni osa-alue. Ennen skannausta ja sen jälkeen suoritetaan useita toimenpiteitä, joilla pyritään takaamaan prosessissa tuotetun digitaalisen manifestaation autenttisuus.

Massadigitoitava aineisto on pääosin pysyvästi säilytettäväksi määriteltyä asiakirjallista kulttuuriperintöä, jonka analoginen manifestatio tullaan hävittämään sen digitaaliseen muotoon muuntamisen jälkeen. Digitaaliseen muotoon muuntaminen saa näin useita erilaisia vaatimuksia lähtien analogisen aineiston käsittelytavoista, päätyen siihen, että aineiston tulisi tarjota todistusvoimainen, autenttinen ja tietosisällöltään täydellinen digitaalinen ilmentymä. Pitkäaikaissäilytykseen on saatettava tallekappale, joka on laadukas ja josta on tulevaisuudessa tuotettavissa erilaisia käyttökappaleita.

Kokonaisuuteen sisältyvistä prosesseista on kaikista laadittu BPMN-mallinnuksen (Business Process Modeling Notation) mukaiset prosessikuvaukset. Raportissa prosessit on kuvattu sanallisesti sekä lisäksi avattu tarkemmin prosessin eri vaiheita ja niissä käytettäviä menetelmiä. Raportin lukemisen yhteydessä on suositeltava tutustua myös BPMN-prosessikuvauksiin.

1.1 Raportin rajaukset

Raportissa on kuvattu digitoinnin kokonaisprosessi seuraavin rajauksin:

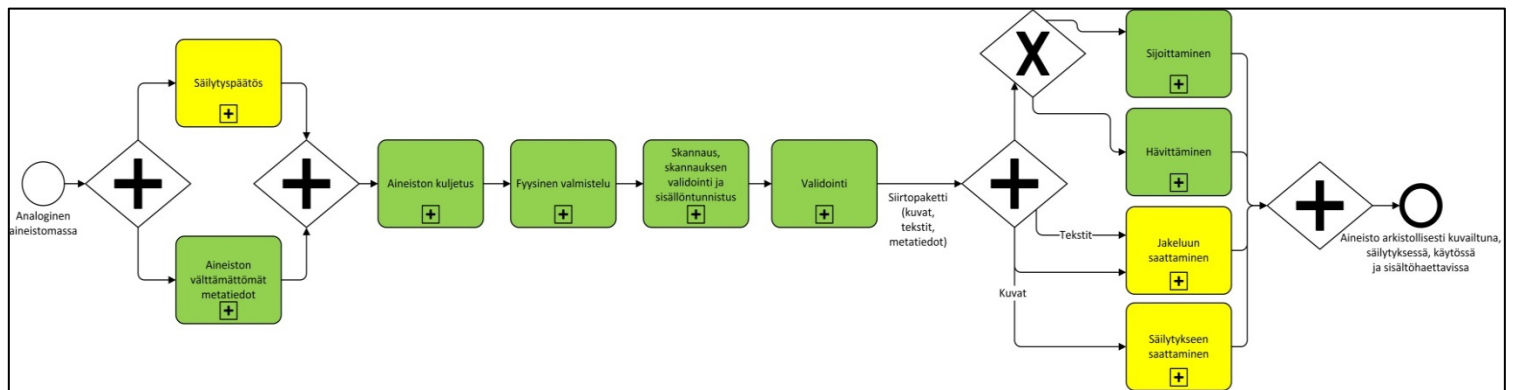
- **Raportti ei ole valmis viranomaisohjeistus**, vaan suunnitteluprojektin näkemys siitä, miten massadigitoinnin prosessi on toteutettavissa tehokkaasti ja laadukkaasti. Raporttia ei ole laadittu ohjeen muotoon, vaan enemmän teknisen toteuttamisen näkökulmasta, mikä heijastuu esimerkiksi käytettyyn terminologiaan.
- Jatkotyössä raportissa esitetyt toimenpiteet tulevat vielä tarkentumaan ja voivat muuttua. Kansallisarkistossa tullaan laatimaan erillinen viranomaisohjeistus (siirtomääräys ja aineiston valmistelun ohjeistus).
- Raportissa ei oteta kantaa digitaalisten ilmentymien säilytysratkaisuun ja siihen, miten ne saatetaan viranomaisten käyttöön.
- Raportissa ja prosessikuvauksissa on alustavasti tunnistettu digitointiin liittyviä toimijoita (viranomainen, Kansallisarkisto, logistiikasta vastaava ja digitoitava taho). Näiden toimijoiden

työnjako voi muuttua jatkovalmistelussa. Raportti ei ota kantaa siihen, miten prosessi tullaan hallinnollisesti järjestämään.

- Raportissa on kuvattu digitoinnissa vaadittavia tietojärjestelmiä ja teknologioita, eikä siinä oteta kantaa käytettäviin tuotteisiin.

1.2 Digitaaliseen muotoon siirtämisen kokonaisuus

Alla olevassa kuvassa (Kuva 1-1, Prosessikokonaisuus) esitetään ei-toiminnallinen kokonaisprosessi, joka sisältää tässä raportissa huomioitavat vaiheet. Vihreällä merkittyjen vaiheiden toteuttaminen on määritelty tässä raportissa. Keltaisiin kohtiin massadigitointi aiheuttaa todennäköisesti vaatimuksia ja tarpeita.



Kuva 1-1, Prosessikokonaisuus

1.3 Prosessit

Suunnittelutyössä on tunnistettu ja määritelty kuusi toiminnallista prosessia, jotka kattavat massadigitoinnin kokonaisprosessin. Jokainen osaprosessi tuottaa taulukossa 1-1 esitetyn tuotoksen, joka on edellytyksenä seuraavan osaprosessin toteuttamiselle.

Taulukko 1-1: Prosessikokonaisuus

#	Prosessi	Aliprosessi	Välituotos	Tuotos
1.	Viranomaisvalmistelu	-	Viranomaisvalmisteltu aineisto	Siirtovalmis aineisto
2.	Logistinen siirto	-	-	Kuljetettu aineisto
3.	Aineiston fyysinen valmistelu	-	-	Digitointivalmis aineisto
4.	Digitointi	4.1. Skannaus	-	Skannattu aineisto
	Digitointi	4.2. Validointi	-	Lähetysvalmis digitoitu aineisto
5.	Digitaalisten ilmentymien siirto	-	-	Vastaanottavaan järjestelmään siirretty digitaalinen manifestaatio
6.	Analogisen aineiston käsittely digitoinnin jälkeen	-	-	Hävitetty analoginen manifestaatio/makasiiniin sijoitettavaksi valmis aineisto

1.4 Aineistotyypit

Massadigitoinnissa käsitellään paperiasiakirjoja sisältäen kortistoaineistot sekä sidokset. Edellä mainitut aineistotyypit voivat sisältää yksittäisiä erikoisaineistoja, kuten taiteltuja, isokokoisia karttoja, jotka valmistellaan muun aineiston yhteydessä ja skannataan erikoisaineistolle tarkoitetulla skannerilla. Erilliset omina sarjoina tai kokoelmina arkistoidut kartta-, piirustus- tai valokuvakokoelmat rajataan kokonaan massadigitoinnin ulkopuolelle. Massadigitoinnin ulkopuolelle jäävät myös esimerkiksi valokuvien negatiivit, diakuvat, lasinegatiivit, mikrofilmit, mikrokortit ja AV-aineisto, jotka vaativat erityisiä valmistelutoimenpiteitä ja skannauslaitteistoa.

Massadigitoinnin alkuvuosina keskitytään helposti valmistettavaan aineistoon mutta myöhemmin myös vaativampiin aineistoihin, joiden valmistelu vie enemmän aikaa. Aineistojen digitoinnin suunniteltua järjestystä suhteessa aineiston ominaisuuksiin on käsitelty tarkemmin tiekarttasuunnitelmassa.

1.5 Käsitteistö

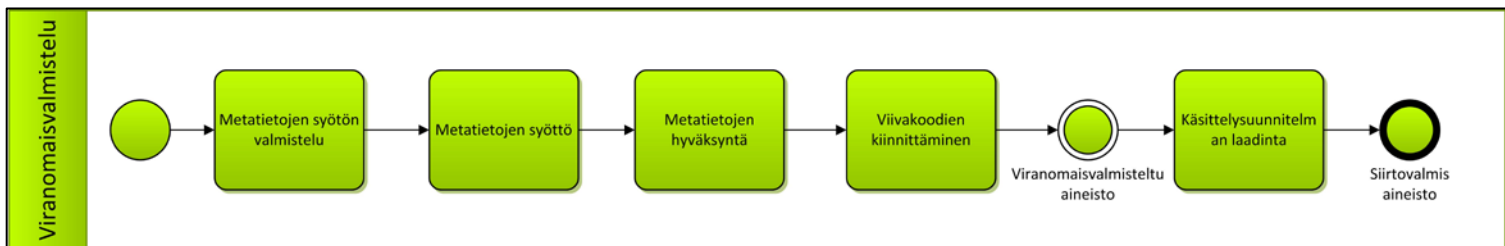
Taulukko 1-2: Käsitteistö

Käsite	Selitys
AHAA	Arkistoaineiston metatietojärjestelmä. Kansallisarkiston arkistoaineiston kuvailu- ja luettelointitietoja sisältävä metatietojärjestelmä.
AHJ	Aineistohallintajärjestelmä. Kansallisarkiston analogisen aineiston sijaintitietojen järjestelmä.
Aineistokokonaisuus	Muista arkistoaineistoista erillään kuvailtava eli itsenäinen sisällönkuvailukokonaisuus (esim. arkisto tai kokoelma), joka voi jakaantua sarjoihin ja/tai arkistoyksiköihin. Kuvailutasona aineistokokonaisuus on ylin (ainoa pakollinen) aineiston sisällönkuvailukokonaisuus.
Analoginen aineisto	Paperisessa tai muussa käsin kosketeltavassa muodossa laadittu, säilytetty ja/tai käytettävä arkistoaineisto. Yleistermi analogiselle aineistolle, joka ei ole tässä raportissa yksiselitteinen arkistokuvailullinen kokonaisuus. Luettavuuden lisäämiseksi on käytetty myös pelkästään käsitettä aineisto. Vrt. digitaalinen aineisto.
Analoginen ilmentymä	Asiakirjan/asiakirjatiedon analoginen versio. Ks. myös analoginen aineisto, joka on lähisynonyympi analogiselle ilmentymälle.
Analoginen manifestaatio	Digitoitavaksi päätetyn teoksen/asiakirjakokonaisuuden analoginen ilmiä, jonka osa on digitaalinen ilmentymä
Analogisen aineiston hävittämisen varmuusaika	Analogisen aineiston digitoinnin jälkeinen aika, jonka kulumisen jälkeen digitoidut ja tuotannonohjausjärjestelmässä tilassa "odottaa hävittämistä" olevat analogiset aineistot voidaan fyysisesti hävittää.
Arkistoyksikkö	Aineistokokonaisuuden tai sarjan osa, luettelointiyksikkö. Kuvailutasoltaan alin tässä raportissa käsiteltävä metatietokokonaisuus, mihin skannaamalla tuotetaan sekä myöhemmässä vaiheessa liitetään aineiston digitaalinen manifestaatio
Asiakirja	Arkistonmuodostajan tehtävien yhteydessä syntynyt kirjallinen tai kuvallinen esitys (tietosisältö) tai sellaista sähköisesti tai muulla vastaavalla tavalla aikaansaatu esitys, joka on luettavissa, kuunneltavissa tai muutoin ymmärrettävissä teknisin apuvälinein.

Bittikarttakuva	Kuvatiedosto, joka muodostuu pikseleistä. Bittikarttakuva voidaan tallentaa useampaan eri tiedostoformaattiin.
Digitaalinen aineisto	Sähköisessä muodossa laadittu tai sähköiseen muotoon teknologian avulla muunnettu arkistoaineisto. Yleistermi digitaaliselle aineistolle, joka ei ole tässä raportissa yksiselitteinen arkistokuvailullinen kokonaisuus. Vrt. analoginen aineisto.
Digitaalinen ilmentymä	Analogisesta asiakirjasta skannaamalla tuotettu sähköinen versio, joka on osa digitaalista manifestaatiota.
Digitaalinen manifestaatio	Digitoitavaksi päätetyn analogisen teoksen/asiakirjakokonaisuuden digitaalinen ilmiasu.
Digitoinnin testitaulu	Testitaulu, jota käytetään digitoinnin laadun määrittelyssä.
Erikoisaineisto	A3-kokoa suurempi tai 65x85 mm kokoa pienempi tai muuten fyysisiltä ominaisuuksiltaan sellainen aineiston analoginen ilmentymä, joka ei sovellu skannattavaksi suurtehoskannerilla.
Kuljetusyksikkö	Analogisten aineistojen siirtoon käytettävä apuväline, eli rullakko.
Kuvan resoluutio	Erottelutarkkuus. Kertoo kuinka monta pikseliä on tietyllä mitalla bittikarttakuvassa. Ilmaistaan tässä dokumentissa arvona ppi (pixels per inch).
Käyttökappale	Käyttökappale tarkoittaa digitointiprosessissa tuotettua digitaalista ilmentymää, joka tarjotaan käytettäväksi esimerkiksi verkkokäyttöliittymän kautta. Yleisesti käyttökappale on tietosisällöltään tallekappaleen kanssa identtinen, mutta informaatio esitetään pakatussa tiedostoformaattissa.
Laatuarkki	Digitoidun manifestaation yhteyteen liitettävä digitointitekniisiä testitauluja omaava paperilomake.
Optinen resoluutio	Analogisen aineiston digitaaliseen kuvamuotoon muuntavan laitteen erottelutarkkuus.
PDA	Personal Digital Assistant. Viivakoodinlukijalla varustettu kannettava kämmentietokone/käsipäätte.
Pikseli	Kuvaelementti, joka on bittikarttakuvan pienin osa.
PPI	Pikseliä per tuuman jana bittikarttakuvassa.
Siirtovalmis aineisto	Logistista siirtoa odottava analoginen aineisto, joka on viranomaisen toimesta valmisteltu ja jolle on tehty käsittelysuunnitelma Kansallisarkiston toimesta.
Sisällön tunnistus	Konekirjoitetun tai tietokoneella tuotetun bittikarttakuvassa esiintyvän tekstin tunnistaminen OCR-menetelmin.
Skannausohjelmisto	Ohjelmisto, joka ohjaa yksittäisen skannauslaitteen toimintaa.
Skannauslinjasto	Skannausinfrastruktuurissa erilaisia laitetyyppejä tarkoittava linjasto.
Skannausresoluutio	Skannauksessa käytettävä resoluutio, joka on suurennuskerroin x haluttu resoluutio (esimerkiksi 12 x 300 = 3600).
Skannausvalmis aineisto	Analoginen aineisto, joka on valmisteltu siten, että se on mahdollista tietojensa ja fyysisen olomuotonsa puolesta digitoida.
Säilytyspäätös	Kansallisarkisto päättää säilytyspäätöksessä, mitkä julkishallinnon asiakirjatiedot on säilytettävä pysyvästi, mikä on niiden säilytysmuoto, ja mitkä analogiset ilmentymät hävitetään tai säilytetään digitoinnin jälkeen. Tässä yhteydessä tarkoitetaan erityisesti päätöstä, jossa määrätään säilytetäänkö alkuperäinen analoginen asiakirja digitoinnin

	jälkeen.
Säilytysyksikkö	Analogisen aineiston säilyttämisen perusyksikkö, jolla on arkistotunnus. Säilytysyksiköllä tarkoitetaan sitä fyysistä välinettä, joka suojaa aineiston analogista ilmentymää.
Tallekappale	Tallekappale tarkoittaa digitointiprosessissa tuotettua digitaalista ilmentymää, joka on teknisiltä ominaispiirteiltään laadukkain digitointiprosessissa tuotettu ilmentymä. Käytetään pitkäaikaissäilytyksessä ns. ”master” – kappaleena.
TAR – paketti	Digitoinnissa syntyneen siirtopaketin muoto, jota sovelletaan siirrossa tuotannosta vastaanottavaan järjestelmään.
Viranomaisvalmisteltu aineisto	Analoginen aineisto, joka on viranomaisen toimesta valmisteltu siihen pisteeseen, että se mahdollistaa Kansallisarkiston käsittelysuunnitelman laadinnan
Ydintieto	Lähde/tietovaranto, missä ylläpidetään tietoa, jota voidaan hyödyntää muissa järjestelmissä sekä prosesseissa

2 Viranomaisvalmistelu (Prosessi 1.)



Kuva 2-1: Viranomaisvalmistelun prosessin vaiheet ylätasolla.

2.1 Taustoitus

Viranomaiset ovat saattaneet aineistonsa siirtokuntoon pysyvää säilyttämistä varten analogisessa muodossa noudattaen Kansallisarkiston antamia määräyksiä ja ohjeita. Arkistolakiin 831/1994 perustuvassa siirtomääräyksessä, *Viranomaisten pysyvästi säilytettävien analogisten asiakirjojen luovuttaminen arkistolaitokseen* (AL/22704/07.01.01.00/2011), ja erillisessä siirto-oppaassa on ohjeistettu keskeiset viranomaisvalmisteluun kuuluvat toimet.

Viranomaisvalmisteluun on kuulunut erilaisia vaiheita, kuten aineiston järjestämistä, seulontaa, kunnostamista, suojaamista arkistokelpoisiin säilytysvälineisiin sekä luettelointia. Viranomaisen on vastannut siirtomuuton järjestämisestä ja siinä tarvittavien suoja- ja muun materiaalin, kaluston sekä henkilökunnan hankinnasta. Viranomaisen on vastannut esimerkiksi aineiston sijoittamisesta paikalleen hyllyyn tai muuhun vastaanottajan niille osoittamaan paikkaan. Vastuu aineistoista on siirtynyt Kansallisarkistolle siinä vaiheessa, kun se on tarkastanut ja todennut tehdyt viranomaisen toimet ohjeistuksen mukaisiksi.



Valtaosa viranomaisten analogisten aineistojen siirtovalmistelusta perinteistä säilyttämistä varten muuttuu valmisteluksi digitointia ja analogisten ilmentymien hävittämistä varten. Toimenpiteet ja tarvittavat resurssit riippuvat aineiston muodosta ja kunnosta sekä valmistelun ja digitoinnin toteutustavasta. Kansallisarkiston tulee päivittää nykyinen siirtomääräys ja siirto-opas vastaamaan tulevia tarpeita. Seuraavaksi kuvataan viranomaisvalmistelun peruseriaatteen, joita tulevassa määräyksessä ja ohjeistuksessa tarkennetaan.

2.2 Kansallisarkiston uusi säilytyspäätös

Kansallisarkisto päättää erillisessä säilytyspäätöksessä viranomaisen pysyvästi säilytettävän asiakirjatiedon digitoinnin jälkeisestä säilytysmuodosta (analoginen ja/tai digitaalinen). Päätöksessä määritellään, millä asiakirjatiedolla on arvoa myös analogisina ilmentyminä. Asiakirjatiedon eri säilytysmuodot ymmärretään sen ilmentyminä – ei kopioina. Kansallisarkiston toimivalta määrätä asiakirjatiedon pysyvästä säilyttämisestä ja säilytysmuodosta perustuu arkistolakiin (8 § ja 14 a §). Säilytyspäätös korvaa aikaisemman Kansallisarkiston antaman seulontapäätöksen, mikäli säilytysmuoto on ollut analoginen. Tästä syystä säilytyspäätös tarvitaan myös asiakirjatiedolle, josta on tehty aiemman arkistolain mukainen seulontapäätös.

Valtaosa viranomaisten hallussa olevasta asiakirjatiedosta on rajattu joko pysyvään tai määräajaiseen säilytykseen jo Kansallisarkiston aiemmilla seulontapäätöksillä. Analogisten ilmentymien hävittämisen mahdollistavan säilytyspäätöksen antamisen yhteydessä Kansallisarkisto määrää tarvittaessa myös arvonmääritystä vaille olevan asiakirjatiedon pysyvästä säilyttämisestä. Pääsääntönä on, että uudessa säilytyspäätöksessä ei muuteta Kansallisarkiston viranomaisen asiakirjatiedolle aiemmin tekemää arvonmääritystä kuin poikkeustapauksissa.

Arkistolain muutosta ennakoiden vuoden 2016 lopussa alkoi analogisten ilmentymien materiaalisuuden arvottamiseen liittyvien kategorioiden ja niiden kriteerien suunnittelutyö. Kaikki pysyvään säilytykseen määrätty asiakirjatieto on merkittävää asiakirjallista kulttuuriperintöä. Osalla asiakirjatiedosta on arvoa myös analogisina ilmentyminä, minkä vuoksi kyseinen asiakirjatieto on tarkoitus säilyttää digitoinnista huolimatta myös analogisessa säilytysmuodossa. Säilytysmuodon arvoon vaikuttavat monet tekijät, kuten asiakirjojen todistusvoimaisuus tai niiden ainutlaatuisuus analogisina ilmentyminä. Uudessa säilytyspäätöksessä päätetään, mihin kategoriaan viranomaisten aineistot sijoitetaan. Alustavan luonnoksen mukaan aineistot jaetaan seuraaviin kategorioihin:

- **Kategoria 1: Kulttuurihistoriallisesti arvokkaat aineistot**
 - Kategoriaan sijoitetaan aineistot, joilla arvioidaan olevan kulttuurihistoriallista arvoa. Aineistojen mahdollinen digitointi tapahtuu erillään massadigitointiprosessista. Aineistot säilytetään digitoinnista huolimatta myös analogisina ilmentyminä. Niiden materiaalisia piirteitä ei saa muuttaa, vaan ne pyritään säilyttämään konservoinnin keinoin.

- **Kategoria 2: Asiakirjatieto, jonka todistusvoimaisuus edellyttää analogisten ilmentymien säilyttämistä**

- Kattegoriaan sijoitetaan aineistot, joiden oikeudellisen todistusvoiman katsotaan edellyttävän analogisten ilmentymien säilyttämistä. Kattegoriaan sijoitetaan myös aineistot, joiden asiakirjatietoa ei voida siirtää sähköiseen muotoon menettämättä todistusvoimaisuutta. Aineistojen mahdollinen digitointi tapahtuu erillään massadigitoointiprosessista, jos niiden skannaamiseen ei ole laitteistoa massadigitoinnin piirissä tai jos aineistoja ei voida digitoida menettämättä asiakirjatietoa. Aineistot säilytetään mahdollisesta digitoinnista huolimatta myös analogisina ilmentyminä. Niiden materiaalisia piirteitä voidaan muuttaa (esimerkiksi sidosten purkaminen digitoinnin valmistelussa), jos se ei vaaranna asiakirjatietoa.
- Kattegoria 3: Digitoinnin jälkeen vain digitaalisena säilytettävä asiakirjatieto
 - Kattegoriaan sijoitetaan aineistot, jotka voidaan digitoida ja niiden analogiset ilmentymät hävittää digitoinnin jälkeen. Suurin osa massadigitoinnin piiriin tulevista aineistoista sijoitetaan kattegoriaan 3.

Kansallisarkisto laatii säilytyspäätöksen viranomaisen Kansallisarkistolle tekemän säilytyspäätösesityksen perusteella. Viranomaisen pitää eritellä esityksessä tiekarttasuunnittelussa tai muuten (viranomaisen oma arvio) digitoitavaksi suunniteltu analoginen aineisto sellaisella tarkkuustasolla (esimerkiksi sarjakokonaisuuksittain), että Kansallisarkisto voi sen pohjalta arvioida seuraavien kohtien 1–4 perusteita ja tarvetta. Lisäksi viranomaisen pitää perehtyä Kansallisarkiston aineistoa koskeviin aiempiin päätöksiin osana esityksen valmistelua.¹

Säilytyspäätöksen keskeiset kohdat tiivistetysti:

1. Kansallisarkisto päättää pysyvästi säilytettävän asiakirjatiedon analogisten ilmentymien säilyttämisestä tai hävittämisestä digitoinnin jälkeen:
 - a. Kulttuurihistoriallisesti arvokkaat aineistot (kattegoria 1) säilytetään digitoinnin jälkeen myös analogisessa muodossa. Aineistoja ei digitoida massadigitoinnin piirissä. Viranomainen esittää säilytyspäätösesityksessä arvion analogisten ilmentymien kulttuurihistoriallisesta arvosta.
 - b. Oikeudellisen todistusvoimansa vuoksi asiakirjatieto (kattegoria 2) säilytetään digitoinnin jälkeen myös analogisessa muodossa. Viranomainen esittää säilytyspäätösesityksessä arvion oikeudellisesta todistusvoimasta.
 - c. Jos asiakirjatietoa ei voida siirtää sähköiseen muotoon (kattegoria 2), säilytetään se analogisessa muodossa.

¹ Kansallisarkiston virastokohtaiset asiakirjatiedon pysyvää säilyttämistä koskevat päätökset ja yleiset määräykset koko julkishallinnolle, valtion- /kuntasektorille, tietyille hallinnonalalle tai virastoryhmälle. Esimerkiksi valtionhallinnon asiakirjoja koskeva arkistolaitoksen yleinen seulantapäätös 3.8.2010 (AL/19273/07.01.01.00/2008).



2. Kansallisarkisto päättää antamallaan säilytyspäätöksillä tai määräyksillä analogisen aineiston korvaamisesta viranomaisen hallussa olevalla syntysähköisellä asiakirjatiedolla tai digitoiduilla asiakirjatiedoilla. Analogiset ilmentymät voidaan hävittää, mikäli digitaalinen ilmentymä vastaa tietosisällöltään kattavasti ja eheästi analogista ilmentymää (esimerkiksi tietojärjestelmästä paperimuotoon tulostetut asiakirjatiedot).
3. Mikäli viranomaisen massadigitointiin siirrettävien aineistojen asiakirjatiedoista ei ole Kansallisarkiston aiemmin tekemää seulontapäätöstä, Kansallisarkisto päättää tarvittaessa näiden asiakirjatietojen pysyvästä säilyttämisestä osana säilytyspäätöstä.
4. Poikkeustapauksissa Kansallisarkisto voi muuttaa uudella säilytyspäätöksellä viranomaisen tallessa olevaa asiakirjatietoa joko pysyvästi tai määräajan säilytettäväksi. Uudessa päätöksessä voidaan esimerkiksi muuttaa ennen vuotta 2017 annetussa päätöksessä pysyvään säilytykseen määrättyä asiakirjatietoa määräajan säilytettäväksi, jos Kansallisarkisto katsoo, että ko. asiakirjatiedolla ei ole enää pysyvää säilytysarvoa. Vastaavasti aikanaan määräajan säilytettäväksi määrättyjä ja vielä tallessa olevia asiakirjoja voidaan muuttaa pysyvästi säilytettäväksi, jos Kansallisarkisto katsoo, että niillä on tutkimuskäyttöä. Aiempia päätöksiä arvioidaan vain poikkeustapauksissa. Pääsääntönä on, että säilytyspäätöksessä ei muuteta Kansallisarkiston asiakirjatiedoille aiemmin tekemää arvonmääritystä (säilytysaikaa).

Säilytyspäätös on edellytys aineiston ottamisesta massadigitointiin. Massadigitoinnin toteuttamisen ja tiekartan aikataulun kannalta Kansallisarkiston tulee pystyä valmistelemaan säilytyspäätökset lyhyessä määräajassa, jotta viranomaiset voivat käynnistää aineistojen valmistelutoimenpiteet. Uuteen säilytyspäätösmenettelyyn siirrytään siinä vaiheessa, kun Kansallisarkisto on saanut valmiiksi tarvittavat päivitykset siirtomääräykseen ja -ohjeisiin. Muutokset astuvat voimaan aikaisintaan vuoden 2018 aikana.

2.3 Viranomaisen valmistelutoimet aineistolle

Massadigitoinnin ensimmäisessä osaprosessissa viranomaisen valmistelee aineiston logistista siirtoa varten. Viranomaisvalmistelu vaatii nykyistä siirtomenettelyä vähemmän toimenpiteitä. Viranomaisen vapautuu useimmista aineiston fyysisen siirron toteutukseen liittyvistä tehtävistä. Osa tehtävävaatimuksista keventyy verrattuna aikaisempiin vaatimuksiin. Seulonta tehdään sarja- tai aineistokokonaisuustasolla ja asiaryhmytyksen mukaan arkistoiduista asia-kirjoista asiaryhmytasolla. Seulontaa ei pääsääntöisesti uloteta säilytysyksikön sisälle: yksittäisiä määräajan säilytettäviä asiakirjatyyppisiä ei poisteta pysyvästi säilytettävän aineiston joukosta, mikäli pysyvästi ja määräajan säilytettävät yksittäiset asiakirjat on sijoitettu säilytysyksikköön keskenään sekaisin, eikä niitä ole koottu selkeisiin säilytysajan perusteella tunnistettaviin asiakirjanippuihin. Sidoksista puretuista asiakirjoista ei myöskään seulota pois yksittäisiä määräajan säilytettäviä asiakirjatyyppisiä. Jos asiakirjatieto säilytetään digitoinnin jälkeen myös analogisessa muodossa, seulonnassa on huomioitava resursoitu säilytystila.

Tulevat muutokset viranomaisvalmistelussa koskevat ennen muuta säilytysyksiköitä, joiden osalta vaatimukset kevenyvät säilytysmuodosta riippumatta. Viranomaisen pitää suojata analoginen aineisto siirron ajaksi niin, että aineisto ei pääse kuljetuksen aikana vaurioitumaan eikä sekoittumaan muiden



aineistojen kanssa. Säilytysyksiköiden ei pidä olla arkistokelpoisia. Säilytysyksiköihin voidaan luetteloinnin yhteydessä merkitä tunniste esimerkiksi lyijykynällä ennen kuin lopulliset viivakoodilliset nimiöt ovat saatavilla.

Luettelointi ja kuvailu tehdään arkistoyksikkötasolle Kansallisarkiston luettelointijärjestelmään (AHAA). Tarkoituksena on, että säilytysyksikkö olisi myös arkistoyksikkö. Viranomaisen suorittamalta luetteloinnilta vaaditaan vähimmäismetatiedot, jotka mahdollistavat tässä raportissa esitetyn prosessin läpiviemisen. Vähimmäismetatiedoilla tässä yhteydessä tarkoitetaan esimerkiksi seuraavia: nimeke, tunniste, alkuvuosi-loppuvuosi sekä käyttörajoitustietoja arkistoyksikkötasolle. Kun luettelointitiedot on hyväksytty, saadaan Kansallisarkiston luettelointijärjestelmästä säilytysyksiköitä identifioivat viivakoodilliset nimiöt. Viranomaisen kiinnittää viivakoodit säilytysyksiköihin.

Mikäli viranomainen on ehtinyt valmistella aineistonsa Kansallisarkiston aiemman siirtomääräyksen ja ohjeistuksen mukaisesti, ei näihin toimenpiteisiin edellytetä muutoksia. Viranomaisen tehtäväksi jää lähinnä siirron logistisen seurannan edellyttämien identifiointitietojen (viivakoodit) lisääminen säilytysyksiköihin. Aineistot, joihin ei vielä ole kohdistettu toimenpiteitä, pitää valmistella uuden siirtomääräyksen ja -ohjeistuksen mukaisesti.

Alustavan suunnitelman mukaan viranomainen varmistaa erillisen tarkistuslistan avulla, että kaikki tarvittavat valmistelutoimet on tehty. Viranomainen toimittaa tarkistuslistan Kansallisarkistolle tarkastettavaksi ja hyväksyttäväksi. Toimenpiteiden päätyttyä viranomaisen hallussa on viranomaisvalmisteltu aineisto. Viranomaisvalmistelun valmiuden todentamisen jälkeen aineistokokonaisuuden metatiedot tuodaan ensimmäisen kerran AHAA-järjestelmästä tuotannonohjausjärjestelmään (Tila: *Viranomaisvalmistelu valmis*).

Tuotannonohjausjärjestelmän rooli tässä vaiheessa on seuraava:

- viranomaisvalmistellun aineiston määrän seurantatyökalu
- työnjohdollinen työkalu, joka mahdollistaa käsittelysuunnitelman laadinnan käynnistämisen

Alla olevassa taulukossa (Taulukko 2-1.) on eriteltyä viranomaisvalmistelun työvaiheita tiivistetysti.

Taulukko 2-1: Viranomaisvalmistelun työvaiheita

Viranomaisvalmistelun työvaiheita
Aineiston kartoittaminen ja yhteen kokoaminen (sijoituspakat ja määrät)
Säilytyspäättöksen mukainen seulonta
Rajaaminen ja järjestäminen
Suojaaminen säilytysyksiköihin (ei tarvitse olla arkistokelpoisia) ja salassapitomerkitöjen tekeminen
Luettelointi ja kuvailu arkistoyksikkötasolle sisältäen käyttörajoitustiedot

Rekisteriselosteen tuottaminen (tarvittaessa)
Nimiöinti eli säilytysyksiköiden identifiointitietojen lisääminen (viivakoodit Kansallisarkiston luettelointijärjestelmästä)
Työvaiheiden läpikäynti tarkistuslistan avulla

2.4 Käsittelysuunnitelma

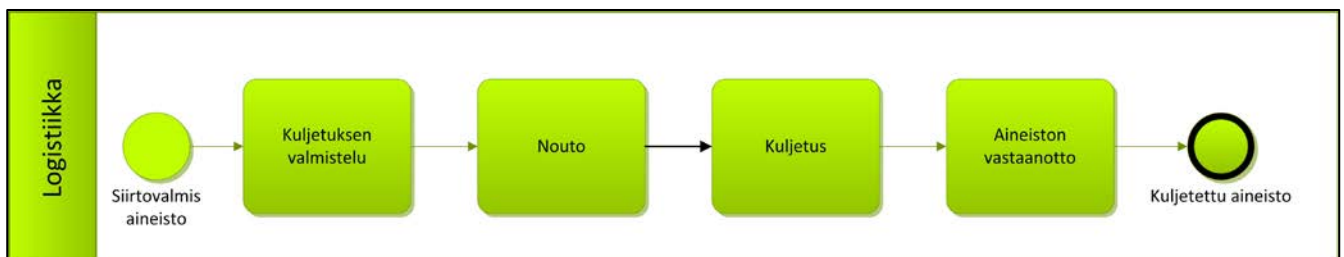
Kun viranomaisen on tehnyt tarvittavat toimenpiteet ja aineiston metatiedot on hyväksytty AHAA:ssa, Kansallisarkiston asiantuntija laatii aineistolle käsittelysuunnitelman digitoinnin toteuttamista varten. Käsittelysuunnitelma laaditaan yhteistyössä viranomaisen kanssa. Sen laadinta voidaan käynnistää samanaikaisesti viranomaisvalmistelun yhteydessä. Osa tiedoista kerätään automaattisesti eri ohjelmistotiedoista (esim. AHAA, AHJ eli aineistohallintajärjestelmä) ja osa vastaavasti laaditaan manuaalisesti. Osa havainnoista ja tiedoista on koko siirrettävää aineistokokonaisuutta koskevaa ja yksiselitteisesti määriteltävissä (mittaluku, kyllä/ei). Osa käsittelysuunnitelmanhavainnoista tehdään aineistotyypeittäin tai kokonaisuuksittain.

Käsittelysuunnitelma sisältää ne viranomaisvalmistellun aineiston keskeiset tiedot, jotka pitää huomioida digitoinnin valmistelun ja toteutuksen suunnittelussa. Käsittelysuunnitelman avulla pitää pystyä arvioimaan digitoinnissa tarvittavaa laitteistoa, henkilöstön määrää ja ajan käyttöä sekä kustannuksia. Toisin sanoen sen avulla pyritään mahdollistamaan digitointipaikan mahdollisimman kustannustehokas toiminta. Sen pohjalta laaditaan myös logistiset siirtosuunnitelmat. Käsittelysuunnitelmaan kirjataan aineistoon liittyviä tietoja, kuten tallennusalusluokitus, kunto, tietosisällön kieli, onko aineisto käsinkirjoitettua vai konekirjoitettua sekä mahdolliset liitemateriaalit.

Käsittelysuunnitelman laadinnan ja tallentamisen jälkeen viranomaisvalmistellun aineiston tila päivitetään tuotannonohjausjärjestelmään (Tila: *Siirtovalmis aineisto*). Tuotannonohjausjärjestelmän rooli tässä vaiheessa on seuraava:

- siirtovalmiin aineiston määrän seurantatyökalu
- työnjohdollinen työkalu, joka mahdollistaa kuljetusten valmistelun käynnistämisen

3 Logistinen siirto (Prosessi 2.)



Kuva 3-1: logistisen prosessin vaiheet ylätasolla

3.1 Taustoitus

Tiekartan mukaisen aikataulun toteuttaminen aineistojen siirtojen osalta viranomaisilta digitointitiloihin vaatii logistista suunnittelua. Digitointitiloissa aineiston välivarastointitilat ovat rajalliset, joten siirrot tulee aikatauluttaa järkevästi ja muodostaa helposti seurattava logistinen prosessi. Logistiikka on vahvasti tuotantosidonnainen ja sen päätehtävä on varmistaa, että digitoitavaa analogista aineistoa on aina saatavilla. Toisaalta on varmistettava, että digitointia odottavan aineiston määrä pysyy kohtuullisena, jotta digitointipaikassa olevaan aineistomäärään kohdentuvat tietopalvelupyynnöt eivät kasva.

Logistiikan suunnittelu kattaa työkalut tuotannonohjauksen optimointiin ja varmistaa, että digitoitavia aineistoja on saatavilla oikea-aikaisesti ja -määräisesti digitointipaikassa käsittelysuunnitelman mukaisesti. Ajojärjestelijä suunnittelee ja valmistele kuljetuksen, jonka pohjalta järjestelmä luo määrämuotoisia kuljetustilauksia. Kuljetustilaukset sisältävät nouto-ohjeet sekä ajankohdan kuljetuksen suorittamiseksi. Ne voidaan lähettää sähköpostin liitteenä tai suoraan tuotannonohjausjärjestelmästä. Tilauksia voidaan seurata reaaliaikaisesti kuorma-autossa sijaitsevan GPS-paikantimen avulla. Ajojärjestely luo tilannekuvaa kuljetustilausten tilasta, sekä havainnollistaa kuljetusten toteutumista.

3.2 Logistiikan prosessikuvaus

Tässä kappaleessa määritellään logistinen prosessi, prosessissa käsiteltävä ja tuotettava tieto, käytettävät varastonseuranta- ja tietojärjestelmät sekä teknologia- ja turvallisuusvaatimukset prosessin läpiviemiseksi.

Logistiikkaprosessi voidaan käynnistää, kun aineisto on tuotannonohjausjärjestelmässä tilassa ”Siirtovalmis aineisto”, eli sille on suoritettu edellisten prosessinvaiheiden mukaiset tehtävät. Termillä tarkoitetaan aineistoja, jotka odottavat kuljetusta digitointipaikkaan.

Logistisen prosessin vaiheet ovat seuraavat:

Kuljetuksen valmistelu (digitoiva taho – logistiikkapäällikkö/ajojärjestelijä)

- Suunnitellaan kuljetus
- Määritetään kuljetuksen ajoitus
- Allokoidaan rullakot
- Määritetään kuljetuksen erityisvaatimukset
 - aineistojen saatavuus viranomaisilla
 - pakkaustarvikkeet
 - erikoiskuljetusalustat
 - olosuhteet (suojaaminen, lämpötila, kosteus, tärinä, kolhut, valo)
- Luodaan pakkaus- ja nouto-ohje
- Luodaan ja lähetetään kuljetustilaus

Kuljetuksen valmistelun jälkeen aineiston tila päivitetään tuotannonohjausjärjestelmään (Tila: *Kuljetusvalmis aineisto*). Sen rooli tässä vaiheessa on seuraava:

- Noutovalmiin aineiston määrän seurantatyökalu
- Työnjohdollinen työkalu, joka mahdollistaa aineiston noudon

Aineiston nouto

- Luetaan pakkaus- ja nouto-ohje
 - osoitetaan ja varmistetaan siirtovalmiista ja siirrettävästä aineistosta
- Varmistetaan että siirtoon varattu kalusto on kunnossa
- Siirretään aineistot (säilytysyksiköt) soveltuviin kuljetusyksiköihin (rullakoihin) ja luetaan säilytysyksiköiden viivakoodit sekä kuljetusyksikön identifioiva viivakoodi
- Suljetaan kuljetusyksikkö
 - kuitataan käsipäätteeltä kuljetusyksikkö valmiiksi
 - lukitaan kuljetusyksikkö
 - suojataan kuljetusyksikkö (huputus)
- Tässä vaiheessa aineiston tila päivitetään tuotannonohjausjärjestelmässä (Tila: *Kuljetuksessa*). Lisäksi aineisto saa tuotannonohjausjärjestelmässä sijaintitiedon: (*Rullakossa X*)
- Siirto ajoneuvoon
- Pakkauslistan luonti ja lähetys

Kuljetuksen suorittaja

- Kuorma-auton ominaisuudet/varusteet:
 - GPS-seuranta
 - paloilmoituslaite
 - paljontorjuntalaitteisto
 - kuormatilan olosuhdeanturointi (ilmankosteus ja lämpötila)
 - kuormatilan jäähdytys/lämmitys (ilmastointilaitte)
 - varashälytin
 - takalaitanostin

Aineiston vastaanotto (digitoiva taho)

- Tarkistetaan kuorma ja verrataan sitä pakkauslistaan
- Ilmoitetaan vauriot, mahdolliset väärinkäytökset, poikkeamat jne.
- Luetaan viivakoodit kuljetusyksiköistä ja siirretään vastaanottovaraston varastopaikalle
 - määritetään varastopaikka (luetaan viivakoodi sopivasta varastopaikasta)
 - siirretään kuljetusyksikkö varastopaikalle odottamaan digitointia/fyysistä valmistelua
- Tässä vaiheessa aineiston tila päivittyy tuotannonohjausjärjestelmässä (Tila: *Aineisto vastaanotettu*). Lisäksi aineiston sijaintitieto päivittyy: (*varastopaikassa Y – rullakossa X*).

3.3 Logistiikan tietojärjestelmävaatimukset ja ohjelmistoarkkitehtuuri

Logistiikan hallinta vaatii eri ohjelmistojen välistä tietojensiirtoa ja logistiikan käyttöliittymää. Massadigitoinnin prosessikuvauksen toiminnoissa edellytetään tuotannonohjausjärjestelmän ja Kansallisarkiston aineistohallintajärjestelmän (AHJ) ja arkistoaineiston metatietojärjestelmän (AHAA) välistä rajapintaa ja tietoliikennettä. Logistiikan hallinnan käyttöliittymä toimisi tuotannonohjausjärjestelmässä. Toiminnot tuotannonohjausjärjestelmässä logistiikan suunnitteluun ja kuljetusten hallintaan sisältyvät vähintään seuraavat ominaisuudet:

- Kuljetustilausten hallinta, luonti (ja suunnittelu)
- Ajojärjestely ja kalustoresursointi (rullakkopooli)
- Nouto- ja vastaanottotapahtumien hallinta
- Poikkeusilmoitukset

Liityntäpinnat muiden ohjelmistojen kanssa:

- Varastonhallinta
- Poikkeamien käsittely
- Seuranta ja tiedonkeruu

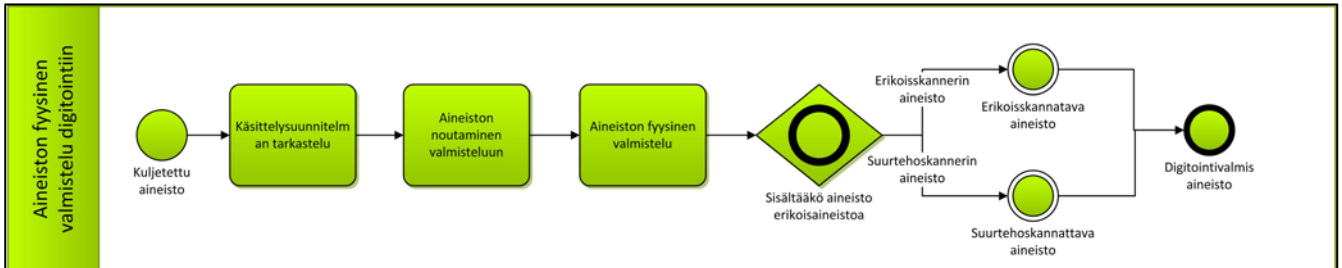
Rullakoiden allokoinnilla hallitaan täysien ja tyhjen rullakoiden sijoittelua ja käyttökapasiteettia, koska rullakoita sijaitsee prosessin monessa eri vaiheessa, kuten esimerkiksi:

- Kuorma-autossa (täysi/tyhjä)
- Valmistelussa viranomaisilla (tyhjä)
- Tehtaalla eri paikoissa (täysi/tyhjä)
- Muut

Logistiikan käyttöliittymään pitää käsittää erillinen mobiilikäyttömahdollisuus, jolla hallitaan nouto- ja vastaanottotapahtumia. Mobiilikäyttöliittymällä päästään käsiksi etänä mm. nouto-ohjeisiin sekä tilaushallintaan. Sillä luetaan viivakoodeja säilytysyksiköistä ja rullakoista. Aineistojen valmistelussa on säilytysyksiköihin tarroitettu viivakooditunnisteet, jotka sisältävät määritellyt metatiedot säilytysyksikön tarkemmasta sisällöstä. Viivakoodi luetaan ja ohjelmistoa (käyttöliittymää) käytetään siihen soveltuvalla kämmentietokoneella (PDA). Kämmentietokone luovutetaan ”kuljetusliikkeen” haltuun, joten sen käytön koulutukseen ja perehdyttämiseen on kiinnitettävä huomiota. Aineistojen noudossa säilytysyksiköt luetaan määritettyyn rullakkoon kuljetettavaksi ja digitointitehtaan vastaanotossa rullakot puretaan kuormasta lukemalle rullakot vastaanotetuiksi tehtaalle.

Logistiikkaohjelmistoon rakennetaan aineistoihin ja kuljetuksiin liittyvien poikkeamien kirjaamisominaisuus. Tämä tarkoittaa sitä, että ohjelmiston kautta voidaan ilmoittaa logistiikkaprosessiin liittyviä havaintoja, esimerkiksi että, kuormassa on havaittu puutteita, rullakko on viallinen tai aineisto on vahingoittunut.

4 Aineiston fyysinen valmistelu digitointiin (Prosessi 3.)



Kuva 4-1: Aineiston fyysisen valmistelun vaiheet ylätasolla

4.1 Taustoitus

Digitoinnin valmistelulla varmistetaan, että digitointi sujuu tehokkaasti analogisen aineiston sisältämää informaatiota vaarantamatta. Puutteellisesti valmistettu aineisto pysäyttää prosessin ja saattaa vahingoittaa asiakirjoja tai skannereita – tai pahimmillaan aiheuttaa asiakirjan koko informaation puuttumisen.

Valmistelutoimenpiteissä pitää ottaa huomioon Kansallisarkiston antama säilytyspäätös sekä aineistotyyppi, jotka asettavat rajaukset tehtäviin toimenpiteisiin. Jos asiakirjatieto säilytetään pysyvästi myös analogisessa muodossa, pitää analogiseen aineistoon kohdistuvien toimenpiteiden ja lisättävien materiaalien olla arkistokelpoisia. Valmistelutoimet pyritään pitämään mahdollisimman vähäisinä ja suoraviivaisina erityisesti analogisessa muodossa hävitettävän aineiston kohdalla.

Aineistot valmistellaan skannattavaksi pääosin suurtehoskannerilla, mistä syystä digitoitavien asiakirjojen fyysisten piirteiden pitää täyttää tietyt reunaehdot. Niiden pitää olla esimerkiksi irrallisina arkkeina. Suurtehoskannerilla voidaan skannata vähintään 65x85 mm kokoisia ja maksimissaan 297x420 mm eli A3-kokoisia arkkeja. Sidokset voidaan skannata purkamisen jälkeen. Paperimateriaalin lisäksi valmisteltavien aineistojen joukossa voi olla muutakin materiaalia, mikä pitää ottaa huomioon valmistelumenetelmien valinnassa. Jos aineistoa ei voida skannata suurtehoskannerilla, valmistellaan aineisto erikoisskannerille. Valmistelun yhteydessä tehdään viime kädessä päätökset, millä laitteella aineisto skannataan. Digitoinnin valmistelussa vaaditaan hyvää materiaalien sekä skannauslaitteiston tuntemusta.

Digitoinnin valmistelun vaiheita paperipohjaiselle aineistolle on kuvattuna tiivistetysti taulukossa 4-1. Taulukossa vihreät kentät kuvaavat vaiheet, jotka tehdään valmistelussa aina. Valkoisissa kentissä olevat vaiheet tehdään tarpeen mukaan. Tarkemmin digitoinnin valmistelua kuvataan seuraavissa valmistelua koskevissa alaluvuissa.

Taulukko 4-1: Massadigitoinnin valmistelu tiivistetysti paperipohjaiselle aineistolle

Massadigitoinnin valmistelu tiivistetysti paperipohjaiselle aineistolle

Valmisteluun toimitetaan kuljetusrullakko, josta säilytysyksikkö kerrallaan valmistellaan aineistoa.

Säilytysyksiköstä luetaan viivakoodi, minkä jälkeen aineistolle tulostetaan ohjausarkki, joka toimii aineiston tunnisteena prosessissa (erilaiset, esimerkiksi eriväriset, digitoinnin jälkeen säilytettäville ja hävitettäville).
Aineisto poistetaan säilytysyksiköstä, joka voidaan hävittää tai kierrättää.
Digitoinnin estävä lika poistetaan.
Niitit ja muut vastaavat liittimet poistetaan.
Aineiston järjestys ja oikea suunta tarkistetaan silmämääräisesti.
Aineistossa olevat erikokoiset liitteet valmistellaan.
Digitoinnin estävät repeämät ja irronneet palaset paikataan.
Digitoinnin estävät taitokset suoristetaan.
Informaation päällä olevat vanhat paikkaukset poistetaan.
Digitoinnin estävät liiman ja teippien jäämät poistetaan.
Sidosmuodossa oleva aineisto puretaan käsin tai leikkurilla, jos se ei vaaranna informaatiota.
Aineiston joukossa oleva erikoisaineisto erotetaan suurtehoskannerille menevästä aineistosta, ja sille tulostetaan erikoisaineiston ohjausarkit (erilaiset, esimerkiksi eriväriset, digitoinnin jälkeen säilytettäville ja hävitettäville).
Aineisto kulkee metallinpaljastimen läpi, millä varmistetaan, että metallisia liittimiä ei ole jäänyt aineistoon.
Aineisto tasataan digitointia varten käyttäen tähän tarjolla olevaa laitteistoa.
Aineisto laitetaan odottamaan skannausta.

4.2 Yleisohjeet kaikille aineistotyypeille

Digitoinnin valmistelua ohjaa luvussa 2.4 kuvattu käsittelysuunnitelma, jonka perusteella valmistelun työjohto antaa tarvittavan ohjeistuksen valmistelutoimenpiteisiin. Etukäteissuunnittelusta huolimatta valmistelussa voi tulla vastaan yksittäisiä asiakirjoja, joiden valmistelutoimenpiteet pitää päättää tapauskohtaisesti.

Valmistelussa analogisen manifestaation jatkokäsittelyä ohjaamaan asetetaan erilaisia ohjausarkkeja. Erilaiset arkit ja niiden käyttötarkoitus on lueteltu taulukossa 4-2: Valmistelussa arkistoyksikköön sijoitettavat arkit.

Taulukko 4-2: Valmistelussa arkistoyksikköön sijoitettavat arkit

Arkki	Tarkempi arkin kuvailu	Käyttötarkoitus	Pakollisuus
Ohjausarkki	Viivakoodin ja aineiston identifioivia tietoja sisältävä paperilomake.	Asetetaan jokaisen arkistoyksikön ensimmäiseksi paperiksi. Digitoinnin jälkeen myös analogisessa muodossa säilytettävän aineiston ohjausarkki on erilainen, esimerkiksi erivärinen, kuin digitoinnin jälkeen analogisessa muodossa hävitettävän aineiston ohjausarkki, joka ohjaa aineistojen käsittelyä skannauksen jälkeen.	Pakollinen
Laatuarkki	Valmistelussa analogisen manifestaation yhteyteen liitettävä digitointitekniisiä testitauluja	Laatuarkin tarkoitus on todentaa digitaalisen manifestaation laatua arkistoyksikkökohtaisesti. Laatuarkin tarpeellisuutta tarkastellaan jatkossa.	Tarvittaessa

	sisältävä paperilomake.		
Erikoisaineiston ohjausarkki	Viivakoodin sisältävä paperilomake, jonka avulla yhdistetään eri skannausprosesseissa skannatut aineistot alkuperäiseen kontekstiinsa.	Kun aineistoa ei ole mahdollista skannata suurtehoskannerilla, asetetaan erikoisaineiston ohjausarkki arkistoyksiköstä toiseen linjastoon ohjatun analogisen ilmentymän kohdalle. Digitoinnin jälkeen myös analogisessa muodossa säilytettävän erikoisaineiston ohjausarkki on erilainen, esimerkiksi erivärinen, kuin digitoinnin jälkeen analogisessa muodossa hävitettävän erikoisaineiston ohjausarkki.	Tarvittaessa

Aineiston fyysinen valmistelu alkaa aineiston toimittamisesta digitoinnin valmistelun työtilaan, jossa aineisto valmistellaan säilytysyksikkö kerrallaan. Tässä vaiheessa tuotannonohjausjärjestelmään päivittyy aineiston tila (Tila: *Siirretty valmisteluun*).

Kun aineiston valmistelu käynnistetään, säilytysyksiköstä luetaan viivakoodi, minkä jälkeen tulostetaan ohjausarkki. Samalla myös aineiston tila päivittyy jälleen tuotannonohjausjärjestelmään (Tila: *Digitoinnin valmistelu*). Arkistoyksikkö otetaan pois säilytysyksiköstä, joka voidaan valmistelun tässä vaiheessa hävittää tai kierrättää. On tärkeää pitää huoli siitä, että arkistoyksikkö ei hajoa ja sekoitu muiden aineistojen kanssa. Arkistoyksiköiden päälle laitettavat ohjausarkit erottavat arkistoyksiköt toisistaan.

Jotta aineiston luettavuus digitaalisessa muodossa säilyy, tarkistetaan silmämääräisesti, että aineisto on esimerkiksi järjestyksessä ja oikein päin. Aineiston visuaalisen tarkastelun yhteydessä tarvittavat valmistelutoimenpiteet tarkentuvat.

Skannausteknisistä syistä valmistelun yhteydessä erotetaan analogisesta aineistosta yksittäiset erikoisaineistot. Jos aineiston joukossa on materiaalia, joka ei sovellu suurtehoskannerilla skannattavaksi, päivitetään tieto tuotannonohjausjärjestelmään. Tämän jälkeen tulostetaan erikoisaineiston ohjausarkit (2 kpl), joista toinen asetetaan erikoisaineistoon ja toinen digitoitavan yksikön siihen kohtaan, josta erikoisaineisto on otettu. Jos valmistelussa tulee vastaan arkistoyksikkö, joka pitää käsitellä kokonaisuudessaan muulla kuin suurtehoskannerilla, aineisto valmistellaan normaalisti. Aineistoon liitetään tavallinen ohjausarkki.

Valmistelussa voi tulla vastaan aineistoa, jota ei voida digitoida, koska sen informaatio on täysin tuhoutunut. Tapauskohtaisesti pitää arvioida, onko mitään enää tehtävissä. Jos informaatiota ei voi pelastaa, pitää aineiston metatietoja päivittää Kansallisarkiston luettelointijärjestelmään AHAA:seen. Vastaavasti, jos digitoinnin valmistelun yhteydessä aineisto vaurioituu ja informaatio osittain tai kokonaan menetetään, pitää luettelointijärjestelmään päivittää asianmukaiset merkinnät.

Jos valmisteltavan aineiston joukosta löytyy massadigitoinnista pois rajattua aineistoa, kuten valokuvien negatiiveja, diakuvia, lasinegatiiveja, mikrofilmejä, mikrokortteja tai AV-aineistoa, pitää aineisto siirtää erilliseen säilytystilaan ja päivittää metatietoja. Arkistoyksikköä ei voida analogisessa muodossa digitoinnin jälkeen hävittää, jos siihen kuuluu massadigitoinnista pois rajattua aineistoa, jonka digitointi pitää suunnitella erikseen.



Aineisto kuljetetaan metallinpaljastimen läpi, millä varmistetaan se, että aineistoon ei ole jäänyt metallisia liittimiä, jotka voivat vaurioittaa skannauslaitteita. Valmisteltu analoginen aineisto siirretään odottamaan digitointia. Suurtehoskannerille menevät asetetaan omaan pinoon (maksimissaan 4 arkistoyksikköä päällekkäin) ja erikoisaineistot omaan pinoon sekä tehdään tilapäivitykset tuotannonohjauksjärjestelmään (Tila: *Digitointivalmis aineisto*).

4.3 Tarkentavat toimenpiteet

4.3.1 Paperiarkit sisältäen kortistokortit

Jos paperiarkit ovat niin likaisia, että niitä ei voida digitoida, ne pitää imuroida pehmeää harjasuulaketta käyttäen arkkinipun reunoilta tai jokainen arkki erikseen riippuen likaisuudesta. Vaihtoehtoisesti ne voidaan puhdistaa mikrokuituliinalla varoen repimästä aineistoa. Hauraan aineiston kohdalla tulee olla varovainen, ettei paperinpalasia tartu puhdistusvälineeseen. Jos aineiston joukosta löytyy homeista aineistoa, käsitellään koko säilytysyksikön sisältö vetokaapissa, jolloin valmistelijalla pitää olla erityinen henkilösuojavarustus. Jos homepuhdistettu analoginen aineisto säilytetään myös digitoinnin jälkeen, pitää säilytysvälineissä ja metatiedoissa olla maininta homevauriosta. Tarkoituksena on kuitenkin viranomaisvalmistelun yhteydessä varmistaa, että homeista aineistoa ei päädy massadigitointiprosessiin.

Erilaiset liittimet, kuten niitit, klemmarit, kuminauhat, muovitaskut, turhat välilehdet, ylimääräiset irtopaperit, pitää poistaa ennen digitointia. Liittimien poistaminen ei saa vaurioittaa informaatiota. Jos liittimet poistetaan esimerkiksi leikkaamalla arkkinipun kulmat pois, on huomioitava, että tämä näkyy mustana alueena digitaalisessa ilmentymässä. Menetelmä voi tehostaa valmistelua mutta vaikuttaa digitoinnin lopputulokseen. Metallinpaljastimen avulla voidaan varmistaa, että aineistoon ei jää vahingossa metallisia liittimiä, jotka estävät digitoinnin ja voivat vaurioittaa laitteistoa.

Aineiston joukossa olevat, kahtia taitetut A2-koon arkit leikataan taitoksesta auki (työvälineenä voi käyttää esimerkiksi paperiveistä tai leikkuria), lukuun ottamatta arkkeja, joissa informaatio jatkuu koko aukeamalla (esim. yli sivun menevä taulukko). Jos ylisivun menevät arkit ovat suurempia kuin A3, käsitellään ne kuten muut erikoisaineistot.

Aineistoon kiinniliimattujen informaatiota sisältävien liitteiden liimauksen tulee olla kunnossa, jotta ne pysyvät paikoillaan skannauksenkin ajan. Irtoamassa olevat liitteet pitää kiinnittää liimalla. Jos liitteet peittävät informaatiota, siirretään ne tyhjään kohtaan sivua tai esimerkiksi arkin tyhjälle taustapuolelle. Jos asiakirjassa ei ole tyhjää tilaa, kiinnitetään liite erilliselle paperiarkille varsinaisen asiakirjan jälkeen tai vaihtoehtoisesti otetaan valokopio liitteestä. Tyhjät liitteet pitää poistaa.

Hyväkuntoisia teippejä tai muita vanhoja paikkauksia ei tarvitse poistaa, jos ne eivät peitä informaatiota. Irtoilevat, repsottavat teipit ja paikat pitää poistaa varovasti aineistoa vahingoittamatta. Digitoinnin jälkeen analogisessa muodossa hävitettävässä aineistossa olevat tahmeat, vanhojen teippien



liimajäämät peitetään esimerkiksi uudella teipillä², jotta arkit eivät tartu kiinni toisiinsa. Digitoinnin jälkeen myös analogisessa muodossa säilytettävässä aineistossa mahdollisesti olevat tahmeat liimajäämät pitää poistaa mekaanisesti. Tarpeen mukaan liimakohtaa voi käsitellä vedellä tai etanolilla kostutetulla pumpulitikulla. Pysyvästi myös analogisessa muodossa säilytettäviin aineistoihin ei saa lisätä teippejä, jotka ajan kuluessa muodostavat ongelmia analogisen aineiston säilyvyydelle.

Pahat repeytymät ja irronneet palaset, jotka haittaavat informaation lukemista ja digitointia, paikataan teipillä silloin, kun analoginen ilmentymä hävitetään digitoinnin jälkeen. Analogisessa muodossa säilytettävät paikataan liimakalvonauhalla ja lämpölusikalla.

Jos paperiarkin reuna on hyvin repaleinen tai rei'itetty, paperiarkit voivat tarttua toisiinsa kiinni. Rispaantunut reuna leikataan pois. Tämä pitää tehdä vaarantamatta informaatiota.

Digitointia haittaavat paperiarkkien taitokset suoristetaan lämpölusikalla tai vastaavalla, jotta digitaaliseen ilmentymään ei synny varjoja. Tarpeen mukaan pitää käyttää suojapaperia lämpölusikan ja arkin välissä.

Joissain tapauksissa alkuperäisen asiakirjan digitointia parempi vaihtoehto on digitoida alkuperäisestä asiakirjasta otettu kopio. Näin voidaan menetellä, jos esimerkiksi asiakirjan sivut ovat takertuneet kiinni muovitaskuun niin, ettei teksti olisi enää luettavissa muovitaskun poistamisen jälkeen. Vaihtoehtoisesti edellä mainitun tyyppisiä asiakirjoja voidaan digitoida erikoisskannerilla.

Valmistelun loppuksi, ennen aineiston siirtämistä digitointipinoihin, asiakirjapinot tasataan digitointia varten käyttäen tähän tarjolla olevaa laitteistoa. Pinojen tasauksen (esimerkiksi tasauskulman) määrittelee käytössä oleva skannausteknologia.

4.3.2 Sidokset

Massadigitoinnin yhteydessä sidokset puretaan, jos se ei vaaranna informaatiota. Sidoksen purussa ei saa menettää yhtään informaatiota – ei esimerkiksi yhtään kirjainta eikä numeroa. Kun sidos on purettu, tehdään muut tarvittavat valmistelutoimenpiteet kuten muillekin paperiarkeille.

Sidokseksi ei katsota jollakin helposti poistettavissa olevalla tavalla (kuten nauhalla, muovi- tai metallihaarukalla) yhteen liitettyjä asiakirjoja. Näissä tapauksissa asiakirjoja yhteen liittävä mekanismi poistetaan ja asiakirjat käsitellään paperiarkkeja koskevan perusohjeen mukaisesti.

Purkamisen tarkoituksena on saada siisti nippu irtoarkeja, joissa ei ole lankoja, liimaa tai reikiä. Purussa sidoksesta irrotetaan kannet ja sidos leikataan tarpeen mukaan pienempiin nippuihin ennen sidoksen leikkaamista käsin tai leikkurilla. Purun yhteydessä on varmistettava, että kansirakenteissa ei ole informaatiota.

Jos sidoksen tietosisältö ei ole suoraan näkyvässä sidoksen ollessa auki, pitää purkamisessa olla erityisen varovainen. Sidos on voitu tehdä niin, että tietosisältö on jäänyt sitomisessa käytettyjen materiaalien

² Digitoinnin valmistelussa käytettävän teipin pitää olla läpikuultavaa, ohutta ja mattapintaista. Teippiä ei saa laittaa useita kerroksia päällekkäin.

alle tai tietosisältö muuten jää syvälle taitoksen sisälle. Jos asiakirjassa oleva teksti näkyy sidoksen ollessa auki, mutta teksti tai osa siitä jää leikkurin leikkuurajan alle, ei sidosta voida purkaa. Tällöin sidos skannataan erikoisaineistona mastoskannerilla.

Eräissä tapauksissa arkkien yhteen sitomisessa käytetty liima on saattanut kiinnittää arkit toisiinsa siten, että tekstiä on jäänyt liiman alle. Liiman alle jäänyt teksti on menetetty jo aikanaan, kun arkit on sidottu sidokseksi. Nämä voidaan leikata leikkurilla, mutta menetetyistä tiedoista pitää olla maininta metatiedoissa luettelointijärjestelmässä.

Jos sidoksessa on paljon tyhjiä sivuja, poistetaan suurin osa tyhjiä sivuista. Tällöin kuitenkin jätetään muutama tyhjä sivu sen jälkeen, kun teksti loppuu tai muutama tyhjä sivu ennen kuin teksti alkaa tai jatkuu uudestaan. Jos sivut on numeroitu, kirjoitetaan ensimmäiselle tyhjälle sivulle lyijykynällä hakasulkeisiin seuraavasti [tyhjiä sivuja, s. xx–xx].

4.3.3 Kartat ja piirustukset

Kartoille ja piirustuksille tehdään tarvittavat toimenpiteet kuten paperiarkeille ottaen huomioon, että niiden materiaalina saattaa olla paperin lisäksi erityismateriaaleja, kuten kuultopaperi, kuultokangas, sinikopio. Materiaaleista esimerkiksi kuultokangas ja kuultopaperi eivät kestä kosteutta, joten tästä syystä muun muassa teippien ja liimajäänteiden poistamisessa näiden materiaalien kohdalla ei saa käyttää nestemäisiä liuottimia.

Kartat ja piirustukset saattavat olla muun aineiston joukossa taiteltuna, jolloin ne pitää suoristaa lämpöpusikalla tai vastaavalla. Tarpeen mukaan pitää käyttää suojapaperia lämpöpusikan ja aineiston välissä. Paperimateriaalia lukuun ottamatta muut erityismateriaalit voivat esimerkiksi kiillottua tai sulaa, jos lämpöä kohdistetaan suoraan materiaalille. Tämä tulee ottaa huomioon myös aineistoja paikattaessa, jos paikkaamisessa käytetään liimakalvonauhaa ja lämpöpusikkaa.

Rullatut kartat ja piirustukset skannataan erikoisaineistona, koska niiden suoristaminen suurtehoskanneria varten on työlästä sekä aikaa vievää ja jossain tapauksissa mahdotonta.

4.3.4 Valokuvat

Valokuville tehdään tarvittavat toimenpiteet kuten paperiarkeille ottaen huomioon muun muassa seuraavat poikkeukset ja rajaukset.

Valokuvia käsiteltäessä pitää käyttää hansikkaita, jotta valokuvaan ei tule sormenjälkiä, jotka näkyvät digitaalisessa ilmentymässä. Hansikkaiden pitää olla puuterittomia. Niissä ei saa olla näppyliitä, jotka voivat naarmuttaa kuvapintaa.

Jos valokuvat ovat hyvin likaisia, ne voidaan puhdistaa kuten muutkin paperiarkit. On kuitenkin varottava naarmuttamasta niiden pintaa. Pehmeä, naarmuttamaton mikrokuituliina on hellävaraisin vaihtoehto.

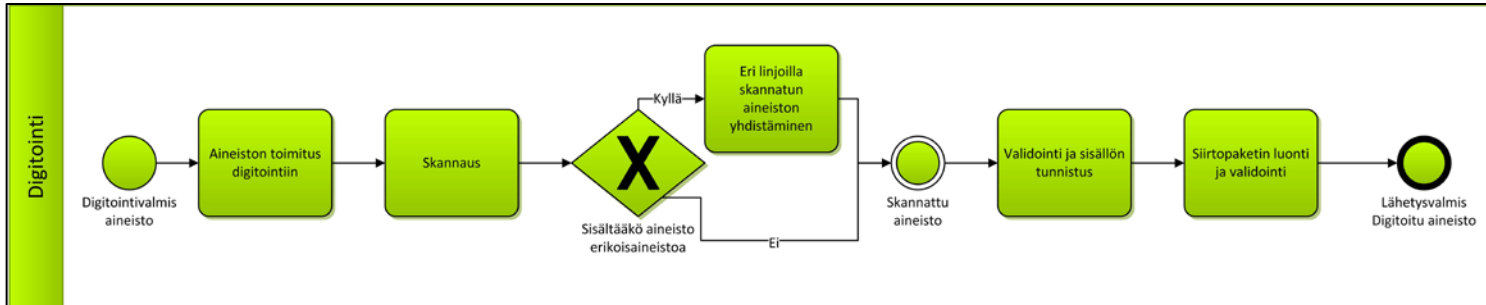
Valokuvien käsittelyssä ei saa käyttää nestemäisiä liuottimia, kuten vettä, joka vaurioittaa valokuvan kuva-aineista. Kosteus vaurioittaa valokuvien rakennetta ja se voi irrottaa esimerkiksi valokuvan pohjamateriaalin, kuten paperin, päällä olevan kuvaemulsion.

Pahat repeytymät ja irronneet palaset, jotka haittaavat informaation lukemista ja digitointia, on paikattava teipillä silloin, kun analoginen ilmentymä hävitetään digitoinnin jälkeen. Analogisessa muodossa säilytettävät paikataan liimakalvonauhalla ja lämpölusikalla taustapuolelta käyttäen vielä erikseen suojapaperia.

Digitointia haittaavat taitokset suoritetaan lämpölusikalla tai vastaavalla taustapuolelta käyttäen suojapaperia.

Valokuvat voivat olla kiinnitettynä muun muassa kartonkiin tai albumiin. Kiinnitystapana on voitu käyttää esimerkiksi liimaa, klemmareita, niittejä, teippejä tai silkkipusseja. Valokuvat voivat olla albumissa muovitaskuissa ja sivuilla on kuvien lisäksi muuta informaatiota. Tapauskohtaisesti pitää valita tarvittavat valmistelutoimenpiteet, jotta aineistot voidaan digitoida suurtehoskannerilla. Tarpeen mukaan aineisto skannataan erikoisskannerilla.

5 Digitointi (Prosessi 4.)



Kuva 5-1: Digitoinnin prosessi ylätasolla

5.1 Taustoitus

Tässä raportissa digitointi pitää sisällään digitointivalmiin analogisen aineiston muuntamisen sähköiseen muotoon (skannaus bittikarttakuvaksi), kuvan laadun validoinnin, kuvan sisällön tunnistuksen OCR -menetelmin, ja siirtopaketin muodostuksen. Tässä luvussa digitointia tarkastellaan pääsääntöisesti toiminnan osalta. Tietotekninen infrastruktuuri esitetään tarkemmin luvussa 8. Kansallisarkiston vaatimukset digitointiprosessille ja siitä syntyville tiedostoille kuvataan erikseen luonnoksessa Kansallisarkiston vaatimuksista hävittämiseen tähtäävään digitointiin.

Jotta aineiston skannaus voidaan suorittaa, on aineiston oltava tuotannonohjausjärjestelmässä tilassa: *Digitointivalmis aineisto*. Digitointivalmiuden saavuttaakseen on analogisen aineiston käytävä läpi prosessikokonaisuuden vaiheet 1-4. Skannaus tarkoittaa toimia, jolla analoginen aineisto muunnetaan sähköiseen muotoon käyttäen välineinä aineiston sähköiseen muotoon muuntamiseen tarkoitettua teknologiaa ja ohjelmistoratkaisuja.

5.2 Laitteistot

Tässä raportissa toimintaa tarkastellaan taulukossa 5-1 esitettävän skannausteknologian näkökulmasta.

Taulukko 5-1: Skannausteknologia

Laite	Teknologia	Käyttötarkoitus
Suurtehoskanneri	Avorata-skanneri, jossa analoginen aineisto ei kulje umpinaisen kotelon läpi. Mahdollistaa yhtäaikaaisesti asiakirjan molempien puolien skannaamisen A3 -kokoon asti.	Suurtehoskannerille valmistellut irtoarkit. Pääsääntöinen skannaustapa.
Suurkokoskanneri	Läpisyöttöskanneri suurikokoiselle aineistolle, joka mahdollistaa maksimissaan ~140cm leveän analogisen aineiston skannaamisen. Mahdollistaa yhtäaikaisesti asiakirjan molempien puolien skannaamisen	Erikoisaineistot. Paperiarkki, joka on suurempi kuin A3.
Mastoskanneri	Skanneri, joka mahdollistaa sidosten skannaamisen purkamatta. Soveltuu myös muun analogisen aineiston skannaamiseen, joka ei sovellu muihin skannerityyppeihin.	Erikoisaineistot. Analoginen aineisto, jota ei ole mahdollista skannata muilla teknologioilla.
Dokumenttiskanneri	Pöytämallinen läpisyöttöskanneri. Mahdollistaa yhtäaikaisesti asiakirjan molempien puolien skannaamisen. Analogiset asiakirjat kulkevat umpinaisen kotelon läpi. Analogisiin asiakirjoihin kohdistuu yleisesti suurempi mekaaninen rasitus kuin avorata-skannerissa.	Erikoisaineistot. Suurtehoskannerin vaihtoehtoinen tapa toteuttaa alle A3 -koon irtoarkkien skannaus. Käytetään harkinnan mukaan.

Edellä esitettyssä taulukossa keskitytään nykytilan skannausteknologiaan, joka on yleisesti käytössä paperiaineiston skannauksessa. Selvityshankkeen yhteydessä toteutetussa markkinakartoituksessa havaittiin myös edellä esitettyjä innovatiivisimpia tapoja aineiston digitointiin, mutta näyttöä kyseisten ratkaisujen käyttöön otosta ei havaittu. Laitteiden kokoonpano on rakennettu tietokoneiden tulosten perusteella ja sillä on arvioitu pystyvän skannaamaan hyvin pitkälti se paperiaineisto, jota viranomaisissa tällä hetkellä on.

Laitteistojen osalta olennaisimpia asioita ovat seuraavat:

1. laitteen tarjoama tekninen valmius tuottaa vaatimusmäärittelyn mukaisia digitaalisia ilmentymiä
2. minkälaisia aineistoja laitteella on mahdollista skannata
3. laitteiston toimintavarmuus ja huoltosykli
4. huollon saavutettavuus
5. laitteen elinkaariennuste

Laitteistojen suhteen arvioidaan tapauskohtaisesti kustannustehokkainta tapaa toteuttaa skannausta myös siitä näkökulmasta, että onko kustannustehokkainta hankkia varalaitteita laiterikkotilanteita varten. Ohjelmistot, mitä skannauksessa käytetään, ovat usein laitesidonnaisia, mutta voivat olla myös erillisiä, toisen toimittajan, ohjelmistoja, jotka mahdollistavat tietyn laitekoonpanon toiminnan.



5.3 Skannaus (Prosessi 4.1)

5.3.1 Taustoitus

Skannausprosessin tavoite on tuottaa yhdellä skannauksella sekä tallekappale että tästä tuotettu käyttökappale.

Lähtökohtaisesti skannaus pyritään aina tekemään taulukossa 5-1 esitetyllä suurtehoskannerilla. Koska tämä ei teknologian asettamien reunaehtojen perusteella ole aina mahdollista, on analoginen arkistoyksikkö mahdollisesti jaettu valmistelussa useampaan skannauslinjastoon skannattavaksi.

5.3.2 Prosessi

Skannaus etenee pääsääntöisesti siten, että skannausoperaattori syöttää suurtehoskannerille, sitä varten valmistellun aineiston.³ Arkistoyksikön sisältöä edeltää aina ohjausarkki ja sitä seuraava laatuarkki. Skannerin viivakoodinlukija lukee valmistelussa asetetusta ohjausarkista viivakoodin, jonka perusteella sen pitää identifioida skannattava analoginen arkistoyksikkö tuotannonohjausjärjestelmästä saatavien tietojen mukaan. Kun ohjausarkki on kulkenut viivakoodinlukijan ohi, muuttuu skannattavan arkistoyksikön tila tuotannonohjausjärjestelmässä (Tila: *Skannauksessa*)

Suurtehoskanneri tuottaa aineistosta talle- ja käyttökappaleet niille määritettyihin hakemistorakenteisiin ja ohjaa nämä levyjärjestelmille odottamaan jatkotoimenpiteitä. Jatkotoimenpiteillä tarkoitetaan aina tallekappaleiden validointia ja lähes aina sisällön tunnistamista käyttökappaleista. Mikäli skannattu arkistoyksikkö ei sisällä erikoisaineiston ohjausarkkeja, on skannaustyö valmis ja arkistoyksikön tila päivittyy tuotannonohjausjärjestelmässä (Tila: *Arkistoyksikkö skannattu*). Jos taas arkistoyksikkö sisältää erikoisaineiston ohjausarkkeja, tapahtuu näiden automaattinen yhdistäminen tässä prosessin vaiheessa. Toisin sanoen, mikäli aineistoa on skannattu eri linjastoilla, pitää nämä yhdistää ohjelmallisesti alkuperäiseen kontekstiinsa korvaamalla erikoisaineiston ohjausarkki sitä vastaavalla kuvatiedostolla asiakirjasta. Kun yhdistäminen on tehty, siirtyy tällainenkin arkistoyksikkö tilaan: *Arkistoyksikkö skannattu*.

Suurtehoskannerissa pitää olla useampia lajittelulokeroita, jotta valmistelussa pinotut arkistoyksiköt voidaan automaattisesti ohjausarkkiensa perusteella jakaa omiin lokeroihinsa. Tämä ominaisuus on tärkeä etenkin silloin, kun analoginen manifestaatio pitää säilyttää sen digitoinnin jälkeen. Lajittelua voidaan hyödyntää myös hävitettävän aineiston suhteen, sillä ohjausarkkeja pitää pystyä hyödyntämään analogisen aineiston käsittelyssä digitoinnin jälkeen (Prosessi 6.). Arkistoyksiköiden lisäksi omaan lokeroonsa pitää ohjata laatuarkit, joita pitää pystyä käyttämään uudelleen. Muilla skannerityypeillä kokonaisuudessaan skannatut arkistoyksiköt pitää asetella skannauksen jälkeen siten, että arkistoyksikön ohjausarkin viivakoodi on luettavissa, vaikka analogisia manifestaatiota olisi pinossa useampia.

³ Poikkeuksen tähän aiheuttaa analoginen aineisto, joka olomuotonsa takia ei sovellu suurtehoskannerille. Tässäkin tapauksessa viivakoodin lukeminen suoritetaan laitteistoa ohjaavalla ohjelmistolla, mikäli mahdollista. Jos tämä ei ole mahdollista, käytetään erillistä lukulaitetta.



Skannausprosessin on kyettävä luomaan sekä talle- että käyttökappale. Siirtopaketin rakenne ja tiedostoformaattit on kuvattu luvussa 5.5. Skannauksessa käytettävän ohjelmiston on luotava ensin tallekappale ja tuotettava tallekappaleesta käyttökappale, jotta prosessissa voidaan varmistua siitä, ettei häviöllisesti pakattua tietoa muunneta myöhemmin häviöttömästi pakattuun muotoon. Edellä esitetty vaatimus syntyy myös siksi, että tallekappaleen ja käyttökappaleen rooli prosessissa eroaa toisistaan. Käyttökappaletta käytetään prosessissa sisällön tunnistuksen työkappaleena, kun taas tallekappaletta pyritään käsittelemään mahdollisimman vähän sen ensimmäisen tallennustapahtuman jälkeen.

Tallekappaleelle lasketaan heti ensimmäisen tallennuksen yhteydessä MD5-tarkistesumma, joka tallennetaan tuotannonohjausjärjestelmään. Digitoinnissa käytettyjen laitteiden ja ohjelmistojen on kyettävä jo tässä vaiheessa tuottamaan tallekappaleen metatietoihin digitaalisen ilmentymän luontia (skannausta) koskevia tietoja, jota myöhemmin käytetään hyväksi kirjoitettaessa digitaalisen tiedoston syntyhistoriaa kuvaavia metatietoja XML-tiedostoon. Tallekappaleiden kopioinnit, siirtämiset ja tallennukset on syytä minimoida. Skannauksen yhteydessä tallekappaleet voidaan siirtää suoraan levyjärjestelmälle, josta ne myöhemmässä vaiheessa toimitetaan vastaanottavalle järjestelmälle käyttökappaleiden sekä validointiprosessissa syntyneiden muiden tiedostojen kanssa.

Skannauksen jälkeen analogisen ja digitaalisen manifestaation käsittely jatkuu eri prosesseissa.

5.4 Validointi (Prosessi 4.2)

5.4.1 Taustoitus

Validoinnilla tarkoitetaan skannauksen jälkeen tapahtuvaa laadunvarmistusta. Validoinnissa esimerkiksi tarkastetaan skannausvaiheessa syntyneiden digitaalisten ilmentymien määritysten mukaisuus sekä digitaalisen manifestaation eheys. Validointi on osa laadunvarmistusta, joka on kokonaisuutena laajempi ja alkaa jo ennen skannausta laitteistojen ja ohjelmistojen toimivuuden todentamisella. Laadunvarmistukseen voi laskea myös datan eheyden varmistamisen prosessin eri vaiheissa. Tässä luvussa keskitytään kuvaamaan skannausprosessin tuotoksena syntyneiden digitaalisten ilmentymien validointia mahdollisten skannausvirheiden näkökulmasta.

Massadigitoinnin tuotannon käynnistämisen vaiheessa pitää jokaista prosessin vaihetta tarkkailla huomattavasti tarkemmin, jotta prosessi saadaan luotettavaksi. Käynnistämisen jälkeen validointiin kuluvan henkilötöiden määrä on todennäköisesti aiempaa vähäisempi.

5.4.2 Prosessi

Jotta validointiprosessi voi käynnistyä, on arkistoyksiköiden oltava skannattu kokonaisuudessaan eli arkistoyksikön on oltava tuotannonohjausjärjestelmässä tilassa: *Arkistoyksikkö skannattu* ja kaikki prosessin edeltävät vaiheet on oltava hyväksytysti suoritettu.

Tallekappaleiden kuvanlaatua validoidaan otannalla. Alustavan arvio mukaan vähintään joka sadas tallekappale pitää validoida visuaalisella tarkastelulla. Mikäli digitoitu arkistoyksikkö sisältää alle 100 tallekappaletta, on ohjelmiston kyettävä ohjaamaan arkistoyksiköstä vähintään yksi tallekappale



operaattorille validoitavaksi. Jos tallekappaleiden validoinnissa havaitaan puutteellinen digitaalinen ilmentymä, merkitään se puutteelliseksi ja koko arkistoyksikkö skannataan tällöin uudestaan. Tuotannonohjausjärjestelmässä arkistoyksikön tila muuttuu: *Uudelleenskannattava*. Samalla varmistetaan, että kyse ei ole systemaattisesta virheestä.

Tallekappaleiden onnistuneen validoinnin jälkeen arkistoyksikön tila muuttuu tuotannonohjausjärjestelmässä (Tila: *Arkistoyksikön kuvanlaatu validoitu otannalla*). Skannausprosessin operaattori suorittaa jatkuvaa laaduntarkkailua skannauksen yhteydessä, mutta skannaustyö suurtehoskannereilla on niin nopeaa, että varsinaisten yksittäisten digitaalisten ilmentymien kuvanlaatua tässä yhteydessä ei voida varmentaa silmämääräisesti.

Sisällön tunnistus suoritetaan tallekappaleiden kuvanlaadun validoinnin rinnalla tai se voidaan vaihtoehtoisesti toteuttaa myös vasta sen jälkeen, kun tallekappaleiden kuvanlaatu on validoitu otannalla. Sisällöntunnistus edellyttää digitaalisia ilmentymiä, joiden tietosisältö on pääsääntöisesti konekirjoitettua. Käsinkirjoitetun tekstin sisältöä ei yritetä massadigitoinnin alkuvaiheessa tunnistaa HTR-menetelmillä. Sisällön tunnistus käynnistyy automaattisesti, kun käyttökappaleita ohjataan tunnistusohjelmalle käsiteltäväksi. Aineistokohtaisesti tunnistusohjelmalle annetaan tunnistuksen vähimmäisarvoja, mitä digitaalisesta ilmentymästä pitää tunnistaa. Kyseiseen arvoon vaikuttaa olennaisesti ilmentymien tietosisällön muoto, eli se, kuinka paljon digitoidussa asiakirjassa on ylipäättään tekstiä, mitä pystytään OCR-menetelmin tunnistamaan.

Kun OCR-ohjelmisto on suorittanut digitaalisen manifestaation sisällön tunnistuksen, siirtyy kyseinen arkistoyksikkö tuotannonohjausjärjestelmässä päivitettyyn tilaan (Tila: *Sisältö tunnistettu*). Mikäli tunnistusarvo jää alle asetetun alarajan, ohjataan digitaalinen ilmentymä operaattorille tarkastettavaksi. Operaattori tarkastaa, onko kuvanlaatu kunnossa, vai johtuiko arvon alittaminen jostain muusta syystä, kuten digitaalisen ilmentymän tietosisällön puutteista OCR-menetelmälle. Mikäli kuvanlaatu on hyväksyttävä, hyväksytään kuva ja tunnistustulos (on siis huomioitava, että tässä yhteydessä ei validoida tunnistuksen tulosta, vaan kuvanlaatua). Mikäli tunnistustuloksen heikkous on johtunut kuvanlaadusta, tallennetaan tästä tieto ja koko arkistoyksikkö ohjataan uudelleen skannattavaksi – toisin sanoen kyseisen arkistoyksikön kuvia ei tarkasteta tässä yhteydessä enempää. Kuvanlaadun tarkastuksen yhteydessä operaattori hyväksyy myös sisällön tunnistuksessa tyhjiksi sivuiksi tunnistetut digitaaliset ilmentymät. Tyhjien sivujen – yleensä kääntöpuolien – tunnistus pitää suorittaa siten, että voidaan olla varmoja, ettei kone poista ennen operaattorille ohjaamista yhtään digitaalista ilmentymää, joka sisältää minkäänlaista informaatiota.

Kun validointi on suoritettu onnistuneesti, siirretään prosessoitavana olleet käyttökappaleet ja niistä tuotetut OCR-tulokset (ALTOXML) tallekappaleiden yhteyteen ja tarkastetaan (tuotannonohjausjärjestelmä), että jokainen prosessin vaihe on suoritettu sekä kaikkia ilmentymiä on tuotettu samansuuruinen määrä. Mikäli tarkastus onnistuu, päivittyy arkistoyksikön tila tuotannonohjausjärjestelmässä (Tila: *Prosessivaiheet sekä kuvat ja tekstit tarkastettu*). Tila toimii käynnistimenä seuraavan vaiheeseen, missä tuotannonohjausjärjestelmä tuottaa MIX-skeeman mukaisen XML-tiedoston digitaalisista tallekappaleista ja validoi tuotetun XML-tiedoston. Tämän



vaiheen jälkeen prosessissa on muodostettu digitoinnin siirtopaketti, joka on kuvattu tarkemmin luvussa 5.5. Siirtopaketista muodostetaan TAR-paketti (tietoa ei kuitenkaan saa pakata) ja paketille lasketaan tarkistesumma. Tämän jälkeen validointivaihe on päätöspisteessään, ja digitaalisen manifestaation tila päivittyy tuotannonohjausjärjestelmään (Tila: *Lähetysvalmis digitoitu aineisto*).

5.5 Laadunvarmistuksen käytänteitä

Skannauksessa käytettäviä laitteita ohjataan asettamalla niille asetukset, jotka ohjaavat laitteen toimintaa. Nämä asetukset ovat laitekohtaisia ja erilaisia riippuen skannausprosessista sekä yksittäisestä laitteesta. Näistä toimista käytetään termejä kalibrointi ja profilointi – kyseisillä toimenpiteillä laitteisto saadaan käyttäytymään halutulla tavalla. Tämän jälkeen skannausohjelmistoon luodaan asetukset, jotka hyödyntävät edellä mainittujen toimintojen tuotoksia.

Yleisesti voidaan todeta, että suurtehoskannereissa ja dokumenttiskannereissa analogisen aineiston värien korkeatasoinen toistuvuus on hankalasti toteuttavissa. Tässä raportissa käsiteltävien aineistojen suhteen värien toistuvuus digitaalisessa ilmentymässä juuri sellaisina kuin ne esiintyvät analogisessa ilmentymässä ei ole niin merkityksellistä kuin se on historiallisen aineiston digitoinnissa. 24-bittisenä digitoitu analoginen ilmentymä sisältää kuitenkin profiloimattomassakin ympäristössä väri-informaatiota, joka voi olla tutkimuksen kannalta tulevaisuudessa merkityksellistä – myös silloin, kun värit eivät olisi toistuneet täydellisesti.

Kuvatiedostoihin pitää kuitenkin tallentaa (kuvan metatietoihin) ICC-profiili, joka myöhemmin mahdollistaa kuvan avaamisen sellaisena, miten se oli skannauksen toimesta muunnettu sähköiseen muotoon. Tässä tulee huomioida, että värit eivät vastaa alkuperäistä asiakirjaa, vaan siitä tehdyn digitaalisen ilmentymän värejä.

Värintoiston lisäksi skannausprosessin tuottamien kuvatiedostojen laatua pitäisi mitata myös muiden kuvanlaatuun vaikuttavien aspektien osalta. Tällaisia ovat muun muassa resoluutio, terävyys, kohina ja valkotasapaino. Mittauksia pitää toteuttaa tietyin väliajoin, jotta varmistutaan, että a) kuvanlaatu on riittävää ja b) kuvanlaatu pysyy tasaisena. Digitointituotannossa mittaus pitää suorittaa jokaisen työvuoron alkaessa. Koska digitoinnissa pitää käyttää massadigitointiin soveltuvia laitteita, on voitava kohtuudella luottaa näiden mittausten välillä syntyvän tasaista laatua. Edellä mainittua mittausta varten on käytettävissä erilaisia testitauluja, esimerkiksi Universal Test Target (UTT). Laitteen tuottamaa kuvaa testitaulusta verrataan ohjelmistojen avulla raja-arvoihin. Raja-arvoja on saatavilla valmiina (esimerkiksi Metamorfoze Preservation Imaging Guidelines⁴ ja FADGI⁵) tai ne voidaan luoda itse. Massadigitointia varten perinteisiä digitoinnin kuvanlaadun laatumäärittäjiä on laskettava, sillä massadigitoinnin mahdollistava laitteisto luo digitoituja ilmentymiä todella nopeasti.

⁴ Metamorfoze Preservation Imaging Guidelines. Viitattu 28.12.2017.

https://www.metamorfoze.nl/sites/metamorfoze.nl/files/publicatie_documenten/Metamorfoze_Preservation_Imaging_Guidelines_1.0.pdf

⁵ FADGI Guidelines. Viitattu 28.12.2017. <http://www.digitizationguidelines.gov/guidelines/digitize-OpenDice.html>

Jotta laatu olisi tasaista, on koko skannausinfrastruktuurin oltava ajan tasalla, kunnossa, puhdas ja asetettu toimimaan siten, että se tuottaa tuotoksia, jotka ovat määritysten mukaisia.

5.5.1 Skannereiden ja skannausohjelmistojen mahdollistama laadunvarmistus

Suurtehoskannereilla on yleisesti kaksi tapaa varmistaa sitä, että kahta asiakirjaa ei mene päällekkäin skannauksesta läpi. Ensimmäinen tapa on mekaaninen, mitä toteutetaan telojen avulla, jotka vetävät paperia paperiradalle. Toinen tapa on ns. ultraäänisensoritekniikka⁶, jotka mittavat skannerin läpimenevien objektien etäisyyttä toisistaan. Jälkimmäinen on yleisesti käytössä suurtehoskannereiden lisäksi myös dokumenttiskannereissa. Tämän pitää olla vaatimus digitointilaitteelle. Käytännössä on todettu, että paperiaineiston skannaaminen suurtehoskannerilla on varsin luotettavaa, kun kyse on ns. tuplasyöttöjen eliminoimisesta.⁷ Suurtehoskannerit mahdollistavat myös jokaisen asiakirjan etupuolen "leimaamisen" laitteeseen integroidulla mustesuihkutulostimella, joka voi sisältää asiakirjan identifioivia tietoja. Leimaa voisi käyttää todentamaan sitä, että kaikki skannattavaksi tarkoitetut asiakirjat on varmasti skannattu. Tämä tarkoittaisi sitä, että asiakirjat olisi skannattava uudelleen toisella laitteella, joka ohjelmallisesti luki jokaisesta asiakirjasta em. identifiointitiedon, ja ilmoittaisi, mikäli asiakirjasta ei kyseistä tietoa löydy.

Skanneria ohjaavan ohjelmiston (skannausohjelmiston) pitää mahdollistaa erilaisten käyttäjätasojen asettaminen. Käyttäjätasoilla pitää pystyä hallitsemaan valikoita siten, että esimerkiksi vain ylläpitäjä pystyy tekemään tietynlaisia valintoja ohjelmistossa. Toisin sanoen osa käyttäjistä ei pysty tekemään muutoksia, jotka voisivat vaikuttaa merkittävästi skannauksen lopputulokseen. Ylläpitäjän pitäisi pystyä tekemään muutoksia sisäverkossa menemättä yksittäisen skannausohjelmiston luokse.

Selvityshankkeessa todettiin, että skannausohjelmistojen kuvanlaadun laadunvarmistus on usein puutteellista tai sitä ei ole. Tämänkin takia on päädytty siihen, että jokaiseen arkistoyksikköön pitäisi todennäköisesti sijoittaa laatuarkki, joka todentaa ko. arkistoyksikön laatua. Laatuarkki pitäisi sisällään ainakin resoluutiota todentavan mittataulun, mittajanan ja harmaasävykiilan. Kansallisarkisto jatkaa edelleen selvittämistä miten keinoälyä voitaisiin hyödyntää automaattisessa kuvanlaadun validoinnissa.

5.6 Digitoinnissa tuotettavat tiedostot

Siirtopaketin pitää sisältää analogisesta manifestaatiosta tuotettuja digitaalisia ilmentymiä alla olevan taulukon mukaisesti:

Tiedoston rooli	Tiedostoformaatti	Pakkaus	Resoluutio	Bittisyvyys	Soveltuva ICC-profiili	Pakollisuus P=pakollinen E=Ehdollinen
Tallekkappale	TIFF 6.0	LZW	300 ppi	24	sRGB (tai	P

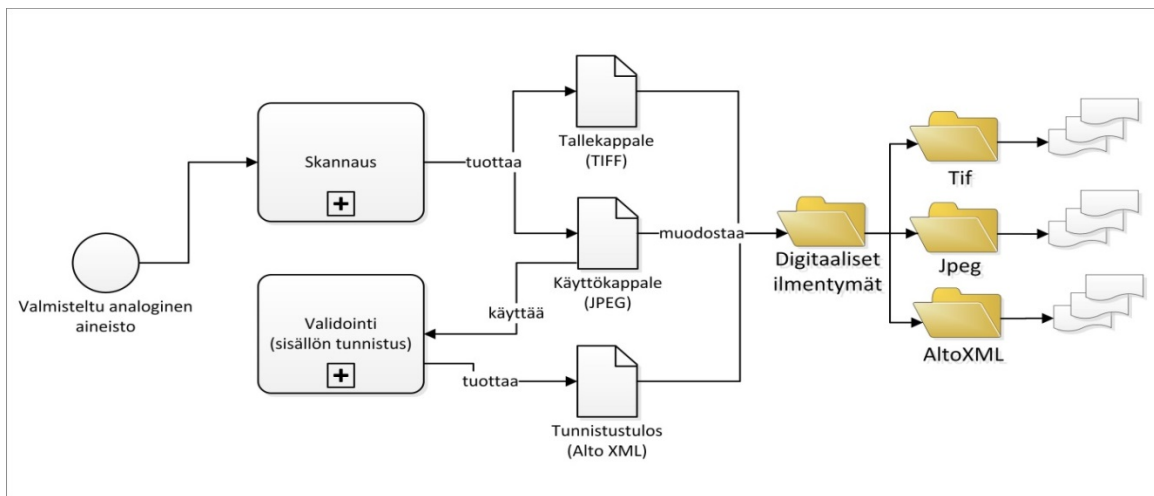
⁶ Wikipedia » Ultrasonic transducer. Viitattu 28.12.2017. https://en.wikipedia.org/wiki/Ultrasonic_transducer

⁷ Ton Boonen, Belastingdienst totesi, että he olivat digitoineet arkistoasiakkaalleen kilometrin paperiaineistoa kahden vuoden ajan käyttäen kahta erillistä suurtehoskanneria. Projektin aikana ei havaittu yhtäkään poikkeamaa skannereiden välillä.
4.1.2018

		(häviötön)			laajempi)	
Käyttökappale	JPEG	Jpeg-laatu 60%	300 ppi	24	sRGB	P
Tunnistettu teksti	ALTO ⁸ (3.0 tai 3.1)	-	-	-	-	E
Digitointiprosessia kuvaavat metatiedot	XML (MIX -skeema ⁹)	-	-	-	-	P

Jokaisesta tietosisältöä sisältävästä asiakirjan sivusta (etu- ja kääntöpuoli) pitää tuottaa saman tietosisällön omaavat talle- ja käyttökappaleet. Kuvien ICC-profiiliin pitää olla kuvan metatiedoissa (embedded). Kuvatiedostojen lisäksi jokaisesta tietosisältöä sisältävästä asiakirjan sivusta on tallennettava oma XML-tiedostonsa Analyzed Layout and Text Object (ALTO) -muodossa. Mikäli digitoidut asiakirjat ovat täysin käsinkirjoitettuja, ei tätä formaattia tarvitse tuottaa. Jokaisesta digitointiprosessissa tuotetusta tallekappaleesta pitää tuottaa MIX:in mukainen XML-tiedosto.¹⁰

Kuvassa 5-2: digitointi ja tuotokset esitetään prosessien tuotokset ja miten niitä käytetään seuraavien prosessien tuotoksien luomisessa. Digitaaliset ilmentymät mahdollistavat tuotannonohjausjärjestelmälle aiemmissa luvuissa esitetyllä tavalla MIX-skeeman mukaisen XML- tiedoston tuottamisen.



Kuva 5-2: digitointi ja tuotokset

5.6.1 Bittisyvyyden vaikutus kuviin

Bittikarttakuvien yhteydessä bittisyvydellä tarkoitetaan bittien lukumäärää pikseliä kohden. Jotta voidaan varmistua kuvien laadullisesta tasaisuudesta, pitäisi digitaalisen bittikarttakuvan olla aina 24bit,

⁸ The Library of Congress » Standards » ALTO. Kongressin kirjaston verkkosivu. Viitattu 19.12.2017.
<https://www.loc.gov/standards/alto/>

⁹ The Library of Congress » Standards » MIX. Kongressin kirjaston verkkosivu. Viitattu 20.12.2017.
<http://www.loc.gov/standards/mix//>

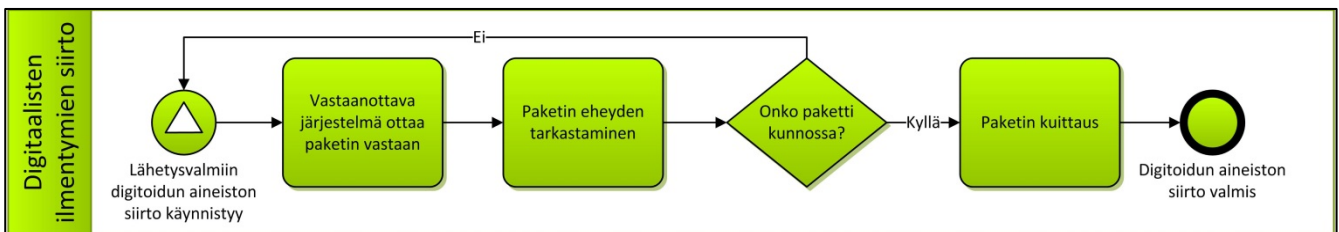
¹⁰ Siirtopaketin muoto ja siihen liittyvät vaatimukset on kuvattu tarkemmin dokumentissa: Kansallisarkiston vaatimukset hävittämiseen tähtäävän digitointiin (luonnos)

RGB-värikuva.¹¹ Edellä mainitusta tallekappaleesta voidaan halutessa tuottaa bittisyvydeltään vähäisempiä tiedostoja käyttökappaleiksi, mutta verkkoinfrastruktuurin ja nopeuksien jatkuvasti kehittyessä, on tämä jo nykyisellään turhaa huomioiden mahdolliset käyttötarpeet. Digitaalisen manifestaation todistusvoima alenee, kun sen bittisyvyttä lasketaan, jolloin se ei välttämättä ole enää tietosisällöltään täydellinen verrattuna analogiseen manifestaatioon. Vaarana on, että asiakirjan tietosisällön ja todistusvoimaisuuden kannalta keskeistä informaatiota häviää. Alla esimerkki bittisyvyyden vaikutuksista digitaalisen ilmentymän tietosisältöön. Kuvassa vasemmalta oikealle: 1. 24bit, RGB-kuva 2. 8bit, harmaasävykuva ja 3. 1bit, mustavalkoinen kuva.



Kuva 5-3: Bittisyvyyden vaikutus kuvan tietosisältöön

6 Digitaalisten ilmentymien siirto (Prosessi 5.)



Kuva 6-1: Digitaalisten ilmentymien siirto ylätasolla

6.1 Taustoitus

Tässä luvussa digitaalisen aineiston siirtoa tarkastellaan pääosin digitoinnin näkökulmasta, eikä tässä luvussa oteta kantaa siihen, miten aineistoa pitäisi välittää eteenpäin esimerkiksi pitkäaikaissäilytysjärjestelmään. Prosessin pitää olla automatisoitu ja ihmisen pitäisi joutua osallistumaan prosessin toteutukseen ainoastaan siinä tapauksessa, mikäli käytettävät järjestelmät joutuvat vikatiloihin. Prosessissa on kaksi järjestelmää: tuotannonohjausjärjestelmä sekä vastaanottava järjestelmä. Vastaanottavan järjestelmän vastuulle jää muun muassa digitaalisten ilmentymien luominen

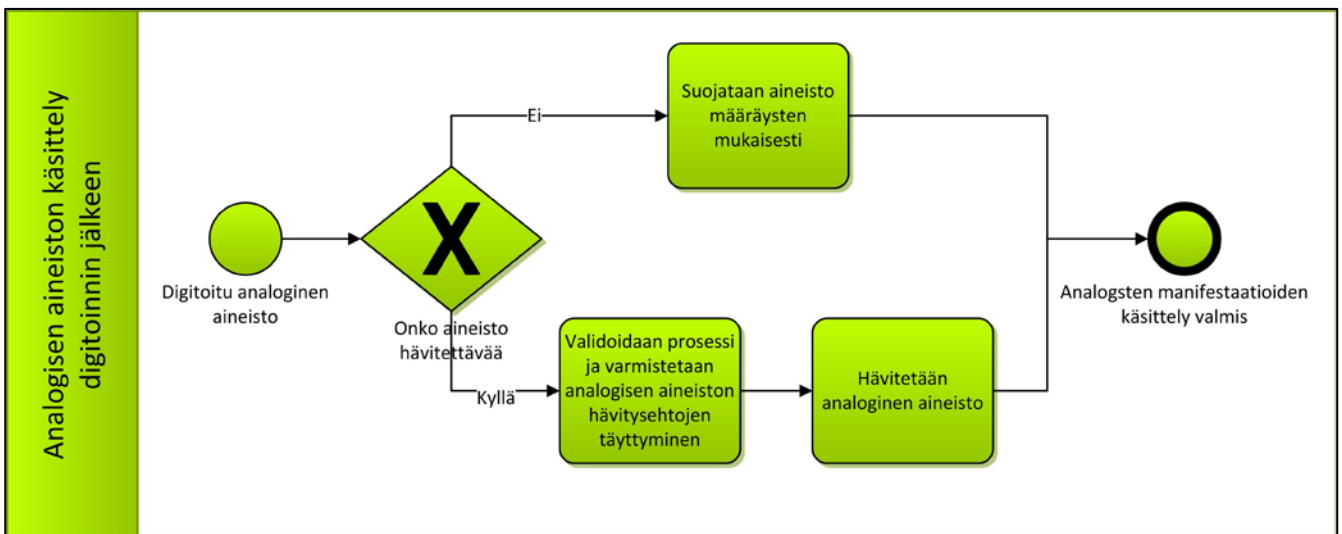
¹¹ bittien määrä pikselin kanava-arvossa = 8 8 8 ja kanava-arvojen määrä pikselissä = 3

AHAA-järjestelmään, käyttökappaleiden välittäminen käyttöliittymiin ja digitoinnissa muodostetun siirtopaketin (tai osan tästä) paketoiminen pitkäaikaissäilytysjärjestelmän vaatimaan muotoon (METS).

6.2 Prosessi

Jotta siirtopaketin lähettäminen käynnistyisi, on aineiston oltava tuotannonohjausjärjestelmässä tilassa *Lähetysvalmis aineisto*. Siirtopaketin siirtäminen käynnistyy automaattisesti, kun prosessi on tullut vaiheeseen, missä siirtopaketti on paketoitu. Tuotannonohjausjärjestelmä käynnistää siirron vastaanottavaan järjestelmään ja tallentaa tilan (Tila: *Siirtopaketti lähetetty*). Vastaanottava järjestelmä ottaa siirtopaketin vastaan ja tarkastaa paketin eheyden tarkistesumman avulla. Mikäli paketti on eheä, lähettää vastaanottava järjestelmä tiedon tästä tuotannonohjausjärjestelmälle ja digitaalisen manifestaation tila muuttuu (Tila: *Digitoidun aineiston siirto valmis*). Mikäli siirtoprosessissa on tapahtunut virheitä (esimerkiksi TAR-paketin tarkistesumma on muuttunut), pyytää vastaanottava järjestelmä paketin uudelleen niin kauan kunnes siirto on suoritettu eheästi. Kun kuittaus eheästi vastaanotetusta paketista saapuu tuotannonohjausjärjestelmälle, poistaa tämä siirtopaketin automaattisesti levyjärjestelmältään.

7 Analogisen aineiston käsittely digitoinnin jälkeen (Prosessi 6.)



Kuva 7-1: Analogisen aineiston käsittely digitoinnin jälkeen

7.1 Taustoitus

Analogisen aineiston käsittely digitoinnin jälkeen on riippuvainen viranomaisvalmistelun yhteydessä annetusta säilytyspäätöksestä. Tämä on prosessikokonaisuuksista viimeinen ja ajallisesti pitkäkestoisin. Se käynnistyy skannauksen jälkeen ja päättyy joko varoajan jälkeiseen hävittämiseen tai aineiston sijoittamiseen makasiiniin.

Tuotannon käynnistys- ja pilotointivaiheessa digitoidun ja hävittämistä odottavan analogisen aineiston hävittämisen varmuusaika on viisi kuukautta. Aineistomääränä tämä hävittämistä odottava kokonaisuus on noin 2 500 hyllymetriä (5 kk X 20,83 tpv X 24 hm). Tuotantovaiheen käynnistyttyä ja

laadunvarmistuksen toimintojen vakiinnuttua hävittämistä odottavan aineiston kokoa voidaan pienentää niin, että se vastaa esimerkiksi yhden kuukauden tuotannon määrää ja olisi aineistomäärältään noin 1 800 hyllymetriä (1 kk X 20,83 X 89 hm). Hävitettävissä olevien aineistojen varmuusaika määritellään myöhemmin, vasta sen jälkeen, kun tässä raportissa esitettyä prosessia on testattu käytännössä. Tässä luvussa ei käsitellä aineiston lopullista sijoittamista makasiiniin siinä tilanteessa, kun myös analoginen aineisto pitää säilyttää. Luvussa ei käsitellä sitä, missä ja miten analoginen aineisto konkreettisesti hävitetään.

7.2 Prosessi

Kun arkistoyksikkö on skannattu sille valitulla skannerilla, ohjaa sille suoritettavia jatkotoimenpiteitä analogisen aineiston fyysisessä valmistelussa asetettu ohjausarkki (Taulukko 4-2: Valmistelussa arkistoyksikköön sijoitettavat arkit).

7.2.1 Analoginen aineisto säilytetään digitoinnin jälkeen

Mikäli analoginen aineisto säilytetään prosessoinnin jälkeen, pitää se koteloida uudelleen, mahdolliset erikoisaineistot pitää yhdistää ohjausarkin avulla alkuperäiseen kontekstiinsa ja tiedot uusista säilytysyksiköistä pitää päivittää AHAA-järjestelmään. Kun edellä mainitut toimenpiteet on tehty, siirretään aineisto odottamaan siirtoa varsinaiseen säilytyspaikkaan ja tästä välitetään sijaintitietoja tuotannonohjausjärjestelmään, jossa aineiston tila jälleen päivittyy (Tila: *Säilytettävän analogisen aineiston käsittely valmis*). Koska säilytysyksiköiden määrät ja koot voivat muuttua prosessissa, on todennäköisesti lopullinen sijoituspaikka määriteltävä vasta, kun tiedot uusista säilytysyksiköistä ovat saatavilla.

Aineiston käsittely on kokonaisuudessaan valmis sen jälkeen, kun digitaalinen manifestaatio saa tuotannonohjausjärjestelmässä tilan *Digitoidun aineiston siirto valmis*. Mikäli tällöin myös aineiston käsittely on valmis, päivittyy arkistoyksikön tila tuotannonohjausjärjestelmässä (Tila: *Aineiston käsittely valmis*). Tämän jälkeen arkistoyksikön keskeiset tiedot (valitut elinkaaritapahtumat) välitetään tuotannonohjausjärjestelmästä AHAA:seen. Tämän jälkeen arkistoyksiköt siirtyvät tuotannonohjausjärjestelmän aktiivipuolelta passiivisiksi.

7.2.2 Analoginen aineisto hävitetään prosessoinnin jälkeen

Mikäli analoginen aineisto hävitetään prosessoinnin jälkeen, voidaan väliaikainen säilytys toteuttaa kevyemmin. Arkistoyksiköt sijoitetaan skannauksen jälkeen pinoihin siten, että ohjausarkin viivakoodi jää selkeästi näkyviin. Viivakoodin on oltava näkyvissä kahdesta syystä:

1. Mikäli digitoituun aineistoon joudutaan palaamaan prosessin myöhäisemmissä vaiheissa havaittujen virheiden takia
2. Aineistoa hävitettäessä jokainen viivakoodi pitää lukea



Arkistoyksiköt sijoitetaan siirtovälineeseen ja siirretään välivarastoon. Kun hävitettävä aineisto on sijoitettu odottamaan hävitystä, tuotannonohjausjärjestelmä tarkastaa, että jokaiselle siirtovälineessä sijaitsevalle arkistoyksikölle on suoritettu jokainen prosessin vaihe vaatimusten mukaisesti. Mikäli hävittämisehdot täyttyvät, tuotannonohjausjärjestelmä ilmoittaa siirtovälineen sisältämät arkistoyksiköt hävitysvalmiiksi, jolloin tuotannonohjausjärjestelmässä arkistoyksikön tila päivittyy (Tila: *Odottaa hävittämistä*).

Varsinainen hävittäminen tapahtuu siten, että luetaan siirtovälineen viivakoodi, mikä todentaa kyseessä olevan oikean siirtovälineen. Tämän jälkeen luetaan jokaisen arkistoyksikön viivakoodi ja aineisto siirretään odottamaan hävittämistä tietoturvasäiliöihin, joista aineisto lopulta fyysisesti hävitetään tietoturvallisesti. Tässä vaiheessa aineiston tila päivittyy tuotannonohjausjärjestelmässä (Tila: *Hävitetty*).

Aineiston käsittely on kokonaisuudessaan valmis sen jälkeen, kun digitaalinen manifestaatio saa tuotannonohjausjärjestelmässä tilan *Digitoidun aineiston siirto valmis*. Tällöin arkistoyksikkö siirtyy tuotannonohjausjärjestelmässä tilaan *Aineiston käsittely valmis*, ja määrätyt tiedot välitetään AHAA:seen. Tämän jälkeen arkistoyksiköt siirtyvät tuotannonohjausjärjestelmän aktiivipuolelta passiivisiksi.

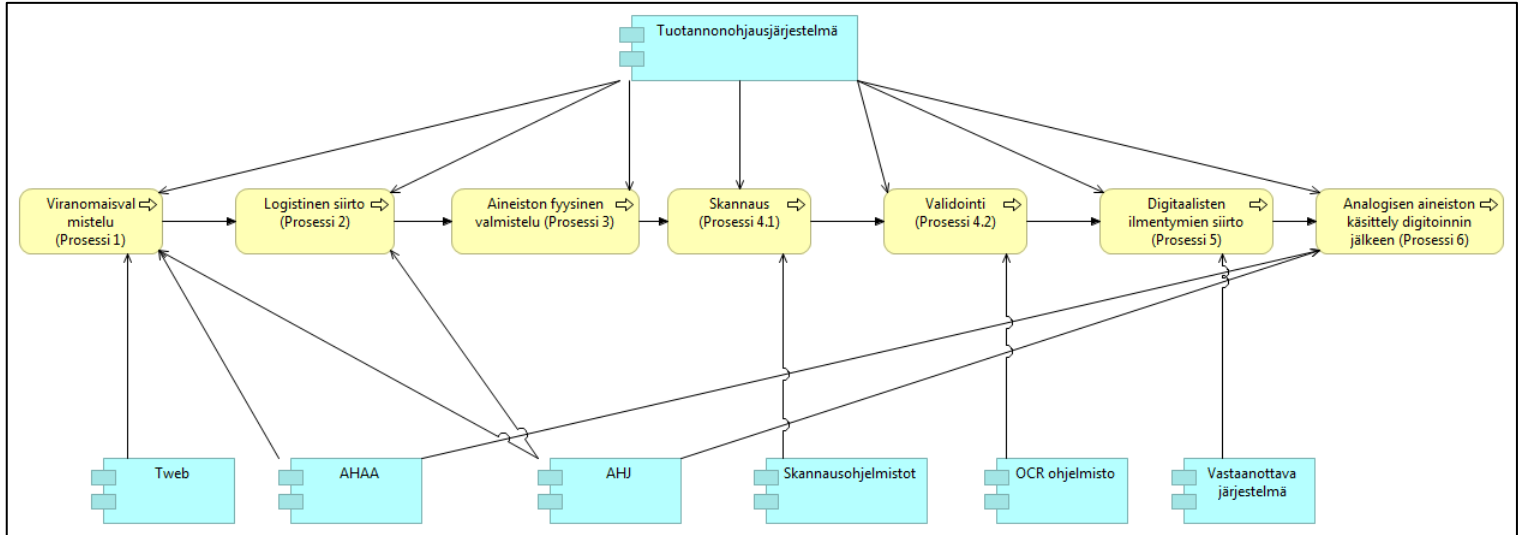
8 Tietotekninen infrastruktuuri ja tuotannonohjausjärjestelmä

8.1 Taustoitus

Tässä raportissa kuvatut prosessit vaativat toteutuakseen useampia eri järjestelmiä, ohjelmistoja ja laitteita. Luvussa 8 kuvataan, miten eri prosessit liittyvät tietojärjestelmiin sekä mitä tietoja eri järjestelmien välillä pitää liikkua, jotta kokonaisprosessi on mahdollista suorittaa. Tässä luvussa ei kuitenkaan kuvata täsmällisesti, miten prosessissa digitaaliseksi muotoon muunnettua aineistoa on mahdollista käyttää käyttöliittymällä eikä eritellä yksityiskohtaisemmin tietoteknisen infrastruktuurin teknologisia vaatimuksia (esimerkiksi konetehoa tai tallennustilan määrää). Lisäksi sisällön analysointia ei käsitellä, koska sen käsitetään yleisesti olevan varsinaisen digitointiprosessin ulkopuolella.

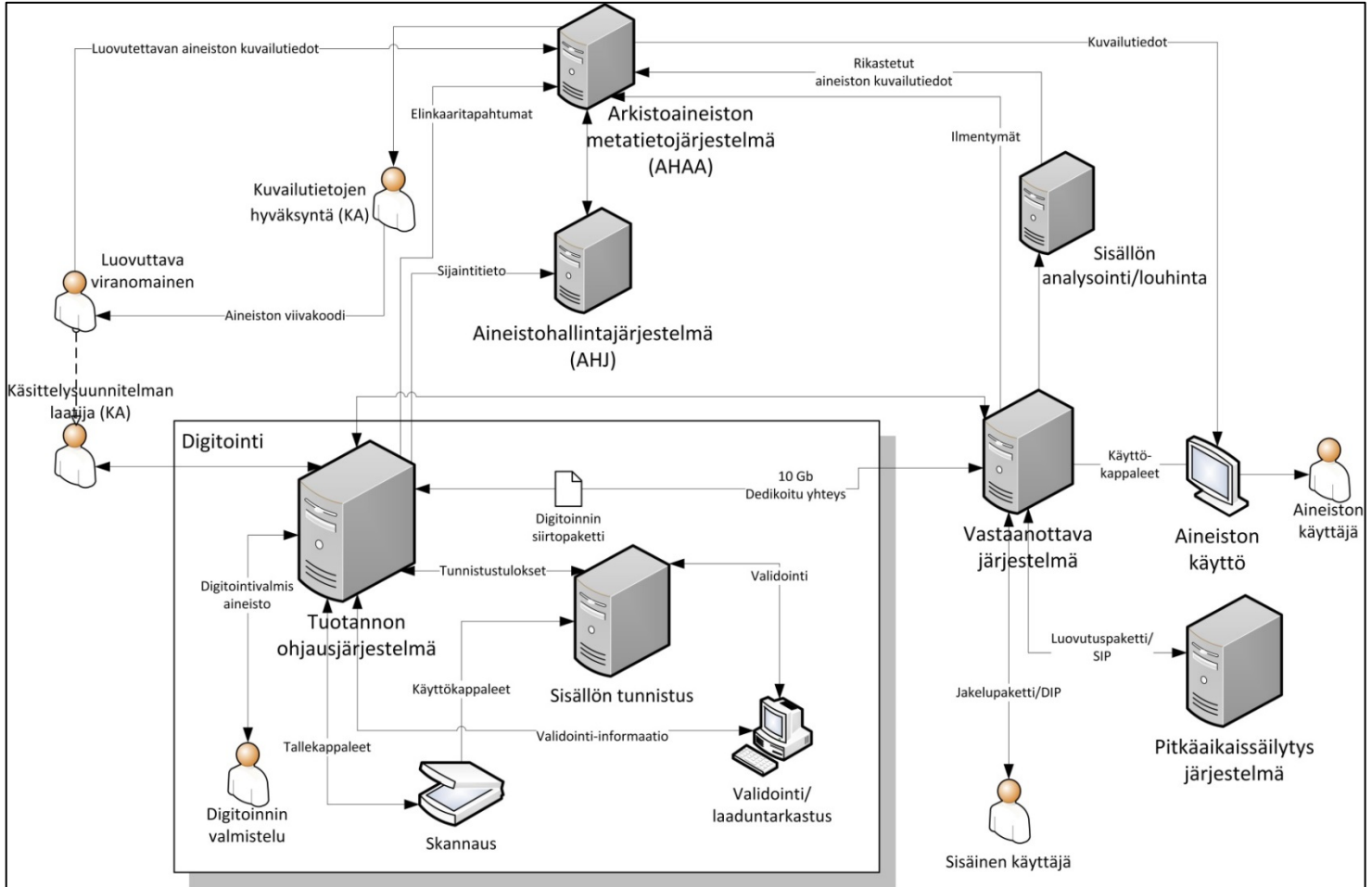
8.2 Järjestelmät, prosessit ja tietovirrat

Alla olevassa kuvassa (Kuva 8-1: Järjestelmät ja prosessit) kuvataan sitä, miten prosessit käyttävät eri tietojärjestelmiä. Kuvaan on varsinaisten tietojärjestelmien lisäksi tuotu myös skannausohjelmistot ja OCR-ohjelmistot kuvan informaatioarvon lisäämiseksi.



Kuva 8-1: Järjestelmät ja prosessit

Kuvassa 8-2 (Tietotekninen infrastruktuuri) kuvataan se järjestelmäkokonaisuus, joka pitää olla, jotta massadigitointia voidaan suorittaa. Kuvassa ilmaistaan myös eri järjestelmien välisiä tietovirtoja. Laatikon ”Digitointi” sisälle on kuvattu ne toiminnot, jotka tulisi kokonaisprosessissa sijoittaa tiloihin, missä digitoinnin tuotantoa tehdään. Suunnitteluvaiheessa on huomioitu se, että laatikon ”Digitointi” sisällä oleva tuotannonohjausjärjestelmä sisältää levykapasiteettia niin paljon, että tuotantoa on mahdollistaa jatkaa viikon ajan, vaikka jokin sen jälkeisistä kriittisistä palveluista (verkkoyhteys, vastaanottava järjestelmä, AHAA, pitkäaikaissäilytysjärjestelmä) olisi poissa toiminnasta. Tämä lisää olennaisesti tietojenkäsittelyyn suunnatun kapasiteetin tarvetta edellä mainituissa järjestelmissä, sillä dataa kasvatetaan tuotannonohjausjärjestelmän puskuriin.



Kuva 8-2: Tietotekninen infrastruktuuri

Taulukossa 8-1 kuvataan suppeasti kuvassa 8 esiintyvien järjestelmien roolit.

Taulukko 8-1: Järjestelmät ja roolit

Järjestelmä	Rooli	Järjestelmän tila	Pakollisuus massadigitoinnin näkökulmasta
Arkistoaineiston metatietojärjestelmä (AHAA)	Ydintietovaranto, missä sijaitsee arkistoaineiston kuvailu- ja luettelointitietojen ydintieto.	Otetaan Kansallisarkistossa käyttöön vuoden 2018 aikana. Vaatii suurella todennäköisyydellä muokkauksia massadigitoinnin näkökulmasta.	Kyllä
Tuotannonohjausjärjestelmä	Koko prosessin hallinnointijärjestelmä, jonka rajapintoja sekä tietoja käytettäisiin muun muassa digitaalisten ilmentymien validointiin, prosessin	Tuotannonohjausjärjestelmä on toteuttamatta. Kansallisarkistossa tuotannonohjausjärjestelmän toiminnallisuuksia toteutetaan useamman eri järjestelmän toimesta ja osaa tässä raportissa kuvattavasta	Kyllä

	validointiin, analogisen aineiston käsittelyn dokumentointiin jne.	toiminnasta ei toteuteta laisinkaan. Kyseisten järjestelmien soveltuvuutta tulee arvioida.	
Aineistohallintajärjestelmä (AHJ)	Ydintietovaranto, missä sijaitsee analogisen aineiston sijainnin ydintieto.	Järjestelmä on käytössä Kansallisarkistossa. Vaatii todennäköisesti pieniä muokkauksia massadigitoinnin näkökulmasta.	Kyllä
Sisällön tunnistus	Tietojärjestelmä, joka sisältää sovelluksia joiden avulla voidaan tunnistaa bittikarttakuvasta sisältöä (esimerkiksi tekstiä).	Sisällön tunnistukseen ei ole käytössä järjestelmää Kansallisarkistossa.	Kyllä
Vastaanottava järjestelmä	Järjestelmä, joka vastaanottaa digitoinnissa syntyneen siirtopaketin, rikastaa ja jakelee sitä sekä käyttöön että säilyttämiseen.	Tässä raportissa kuvattavan siirtopaketin vastaanottavaa järjestelmää ei ole käytössä toistaiseksi. Järjestelmä on suunniteltu olevan SAPA-palvelu.	Kyllä
Pitkäaikaissäilytysjärjestelmä	Järjestelmä, joka säilyttää digitoidusta aineistosta tallekappaleet	Järjestelmä on käytössä Kansallisen digitaalisen kirjaston pitkäaikaissäilytysjärjestelmänä (KDK-PAS). Vaatii mahdollisesti kehitystyötä massadigitoinnin näkökulmasta.	Kyllä ¹²
Sisällön analysointi	Järjestelmä, joka käyttää hyväkseen sisällön tunnistuksen tuloksia ja rikastaa "aineiston" metatietoja.	Sisällön analysointiin ei ole käytössä järjestelmää Kansallisarkistossa. Järjestelmä voidaan toteuttaa irrallaan massadigitoinnista.	Ei

8.3 Tuotannonohjausjärjestelmä

Kansallisarkiston digitointiprosessia nykyisellään hallinnoidaan useammalla erilaisella käyttöliittymällä. Tuotannonohjausjärjestelmän kehittämistyössä tulisi keskittyä näiden toimintojen tuomiseen yhteen järjestelmään sekä lisätoiminnallisuuksien toteuttamiseen. Tuotannonohjausjärjestelmän ensisijainen tarkoitus on mahdollistaa tuotannon suorittaminen mahdollisimman tehokkaasti, mutta sen rooliin kuuluu myös prosessin seuraaminen, elinkaaritapahtumien dokumentointi ja prosessin raportointi.

Tuotannonohjausjärjestelmän roolin laajuuteen vaikuttaa se mikä rooli digitointiin ja sisällön tunnistamiseen valikoituneilla ohjelmistoilla on ja mitä toimia ne hoitavat. Tuotannonohjausjärjestelmän käyttö alkaa jo prosessissa 1. ja sillä seurataan sekä analogisen aineiston että digitaalisen aineiston liikkeitä siihen saakka, kunnes analoginen aineisto joko hävitetään tai sijoitetaan uudelleen ja digitoinnissa syntyneet siirtopaketit on vastaanotettu eheänä vastaanottavassa järjestelmässä. Koneoppimista olisi mahdollista hyödyntää tuotannonohjausjärjestelmälle ihmisen toimesta tuotettavan

¹² Käytännössä massadigitoinnin näkökulmasta tarvitaan vain digitoidun aineiston vastaanottava järjestelmä

datan pohjalta, toisin sanoen opetusaineiston karttumisen jälkeen. Jotta koneoppiminen olisi mahdollista, tarkoittaisi se hyvin tarkkaa ajallista seuraamista aineiston eri työvaiheista. Edellä mainitun tiedon perusteella olisi mahdollista rakentaa tuotannonohjausjärjestelmään algoritmi, joka osaisi ennakoita kyseiselle analogiselle aineistokokonaisuudelle tarvittavaa henkilöresurssin määrää, tarvittavaa laiteinfrastruktuuria ja mahdollisia läpimenoaikoja (resurssien allokointi). On kuitenkin huomioitava, että kyseinen algoritmi tarvitsisi pohjadatakseen useampia kokoelmia ja tarkkaan mietittyjä käsittelysuunnitelman parametreja viranomaisvalmisteluvaiheesta.

Alla olevassa taulukossa on kuvattu arkistoyksikköön kohdistuvat tilat tuotannonohjausjärjestelmässä. Taulukon oletuksena on se, että prosessi suoritetaan virheettää. Virhetiloista on maininta prosessin kuvauksessa. Mikäli aineisto säilytetään analogisena sen digitoinnin jälkeen, siirtyy se tilan 16 jälkeen suoraan odottamaan tilaa 19. Mikäli taas aineistoa ei säilytetä analogisena, jää kohta 19 pois vaadituista statuksista.

Taulukko 8-2: Suunnitelma aineiston statuksista Tuotannonohjausjärjestelmässä

#	Tila tuotannonohjausjärjestelmässä	Toteutuu prosessissa
1	Viranomaisvalmistelu valmis	Viranomaisvalmistelu
2	Siirtovalmis aineisto	Viranomaisvalmistelu
3	Kuljetusvalmis aineisto	Logistinen siirto
4	Kuljetuksessa	Logistinen siirto
5	Aineisto vastaanotettu	Logistinen siirto
6	Siirretty valmisteluun	Aineiston valmistelu digitoitavaksi
7	Digitoinnin valmistelu	Aineiston valmistelu digitoitavaksi
8	Digitointivalmis aineisto	Aineiston valmistelu digitoitavaksi
9	Skannauksessa	Skannaus
10	Arkistoyksikkö skannattu	Skannaus
11	Arkistoyksikön kuvanlaatu validoitu otannalla	Validointi
12	Sisältö tunnistettu	Validointi
13	Prosessivaiheet sekä kuvat ja tekstit tarkastettu	Validointi
14	Siirtopaketti paketoitu	Validointi
15	Siirtopaketti lähetetty	Digitaalisten ilmentymien siirto
16	Digitoidun aineiston siirto valmis	Digitaalisten ilmentymien siirto
17	Odottaa hävittämistä ¹³	Analogisen aineiston käsittely digitoinnin jälkeen
18	Hävitetty ¹³	Analogisen aineiston käsittely digitoinnin jälkeen

¹³ Ei toteudu aineiston suhteen, jonka analogista manifestaatiota ei hävitetä.

19	Säilytettävän analogisen aineiston käsittely valmis ¹⁴	Analogisen aineiston käsittely digitoinnin jälkeen
20	Aineiston käsittely valmis	Analogisen aineiston käsittely digitoinnin jälkeen

Prosessissa syntyneiden digitaalisten ilmentymien autenttisuus ja eheys on erilaisten käsittelyprosessissa suoritettavien vaiheiden ja näiden dokumentoinnin summa. Tämän takia on olennaista kirjata sekä analogisen että digitaalisen aineiston käsittelyyn liittyviä vaiheita ja tallentaa näitä ns. elinkaaritapahtumiksi osaksi aineiston kuvailutietoja AHAA-järjestelmässä. Tuotannonohjausjärjestelmässä pitää voida määrittää tiettyjen käsittelyvaiheiden tallentaminen automaattisesti elinkaaritapahtumiksi AHAA:ssa tai toisaalta prosessin sijaintitiedoiksi AHJ:ssa. Tuotannonohjausjärjestelmässä seurataan aineiston liikehdintää huomattavasti AHJ:ta tarkemmalla tasolla, mutta tätä tietoa ei ole tarkoituksenmukaista päivittää kaikkien vaiheiden osalta AHJ:n.

9 Yhteenveto tiedostoformaateista

9.1 Yhteenveto kuvaformaateista

Dataa voidaan yleisesti tallentaa moneen eri formaattiin. Arkistosektorilla on ”de facto -standardeja”, jotka perustuvat yleisesti siihen, että ne ovat laajalti käytettyjä ja hyvin tuettuja. Tämä tarkoittaa esimerkiksi sitä, että pakkaamaton TIFF (6.0) -tiedosto on selkeästi käytetyin tallekappale (archival master), jota käytetään, kun tähtäimenä on pitkäaikaissäilytys. TIFF-tiedoston puolesta puhuu sen laaja tuki ja monessa suhteessa ylivoimaisuus kuvanlaadun suhteen. Tämä on nähty äärimmäisen tärkeänä asiana, kun on tallennettu pysyvään säilytykseen määrättyä asiakirjallista kulttuuriperintöä. Osa pysyvään säilytykseen määrätystä analogisista asiakirjoista voidaan nähdä myös ”teoksina” (kartat, piirustukset jne.). Tässä raportissa käsitellään pääsääntöisesti analogista aineistoa, joka on tuotettu joko kirjoituskoneella ja/tai tietokoneella, joten viitekehys voidaan nähdä erilaisena verrattuna edellä mainittuihin. JPEG-formaatti on aina häviöllinen, mutta erittäin paljon käytetty ja erittäin hyvin tuettu. Se on yleisin käyttökappaleen formaatti pienestä koostaan ja hyvästä selain tuestaan johtuen. Lisäksi JPEG mahdollistaa melkein kaikkien tarvittavien metatietojen kirjoittamisen tiedoston ”tageihin”. Tässä suhteessa JPEG kuitenkin eroaa TIFF (Tagged image file format) -formaattista, sillä yleisesti voidaan todeta, että TIFF-tiedostoissa metatiedot kirjoitetaan ”Tif-tageihin”, kun taas JPEG:ssä käytetään eri skeemaa – esimerkiksi EXIF.

Kansainvälisen vertailun perusteella yleisimmin käytetyt tiedostoformaatit kuville ovat TIFF, JPEG ja PNG.¹⁵ Niistä yleisimmät tallekappaleet ovat TIFF (88 % kansainväliseen selvitykseen vastanneista ilmoitti TIFF:in tallekappaleekseen) ja PNG, jota käytetään lähinnä mikrofilmien ja -korttien digitoinnissa. JPEG on yleisin käyttökappaleen formaatti. Eri formaatit vaikuttavat merkittävästi kuvatiedostojen

¹⁴ Ei toteudu aineiston suhteen, jonka analoginen manifestaatio hävitetään.

¹⁵ *Digitization Survey* (National Archives of Finland, 2017).

datamääriin. Library of Congress on tuottanut eri kuvaformaateista tiiviin yhteenvedon, joka on vapaasti saatavilla verkossa.¹⁶ Kansallisen digitaalisen kirjaston (KDK) standardisalkussa TIFF, PNG ja JPEG on määritelty kaikki kuvatiedostojen säilytyskelpoisiksi tiedostoformaateiksi. Säilytyskelpoisiksi hyväksytään sellaiset tiedostoformaatit, joissa tietosisällön säilyminen ja ymmärrettävyys voidaan taata pidemmällä aikavälillä.¹⁷

Massadigitoinnin suunnitteluprojektissa on otettu lähtökohdaksi tuottaa digitoinnin tuloksena digitaalisen ilmentymän tallekappale TIFF-formaatissa ja käyttökappale JPEG-formaatissa (ks. tarkemmin luku 5.6).

9.2 Datamäärät

Digitoinnissa syntyvät datamäärät ovat riippuvaisia monesta tekijästä. Suurimmat yksittäiset tekijät ovat analogisten asiakirjojen fyysinen koko ja valittu tiedostoformaatti. Koska massadigitoinnissa ei tulla nostamaan resoluutiota analogisten ilmentymien fyysisten ominaisuuksien mukaan (kuten yleisesti arkistosektorin aineiston skannauksessa tehdään), tarkoittaa tämä sitä, että esimerkiksi A5-kokoinen asiakirja on datana puolet pienempi kuin A4-kokoinen asiakirja.

Kuvan tallennusformaatin, resoluution ja pakkauksen määrän valinnalla on myös merkittävä vaikutus prosessissa syntyvien digitaalisten ilmentymien vaatimiin datamääriin.

9.3 Kuvaformaatit

TIFF (Tagged Image File Format) on tagi-pohjainen tiedostoformaatti, joka on tarkoitettu bittikarttakuvien tallentamiseen ja varastointiin.¹⁸ Tiedostoformaatti on pitkäikäinen ja sen tuorein kuudes versio on ollut käytössä vuodesta 1992. TIFF-tiedostot voidaan tallentaa pakkaamattomina tai pakattuina. Yleisin pakkausformaatti on LZW (Lempel-Ziv-Welch), joka mahdollistaa häviöttömän pakkauksen.¹⁹ Tiedostoformaatin etuja ovat sen hyvä soveltuvuus pitkäissäilytykseen ja migraatioihin sekä yleisesti saatavilla olevat sovellusratkaisut, jotka laskevat kustannuksia. TIFF-formaatin heikkous on pakkaamattomien tiedostojen suuri koko ja verkkoselaintuen puute. Tästä johtuen TIFF-tiedostoista muodostetaan yleensä JPEG-tiedostoja kuvien esittämistä ja jakelua varten.²⁰ Kuvatiedostojen suuresta koosta huolimatta TIFF-tiedostoja ei ole kansainvälisen selvityksen perusteella yleisesti pakattu, kun

¹⁶ 'Format Descriptions for Still Image' <https://www.loc.gov/preservation/digital/formats/fdd/still_fdd.shtml> [accessed 10 November 2017].

¹⁷ Kansallinen digitaalinen kirjasto (KDK): Säilytys- ja siirtokelpoiset tiedostomuodot, v1.5.1.

¹⁸ TIFF-formaatista tarkemmin esim. 'TIFF, Revision 6.0', 2009
<<https://www.loc.gov/preservation/digital/formats/fdd/fdd000022.shtml>> [accessed 10 November 2017].

¹⁹ LZW-pakkauksesta esim. 'LZW Compression Encoding', 2012

<<https://www.loc.gov/preservation/digital/formats/fdd/fdd000135.shtml>> [accessed 10 November 2017].

²⁰ *Raster Still Images for Digitization. A Comparison of File Formats*, 3 vols (Federal Agencies Digitization Guidelines Initiative. Still Image Working Group, 2014), part 3, 9; 'TIFF, Revision 6.0'.

niitä on käytetty tallekappaleina. Pakkaamaton TIFF on selkeästi yleisin tallekappaleen tiedostoformaatti tallennettujen kuvatiedostojen määrästä huolimatta.²¹

PNG²² (Portable Network Graphics) on bittikarttakuvien tallennusformaatti, joka kehitettiin korvaamaan patentoidun GIF -formaatin verkkoselaimissa. Vuoden 2005 jälkeen PNG -formaattia ei olla tuettu yleisimmissä selaimissa.²³ Formaatti tukee 48 bittistä RGB -väritilaa, mutta nykyään PNG:tä käytetään lähinnä mikrofilmien ja -korttien digitoinnissa.²⁴ PNG -formaatin käyttö näyttäisi perustuvan skannauslaitteiston mahdollistamiin tiedostoformaatteihin.

JPEG²⁵ (Joint Photographic Experts Group) on bittikarttakuvan tiedostoformaatti, joka tallentaa tiedon aina pakattuna. Formaatti on yleisin levitys- ja käyttötiedosto arkistosektorilla. JPEG on yleisesti käytössä kameroissa, matkapuhelimissa sekä useissa muissa laitteissa, joilla on mahdollista tuottaa kuvatiedostoja ja sillä on erinomainen ohjelmisto- sekä selaintuki. Lisäksi JPEG-formaatin etuja ovat sen pieni tiedostokoko. TIFF-formaatti on ominaisuuksiltaan parempi pitkäaikaissäilytykseen kuin JPEG.²⁶ Toisaalta pitkälti kustannussyistä esimerkiksi Alankomaiden Kansallisarkisto sekä Yhdysvaltojen Kansallisarkisto ovat päätyneet hyväksymään myös JPEG – pakatun tiedoston tallekappaleeksi.²⁷ Koska JPEG-formaatti on häviöllinen se voi olla ongelmallinen tallekappaleena migraatioista johtuen. Se ei ole myöskään ihanteellinen formaatti OCR-tekniikoiden kannalta, koska sen pakkaus voi pyöristää merkkien ja sanojen reunoja, jolloin OCR:n tarkkuus laskee.

JPEG 2000²⁸ aallokemuunnokseen perustava pakkausstandardi kuville ja videolle. JPEG 2000 ei ole täysin vapaa patenteista, mutta keskeiset patentin haltijat ovat luopuneet oikeuksistaan standardin keskeisiin osiin. Formaatin käyttäjien tulee itse selvittää mahdolliset patentin haltijat. JPEG 2000 mahdollistaa häviöttömän, että häviöllisen pakkauksen. Sen keskeisin etu on, että häviötön pakkaus mahdollistaa merkittävästi pienemmän kuvatiedoston koon verrattuna esimerkiksi TIFF-formaattiin. JPEG 2000 huonoja puolia on, että skannerit eivät yleisesti tuota JPEG 2000 – kuvia. Niiden tuottamiseen tarvitaan erilliset sovellukset, joita on usein kehitettävä vastamaan digitoivan organisaation tarpeita. JPEG 2000

²¹ *Digitization Survey* Yhdysvaltain kansallisarkisto (National Archives and Records Administration tallentaa kaikki 390 miljoonaa kuvatiedostoaan TIFF-formaatissa.

²² PNG. <https://www.w3.org/TR/2003/REC-PNG-20031110/> Viitattu 28.12.2017

²³ Tarkemmin PNG-formaatista esim. 'PNG, Portable Network Graphics', 2009

<<https://www.loc.gov/preservation/digital/formats/fdd/fdd000153.shtml>> [accessed 10 November 2017].

²⁴ Esimerkiksi Viron kansallisarkisto digitoi mikrofilmensä PNG-formaatissa. Formaatti on ollut käytössä Australian ja Kanadan kansallisarkistojen master-tiedostona kuville. Australia on kuitenkin luopunut siitä ja siirtynyt TIFF-tiedostoihin. *Digitization Survey* (National Archives of Finland, 2017); *Raster Still Images for Digitization. A Comparison of File Formats*, part 3, 9.

²⁵ JPEG. <https://jpeg.org/jpeg/index.html> Viitattu 29.12.2017

²⁶ *Raster Still Images for Digitization. A Comparison of File Formats*, part 3, 9.

http://www.digitizationguidelines.gov/guidelines/FADGI_RasterFormatCompare_p3_20140417_r.pdf Viitattu 5.1.2018

²⁷ Joop Korswagenin (Alankomaiden Kansallisarkisto) esitys Alankomaan Kansallisarkiston digitointikriteereistä

Kansallisarkiston vierailulla 4.1.2017 Osa Alankomaiden Kansallisarkiston tallekappaleista on: Jpeg, 24bit, 300ppi, EciRGBv2, JPEG-laatu: 90%.

²⁸ <https://jpeg.org/jpeg2000/index.html>



on rakenteeltaan suhteellisen monimutkainen eikä sitä ole otettu yleisesti käyttöön siitä huolimatta, että sen pakkausalgoritmit ovat tunnustettu toimiviksi. Yleisimmät verkkoselaimet eivät tue formaattia.

9.4 Optinen tekstintunnistus (OCR)

Digitoinnin yhteydessä tuotetuista kuvatiedoista on mahdollista tallentaa niiden sisältämää tekstiä optisella tekstintunnistuksella erillisiin tiedostoihin. Yleisimmät OCR-formaatit ovat HMTL, PDF/A ja XML. Kansallisarkisto toteutti 2016 selvityksen OCR-formaateista, jonka perusteella KDK:n määrittelemien pitkäaikaissäilytyksen tiedostoformaattien ehdot täyttivät muun muassa ABBYY FineReader XML, Alto XML, hOCR, Page XML ja PDF/A. Näistä Alto XML oli selvityksen suosittelema formaatti, sillä se on halutessa mahdollista linkittää METS-skeeman avulla digitoituun kuvatiedostoon.²⁹

9.5 Metadata

Digitointiprosessi voi tuottaa metadataa useammassa eri muodossa. Osa metadatasta on mahdollista upottaa kuvatiedoston ”tageihin”. Käytetty kuvaformaatti asettaa rajat ”tageihin” tallennettavalle metadataalle.

Pitkäaikaissäilytys edellyttää, että tallennettavat tiedostot kuvaillaan Metadata Encoding & Transmission Standard (METS) – standardin mukaisesti. Sen sisällä jokaiselle digitaaliselle objektille on muodostettava PREMIS-muotoinen metatieto-objekti. Digitoitujen kuvien tekninen metadata pitää esittää MIX-skeemassa PREMIS:n sisällä. Tämän metadata tiedoston on oltava XML-muotoinen, joten digitointiprosessissa pitää tuottaa XML-skeema, joka sisältää MIX-skeeman mukaiset tiedot ja siitä on mahdollista muodostaa PREMIS-paketti pitkäaikaissäilytystä varten.

²⁹ Sami Nousiainen, *Report on File Formats for Hand-Written Text Recognition (HTR) Material* (National Archives of Finland, 2016), 52–55 Raportti sisältää tarkemmat kuvailut yleisimmistä formaateista liittyen bittikarttakuvasta tunnistetun tekstin tallennukseen.