



KANSALLISARKISTO

Massadigitoinnin suunnitteluprojekti

Vipuvoimaa  
EU:lta  
2014–2020



Euroopan unioni  
Euroopan aluekehitysrahasto

# Massadigitoinnin suunnitteluprojektin loppuraportti

---



## Sisällysluettelo

Yhteenvedo projektin tuloksista .....	3
1 Johdanto.....	5
1.1 Projektin tausta ja tavoitteet .....	5
1.2 Projektin rajaukset .....	6
1.3 Keskeiset käsitteet.....	7
2 Massadigitoinnin tilanne kansainvälisesti .....	8
3 Digitoitavat aineistot.....	10
3.1 Valtionhallinnon analogisten asiakirjojen määrä.....	10
3.2 Digitoitavat aineistot .....	12
3.3 Digitoinnin aikataulu ja järjestys .....	14
4 Massadigitoinnin prosessi.....	15
4.1 Prosessin ennakkoehdot .....	17
4.2 Viranomaisvalmistelu .....	17
4.3 Logistinen siirto .....	18
4.4 Digitoinnin valmistelu.....	19
4.5 Digitointi .....	20
4.6 Digitaalisten ilmentymien siirto .....	22
4.7 Analogisen aineiston käsittely digitoinnin jälkeen.....	22
5 Massadigitoinnin tuotanto.....	23
6 Kustannukset ja hyödyt.....	27
6.1 Digitoinnin kustannukset.....	27
6.2 Digitoinnin hyödyt.....	28
6.3 Kustannus- ja hyötyanalyysi .....	29
7 Massadigitoinnin riskit .....	32
8 Massadigitoinnin riippuvuudet ja jatkotoimenpiteet.....	33
8.1 Henkilöstö ja osaaminen .....	34
8.2 Tietojärjestelmät .....	36
8.3 Ehdotetut jatkotoimet.....	37



## Yhteenveto projektin tuloksista

Massadigitoinnin tavoitteena on saattaa hallinnon paperiaineistot digitaaliseen muotoon ja siten paremmin saataville ja käytettäväksi. Aineistojen digitaalisella säilyttämisellä ja paremmalla käytettävyydellä tavoitellaan asiakirjahallinnon ja arkistotoimen tehtävien merkittävää tehostumista valtionhallinnossa ja kustannussäästöjä.

Massadigitoinnin suunnitteluprojektissa on määritelty uusi hävittämiseen tähtäävä digitointiprosessi, jossa digitoinnin jälkeen asiakirjatiedon analoginen ilmentymä voidaan hävittää. Asiakirjan tietosisältö ei häviä, vaan ainoastaan asiakirjan säilytysmuoto muuttuu. Keskeisenä vaatimuksena määrittelyssä on ollut Kansallisarkiston lakisääteinen tehtävä varmistaa kansalliseen kulttuuriperintöön kuuluvien viranomaisten asiakirjojen ja niihin sisältyvien tietojen säilyminen.

Suunnitteluprojektin käynnistyessä suoritettiin laaja aineistokartoitus. Sen tuloksena Kansallisarkisto sai aiempaa huomattavasti tarkemman kuvan valtionhallinnon arkistoitavien aineistojen määrästä ja niiden digitointiin vaikuttavista ominaisuuksista. Kartoituksessa hallinnon arkistojen aineistomäärä nousi miltei 400 hyllykilometriin, joka on huomattavasti aiempia arvioita korkeampi. Määrä on miltei kaksi kertaa niin paljon kuin Kansallisarkiston hallussa oleva nykyinen aineistomäärä. Digitoitavaksi aineistomääräksi on projektissa arvioitu 212 hyllykilometriä. Valtaosa digitoinnin ulkopuolelle jäävistä aineistoista on määrääjän säilytettäviä aineistoja. Digitoitavan aineiston määrä tarkentuu massadigitoinnin jatkovalmistelussa.

Digitoitava aineisto on hajaantuneena virastoissa ja niiden eri arkistotiloissa. Prosessin suunnittelussa on pitänyt ottaa huomioon, miten aineisto saadaan viranomaisten arkistotiloista siirrettyä hallitusti keskitettyyn digitointiin. Digitointiprosessissa viranomaiselta edellytetään Kansallisarkiston nykyistä siirtomenettelyä huomattavasti vähemmän omia toimia aineiston



valmistelussa ja siirrossa. Uuden toimintamallin lähtökohtana on hoitaa aineiston valmistelun tehtävät keskitetysti, mikä lisää toiminnan tehokkuutta.

Digitoinnin kustannukset koostuvat valtaosin aineiston valmistelun ja skannauksen henkilötyöstä, jolloin aineiston ominaisuuksilla on merkittävä vaikutus työmäärään ja siten kustannuksiin. Keskeisenä haasteena digitoinnissa on varmistaa asiakirjatiedon eheyden ja autenttisuuden säilyminen eli toteuttaa digitointi laadukkaasti siten, että lopputuloksena on analogisessa muodossa olevasta asiakirjasta tuotettu digitaalinen ilmentymä, jossa tietosisältöä ei ole hävinnyt. Tämä pyritään varmistamaan tuottamalla digitoinnissa riittävän laadukas kuva, jonka laadun määrittelyssä on nojaututtu arkistosektorilla kansainvälisesti käytettyihin laatuvaatimuksiin.

Massadigitoinnin hyötyjen tarkastelussa lähtökohtana on verrata nykyistä analogista säilyttämistä digitaalisen säilyttämisen skenaarioon. Nykytilassa viranomaiset siirtävät vähitellen Kansallisarkistoon 40 vuotta tai sitä vanhempia analogisia asiakirjoja, jolloin virastoissa säilyy vielä pitkän aikaa paperiarkistot ja niihin liittyvät kustannukset. Digitaalisessa skenaariossa viranomaisten siirtymistä täysin digitaaliseen tiedonhallintaan nopeutetaan merkittävästi massadigitoinnilla. Näin viranomaiset vapautuvat paperiarkistojen hoidosta huomattavasti nopeutetussa aikataulussa. Massadigitoinnin päätyttyä vuonna 2030 digitaalisen säilyttämisen vuosittainen säästö verrattuna analogisen säilyttämisen skenaarioon on arviolta 14 miljoonaa euroa.

Massadigitoinnin toteuttaminen on yli kymmenen vuoden ohjelma, joka edellyttää hyvää koordinoitua, mutta ennen muuta vahvaa sitoutumista. Viranomaisten tulee sitoutua yhdessä määriteltäviin digitoinnin aikatauluihin ja aineiston valmistelun ohjeistukseen, Kansallisarkiston pitää sitoutua digitointiprosessin ohjausvastuuseen ja toimintaa ohjaavien ministeriöiden tulee sitoutua yhteisesti määriteltyihin tavoitteisiin sekä toiminnan jatkuvuuteen.



## 1 Johdanto

### 1.1 Projektin tausta ja tavoitteet

Massadigitoinnille on määritelty raamit valtioneuvoston 21.6.2017 periaatepäätöksessä, jonka mukaan viranomaisten hallussa oleva arkistoitava analoginen asiakirja-aineisto digitoidaan vuoteen 2030 mennessä. Päätöksen mukaan asiakirjatieto arkistoidaan vain digitaalisessa muodossa, jolloin suurin osa analogisessa muodossa olevista asiakirjoista hävitetään digitoinnin jälkeen.

Massadigitoinnilla tarkoitetaan tässä yhteydessä aineistojen laajamittaista digitointia, joka on organisoitu ja toteutettu teollisena prosessina. Massadigitointi eroaa muusta digitoinnista erityisesti siinä, että digitoinnin prosessia ei räätälöidä aineistotyyppi- tai organisaatiokohtaisesti, mikä on tyypillistä projektimuotoiselle digitoinnille.

Periaatepäätöksen asettamat reunaehdot ovat määritelleet massadigitoinnin suunnittelua. Projektin tavoitteena oli luoda toteutussuunnitelma digitointiprosessista, joka alkaa aineiston siirtämisestä viranomaiselta digitoitavaksi ja päättyy siihen, että aineisto on muutettu digitaaliseen muotoon ja analoginen ilmentymä asiakirjasta on hävitetty.

Projektin tehtäviksi määriteltiin seuraavat kokonaisuudet:

- tuottaa tiekartta julkisen hallinnon paperiasiakirjojen digitoinnin suunnitelmaksi, joka esittää missä järjestyksessä ja millä aikataululla aineistojen digitoinnissa edetään
- kartoittaa tiekartan pohjaksi valtionhallinnon organisaatioiden hallussa olevien sekä valtionhallinnosta ja kunnilta maakunnille siirtyvien aineistojen määrät ja ominaisuudet
- suunnitella massadigitoinnin toteuttamisen prosessi ja määritellä massadigitoinnissa käytettävät teknologiat
- laatia massadigitointia koskevat kustannus- ja hyötylaskelmat



- määrittellä logistinen malli miten aineistot siirretään valtionhallinnon virastojen arkistoista keskitettyyn digitointiin
- selvittää massadigitoinnin kansainvälistä tilannetta ja digitoinnin parhaita käytäntöjä, joita on käytössä.

Projekti toteutettiin Euroopan aluekehitysrahaston osarahoituksella ajalla 2.6.2017–31.12.2017. Projektin päätoteuttajana oli Kansallisarkisto ja osatoteuttajana Kaakkois-Suomen ammattikorkeakoulu, Xamk, joka vastasi logistiikan osakokonaisuuden selvitystyöstä. Projektin rahoitukseen sekä ohjausryhmän työskentelyyn osallistui myös Mikkelin kaupunki.

Tämän raportin tavoitteena on antaa yleiskuva massadigitoinnin toteuttamisen kokonaisuudesta. Raportin liitteissä on kuvattu yksityiskohtaisemmin eri selvitysten ja suunnitelmien toteuttamista. Liitteiden tavoitteena on antaa yksityiskohtaisempi kuvaus massadigitoinnin määrättyistä osa-alueista.

## 1.2 Projektin rajaukset

Projekti määriteltiin luonteeltaan tekniseksi selvitysprojektiksi, jossa ei oteta kantaa massadigitointiin liittyviin hallinnollisiin tai lainsäädännöllisiin kysymyksiin. Projekti toteutettiin seuraavin rajauksin:

- Projektissa ei oteta kantaa tulevan massadigitoinnin erilaisiin hallinnollisiin toteuttamistapoihin (ulkoistaminen, virkatyö tms.).
- Projektissa ei oteta kantaa aineiston mahdollisen omistajuuden siirtymiseen digitaaliseen muotoon muuttamisen seurauksena.
- Projektissa ei oteta kantaa digitaalisen aineiston tulevaan säilytysratkaisuun ja sen kehittämiseen.
- Projektissa ei tuoteta kuvauksia liittyen kuntien aineistoihin, jotka eivät ole siirtymässä maakuntien omistukseen.
- Projektissa ei oteta kantaa tulevan massadigitoinnin rahoittamistapaan ja vaihtoehtoihin.



Tiukan toteuttamisaikataulun vuoksi projekti rajattiin koskemaan ainoastaan digitaalisen muotoon muuttamisen prosessia, jolloin digitaalisten aineistojen säilyttämiseen ja käyttöön liittyvä kokonaisuus rajattiin ulkopuolelle. Rinnakkain projektin kanssa on ollut käynnissä sähköisen arkistoinnin palvelukokonaisuuden (SAPA) arkkitehtuuryö, jonka tavoitteena on ollut määrittää palvelun ratkaisuarkkitehtuuri. Työtä on vetänyt opetus- ja kulttuuriministeriö ja siihen ovat osallistuneet valtiovarainministeriö, Kansallisarkisto ja tieteen tietotekniikkakeskus CSC Oy.

Massadigitoinnin suunnitteluprojektin tavoitteisiin kuului selvittää, mitä kunnilta maakunnille siirtyviin tehtäviin liittyviä aineistoja tullaan digitoimaan. Projektin aikana maakuntaudistuksen kokonaisaikataulua lykättiin vuodella. Tämän seurauksena maakuntaudistuksesta vastaavilta tahoilta ei ole saatu linjauksia maakuntien aineistojen digitoinnista. Maakunnille siirtyvien aineistojen digitointi jää siten projektin jälkeen linjattavaksi.

### 1.3 Keskeiset käsitteet

- **Aineisto:** asiakirja tai asiakirjojen kokonaisuus. Ei ole yksiselitteinen arkistokuvailullinen kokonaisuus.
- **Digitointi:** prosessi, jossa analogisessa muodossa oleva informaatio muutetaan digitaaliseen muotoon
- **Massadigitointi:** aineistojen laajamittainen digitointi, joka on organisoitu ja toteutettu teollisena prosessina.
- **Skannaus:** toimenpide, jolla analogisessa muodossa oleva informaatio muunnetaan digitaaliseen muotoon käyttäen välineinä tähän tarkoitettua teknologiaa ja ohjelmistoratkaisuja.
- **Säilytyspäättös:** Kansallisarkisto päättää säilytyspäätöksessä, mitkä julkishallinnon asiakirjatiedot on säilytettävä pysyvästi, mikä on niiden säilytysmuoto, ja mitkä analogiset aineistot hävitetään tai säilytetään digitoinnin jälkeen. Tässä yhteydessä tarkoitetaan erityisesti päätöstä, jossa määrätään, säilytetäänkö asiakirjatieto myös analogisessa muodossa digitoinnin jälkeen.



- **Tiekartta:** aineistojen digitoinnin suunnitelma, joka ohjaa viranomaisten hallussa olevien analogisten aineistojen digitoinnin järjestystä ja aikataulua sekä kuvaa hallinnonalakohtaisesti digitoitavat aineistotyytit ja aineistomäärät.

## 2 Massadigitoinnin tilanne kansainvälisesti

Kansallisarkisto toteutti loka–joulukuussa 2017 selvityksen, jonka tarkoituksena oli kartoittaa massadigitoinnin tilaa kansallisarkistosektorilla. Sen keskeinen tulos on, että tässä raportissa tarkoitettua massadigitointia ei ole toteutettu arkistokontekstissa missään muualla maailmassa. Asiakirjatiedon analogista säilytysmuotoa ei ole hävitetty digitoinnin jälkeen, joten kansainvälisiä vertailukohtia hyvistä käytänteistä ei ole saatavilla.

Selvitys toteutettiin verkkokyselynä sekä kahdenvälisinä keskusteluinä Ruotsin, Norjan, Alankomaiden ja USA:n kansallisarkistojen kanssa. Lisäksi kerättiin eri maiden digitoinnin laatumäärittäjiä, digitointiohjeita sekä digitoinnin yhteyshenkilöiden kontaktitietoja. Kyselyyn vastasi yhteensä neljäkymmentä arkistoa, kirjastoa ja yritystä. Mukana olivat muun muassa kaikki Pohjoismaat, Yhdysvallat, Australia, Alankomaat, Belgia ja Itävalta.

Selvityksen perusteella Kansallisarkiston nykyinen digitointi on vuosituotannon perusteella määrällisesti kansainvälisellä huipulla Yhdysvaltojen jälkeen ainakin vuoden 2017 lukujen valossa. Yhdysvallat digitoi noin kolmekymmentä miljoonaa tiedostoa vuosittain, ja Kansallisarkisto digitoi kahdeksantoista miljoonaa tiedostoa vuonna 2017. Muita merkittäviä määriä aineistoa digitoivia – yli kymmenen miljoonaa kuvaa vuodessa – tahoja ovat Ruotsin, Tanskan ja Australian kansallisarkistot. Digitoinnin määrät ovat suuntaa antavia ja voivat vaihdella huomattavasti eri vuosina projekteista johtuen.

Keskeistä skannauksen laatukriteereille ja tallekappaleen formaatille on pyrkimys säilyttää alkuperäisen asiakirjan informaatio häviöttömästi. Tästä johtuen reilut kahdeksankymmentä





prosenttia selvitykseen vastanneista tahoista tallensi digitoimansa tiedostot pakkaamattomassa TIFF-formaatissa. Merkittävä enemmistö skannasi tiedostot 24-bittisinä värikuvina 300 dpi:n resoluutiolla. Joissain tapauksissa käytettiin 16- tai 8-bittistä harmaasävydigitointia. Alankomaiden ja Yhdysvaltojen kansallisarkistot tallentavat myös JPEG-tiedostoja TIFF-tiedostojen lisäksi. Kumpikaan ei ole hävittänyt aineistoa analogisessa muodossa. Kansallisarkistoista vain Iso-Britannia käytti muuta kuin pakkaamatonta TIFF-tiedostoa tallekappaleenaan. Se on siirtynyt JPEG2000-formaattiin, joka mahdollistaa häviöttömän pakkauksen ja siten TIFF-tiedostoja pienemmän koon. Formaatti on ongelmallinen, koska se ei ole täysin vapaa patenteista, sillä ei ole selaintukea ja sen tuottamiseen tarvittavat ohjelmistot ovat kalliita.

Vain neljä kyselyyn vastanneista arkistoista oli laatinut ohjeistuksen analogisen säilytysmuodon hävittämiseksi digitoinnin jälkeen. Ne koskivat lähinnä terveydelle vaarallisia tai räjähdysherkkiä aineistoja. Australian kansallisarkisto oli ainoana vastaajista laatinut ohjeistuksia hävittämisen mahdollistavalle digitoinnille. Uusi-Seelanti ei osallistunut kyselyyn, mutta se on julkaissut toukokuussa 2017 hävittämisohjeen.

Suomen lisäksi kolme maata - Belgia, Norja ja Portugali – harkitsevat hävittämisohjeen tekemistä, koska niiden kansallinen laki sallii nykyään digitoinnin jälkeisen hävittämisen. Näistä kolmesta Belgia ja Norja olivat kuitenkin kriittisiä hävittämistä kohtaan. Alankomaiden, Islannin, Ruotsin ja Tšekin kansallisarkistot sekä Euroopan komission historiallinen arkisto ja Lüneburgin hansakaupunginarkisto painottivat vastauksissaan, että ne eivät arkistoina ole valmiita hävittämään asiakirjallista analogista kulttuuriperintöä. Selvitykseen vastanneissa maissa ei ole vielä vastaanotettu tai suunniteltu hyvin nuoren aineiston digitointia, vaan digitointi on keskittynyt arkistoissa jo olevaan, kaikkein käytetyimpään tai kulttuurihistoriallisesti arvokkaimman aineiston käytettävyyden parantamiseen.



Massadigitoointiprojektin lähimmät vertailukohtat lähitulevaisuudessa ovat ainakin Alankomaat, Ruotsi, Yhdysvallat ja Norjan *Helse arkivetin* potilasasiakirjojen digitointi. Jälkimmäisessä tapauksessa suunnitelmissa on ollut analogisten originaalien hävittäminen digitoinnin jälkeen. Alankomaiden kansallisarkisto suunnittelee digitoivansa seuraavan vuosikymmenen kuluessa noin 130 hyllykilometriä käytettyä aineistoaan. Ruotsin kansallisarkisto pyrkii saamaan 10 % aineistostaan digitoituksi seuraavan vajaan kymmenen vuoden aikana. Yhdysvallat on käynnistämässä projektia, jossa suunnitellaan suuren aineistomäärän digitointia hajautetusti.

Digitointikyselyn yhteydessä vastaajille kerrottiin lyhyesti massadigitointihankkeen suunnitelmista. Kyselyä välitettiin eri kansallisarkistojen digitoinnista vastaaville henkilöille sähköpostein sekä ICARUS-verkoston ja sosiaalisen median kautta. Tätä kautta tieto suunnitteilla olevasta digitoinnista on levinnyt laajasti arkistosektorilla. Kysely avattiin lähemmäs 900 kertaa. Kyselyn tulokset toimitetaan englanninkielisinä kaikille siihen vastanneille tahoille. Tieto massadigitoinnin suunnitteluprojektista, digitoitavan aineiston määrästä, karkeasta aikataulusta ja analogisen aineiston hävittämisestä digitoinnin jälkeen on levinnyt kaikkiin EU-maihin.

*Kansainvälisen selvityksen tulokset on kuvattu tarkemmin raportin liitteessä 2.*

### 3 Digitoitavat aineistot

#### 3.1 Valtionhallinnon analogisten asiakirjojen määrä

Valtionhallinnon analogisten asiakirja-aineistojen määrät ja aineiston digitoitavuuteen vaikuttavia ominaisuuksia kartoitettiin kaksivaiheisella aineistonkartoituskyselyllä. Kyselyt olivat kokonaisuudessaan laajoja ja niiden avulla saatiin aiempaa tarkempi kuva valtionhallinnon arkistoinnin tilasta ja analogisten asiakirjojen määrästä. Kyselyn ensimmäisen vaiheen vastausprosentti oli 92 %. Toisessa vaiheessa vastaukset saatiin kaikilta kyselyn vastaanottajilta.



Aineistokartoituskyselyn ensimmäisen vaiheen tulosten perusteella valtionhallinnon organisaatioiden hallussa on yhteensä noin 387 000 hm analogista asiakirja-aineistoa, josta noin 171 500 hm on pysyvästi säilytettävää ja 209 600 hm määräajan säilytettävää.<sup>1</sup> Määrä on huomattavasti suurempi kuin aiemmat arviot aineistomääristä ovat olleet. Suurin osa analogisista aineistosta on paperiasiakirjoja.

Aineistokartoituksen toinen vaihe toteutettiin osalle valtionhallinnon organisaatioita. Kysely kohdennettiin etupäässä aineistomääriltään suurimmille organisaatioille. Toisen vaiheen kyselyn vastaajat edustivat aineistomääriltään noin 80 % kyselyn ensimmäisen vaiheen vastaajista. Kyselyn toisessa vaiheessa kartoitettiin aineiston digitoitavuuteen vaikuttavia ominaisuuksia: tietoja aineistojen fyysisistä ominaisuuksista, laadusta, kunnosta sekä kuvailun ja järjestämisen tasosta. Kyselyn vastaukset annettiin arkistotasolla<sup>2</sup>. Tuloksena saatiin organisaatiotasoa yksityiskohtaisempaa tietoa aineiston ominaisuuksista.

Aineistokartoituksen toisen vaiheen tuloksissa saatu aineiston kokonaismäärä oli 266 000 hm, josta noin 124 000 hm on pysyvästi säilytettävää ja noin 140 000 hm on määräajan säilytettävää. Aineistotyypeistä noin 86 % on paperiasiakirjoja (sisältää kortistot), noin 7 % on sidoksia ja noin 7 % on erikoisaineistoja (kuten karttoja, piirustuksia, mikrofilmejä ja valokuvia) ja tyyppittelemättä jääneitä vastauksia. Tulosten perusteella noin 9 % aineistoista on järjestämättä ja noin 4,3 % aineistoista on vailla arvonmäärittystä. Digitoinnin suunnittelun kannalta merkittävää tietoa saatiin mm. irtonaisten paperiasiakirjojen määrästä (noin 50 700 hm), yhteen liitettyjen

---

<sup>1</sup> Valtaosa viranomaisten hallussa olevista analogisista aineistoista on rajattu joko pysyvään tai määräajaiseen säilytykseen jo Kansallisarkiston aiemmilla päätöksillä. Lisäksi suurin osa pysyvästi säilytettävästä aineistosta on jo seulottu ja erotettu määräajan säilytettävästä. Kansallisarkiston tavoitteena on ollut 1980-luvulta lähtien määrätä viranomaisten paperiasiakirjojen vuosikasvusta noin 10–15 prosenttia pysyvään säilytykseen, ja ottaa huomioon myös kustannustekijät. Kustannusnäkökohdat eivät kuitenkaan saa ensisijaisesti ratkaista, mitä asiakirjoja säilytetään pysyvästi. Asiakirjoista on pystyttävä seulonnan jälkeenkin toteamaan toiminnalliset, ajalliset, paikalliset ja asialliset syy- ja seuraussuhteet.

<sup>2</sup> Arkistolla tarkoitetaan arkistonmuodostajan toiminnan tuloksena syntyneitä asiakirjojen kokonaisuutta (ei arkistotilaa tai -rakennusta)



paperiasiakirjojen määrästä (noin 165 000 hm), erikoispaperien määrästä (noin 12 600 hm) ja käsinkirjoitetun aineiston määrästä (noin 83 000 hm).

Ensimmäisen vaiheen kyselyssä mukana oli myös maakuntauudistuksen kohteena olevat kunnalliset organisaatiot. Näiden organisaatioiden kohdalla kyselyssä kartoitettiin sitä aineistomäärää, joka toiminnan jatkumisen ja tehtävien hoitamisen kannalta on siirrettävä maakuntauudistuksen myötä aloittaville uusille maakunnille. Uusille maakunnille siirtyviä aineistoja on tulosten perusteella arviolta noin 22 000 hm. Koska toistaiseksi ei ole tehty selkeää määrittelyä siitä, mitä aineistoja tuleville maakunnille siirtyy, oli maakuntakyselyyn vastaajille vastausten antaminen haastavaa. Tämä vaikuttaa tulosten luotettavuuteen ja se on huomioitava maakuntakyselyn tuloksia käsiteltäessä.

*Aineistokartoitusten tulokset on kuvattu tarkemmin liitteessä 1.1.*

### 3.2 Digitoitavat aineistot

Massadigitoitavan aineiston määrä, aineistotyytit ja aikataulu on esitetty aineistokartoitusten pohjalta laaditussa tiekartassa. Tiekartta on suunnittelun väline ja sitä tullaan päivittämään ja ylläpitämään massadigitoinnin suunnittelun ja toteutuksen ajan. Massadigitointiin ja tiekarttaan sisältyvää aineistoa arvioitiin olevan valtionhallinnossa yhteensä 212 000 hm. Massadigitointiin soveltuva aineisto jaettiin kolmeen aineistotyyppikohtaiseen pääluokkaan: A4/Folio-kokoiset paperiaineistot (72 %), kortistoaineistot (14 %) ja sidokset (14 %).

Aineistokartoitusten myötä kävi ilmi, että osa viranomaisten hallussa olevasta analogisesta aineistosta ei sovellu massadigitointiprosessiin, joka edellyttää aineistoilta nopeaa ja helppoa digitoitavuutta. Massadigitoinnin (212 hyllykilometriä) ulkopuolelle on rajattu seuraavat aineistot:



- Erikoisaineistot (kuten kartat, piirustukset, mikrofilmit ja -kortit sekä AV-aineistot) omina kokoelminaan. Erikoisaineistoja, kuten karttoja ja piirustuksia, voi sisältyä massadigitointiin silloin kuin ne ovat muun aineiston seassa (esim. samoissa säilytysyksiköissä tai samoissa akteissa).
- Fyysisiltä ominaisuuksiltaan massadigitointiin sopimattomat aineistot (esimerkiksi erikoispaperit, erityisen hauraat aineistot ja homeiset aineistot).
- Tietosisällöltään toisiaan vastaavat analogisessa ja digitaalisessa muodossa olevat aineistot
- Huomattava osa määrääjän säilytettävistä asiakirjoista
- Suuri osa turvaluokitellusta aineistosta sekä suojaustasojen STI I ja ST II aineistot.
- Tiettyjen organisaatioiden aineistot: Jos organisaatiolla ei ole arkistolain mukaista siirtovelvollisuutta siirtää aineistoja Kansallisarkistolle, organisaatiolla on poikkeuslupa olla luovuttamatta aineistoja Kansallisarkistolle, organisaation hallussa on vain hyvin vähäinen määrä pysyvästi säilytettävää aineistoa tai organisaation aineistot ovat esimerkiksi hitaasti digitoitavia erikoisaineistoja tai muuten fyysisiltä ominaisuuksiltaan digitointiin sopimattomia.

Massadigitoinnin suunniteltu aineistomäärä 212 hyllykilometriä on kansainvälisesti katsoen arkistosektorilla hyvin mittava aineistomäärä. Digitoitavaksi suunniteltu aineistomäärä on moninkertaisesti suurempi kuin toistaiseksi aineistoja eniten digitoineen Yhdysvaltojen kansallisarkiston digitoitu aineistomäärä. (Massadigitoinnissa 212 hyllykilometrin digitoinnista syntyisi arviolta 3,4 miljardia kuvaa, Yhdysvaltojen kansallisarkisto on digitoinut yhteensä 390 miljoonaa kuvaa, vuosituotanto on 30 miljoonaa).

Täyden tuotannon aikana massadigitoinnissa on suunnitelmassa digitoida 22 hyllykilometriä vuodessa. Tämä tarkoittaa, että vuodessa syntyy noin 350 miljoonaa kuvaa, mikä on enemmän aineistoa kuin esimerkiksi kaikissa Pohjoismaiden kansallisarkistoissa on tähän mennessä digitoitu yhteenlaskettuna. Yhteenvetona voi todeta, että massadigitoinnin suunnitellut aineistomäärät eivät ole samassa mittaluokassa verrattuna arkistojen nykyisiin digitoinnin määriin.

### 3.3 Digitoinnin aikataulu ja järjestys

Viranomaisten aineistojen digitoinnin järjestystä ja aikataulua ohjaavassa tiekartassa viranomaisten aineistot on jaettu neljään ryhmään massadigitoinnin aikataulun ja aineistojen digitoitavuuteen vaikuttavien kriteerien perusteella. Kriteerit pohjautuvat aineistokartoituksissa kerättyihin tietoihin, kuten irtonaisten paperien määrään sekä järjestämisen ja kuvailun tasoon. Esitettyjen kriteerien tarkoituksena on ohjata viranomaisten aineistojen sijoittamista ryhmiin. Kriteerit eivät ole ehdottomia, vaan lopullinen järjestys tulee perustumaan kokonaisarvioon. Massadigitoinnissa digitoitavien aineistojen aikataulu ja tiekartan ryhmäjako aineistomäärineen ja kriteereineen on esitetty taulukossa 1.

Taulukko 1. Massadigitoinnissa digitoitavien aineistojen aikataulu ja tiekartan ryhmäjako (ryhmät 1 – 4)

Aikataulu	7/2019–7/2021	7/2021–2023	2024–2026	2027–7/2030
<b>Ryhmät</b>	1. ryhmä	2. ryhmä	3. ryhmä	4. ryhmä
<b>Määrät</b> (hyllykilometriä = hkm) <b>Yht. 212 hkm</b>	12 hkm	~52 hkm	~67 hkm	~81 hkm
<b>Kriteerit</b>	<ol style="list-style-type: none"><li>Aineistojen nopea ja helppo digitoitavuus</li><li>Viranomaisen kyvykkyys saattaa aineistot digitoinnin siirtokuntoon</li><li>Aineiston suuri käyttötarve</li></ol>	<ol style="list-style-type: none"><li>Aineiston suuri käyttötarve</li><li>Organisaatio- ja arkistotilamuutokset</li><li>Toimintansa pitkälle digitalisoineet organisaatiot → loppujen paperien digitointi</li></ol>	<ol style="list-style-type: none"><li>Aineistot hitaasti digitoitavia</li><li>Aineistojen vähäinen käyttötarve</li></ol>	<ol style="list-style-type: none"><li>Aineistot hitaasti digitoitavia</li><li>Aineistojen vähäinen määrä / ei voida luopua arkistotiloista</li><li>Organisaatio ei vielä toimi digitaalisesti</li></ol>



Tiekartan laadinta jatkuu kevään 2018 aikana, jolloin yhteistyössä valtionhallinnon viranomaisten kanssa valmistellaan organisaatio- ja arkistokohtainen digitointijärjestys. Tämän työn tuloksena konkretisoituu millä aikataululla kunkin organisaatiot aineistot tultaisiin digitoimaan. Tiekartan valmistelussa ovat tähän mennessä olleet mukana vain valtionhallinnon organisaatiot, sillä kunnilta maakunnille siirtyvien aineistojen digitoinnista ei ole tehty linjauksia.

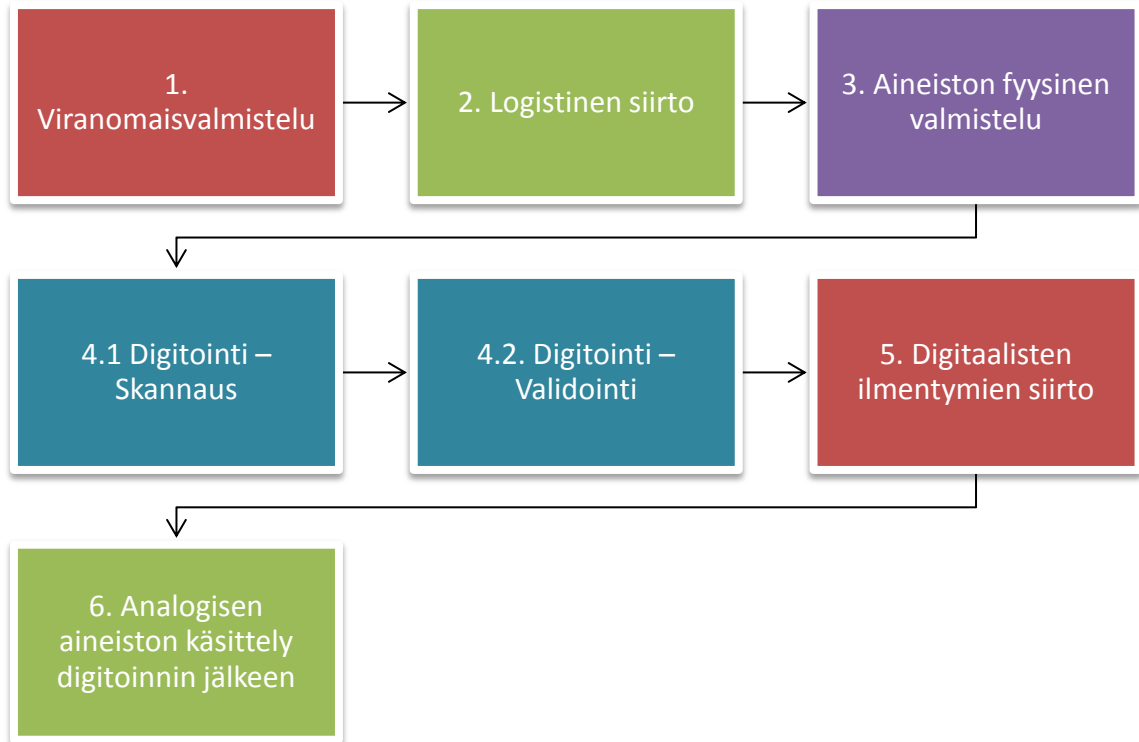
*Digitoitavat aineistot ja niiden digitointijärjestys on kuvattu tarkemmin liitteessä 1.2*

#### 4 Massadigitoinnin prosessi

Massadigitoinnin suunnittelussa on määritelty yksi prosessikokonaisuus, jonka mukaisesti kaikki massadigitointiin kuuluvat aineistot tullaan digitoimaan. Prosessissa on tunnistettu seuraavat toimijat ja niiden tehtävät:

- *Viranomainen* omistaa digitoitavaksi tulevan aineiston. Viranomainen vastaa aineiston viranomaisvalmistelun toimenpiteistä.
- *Kansallisarkisto* vastaa viranomaisten valmistelun ohjauksesta. Kansallisarkisto koordinoi ja valvoo tiekarttasuunnitelman aikataulun toteutumista sekä viranomaisvalmistelun että digitoinnin etenemisen osalta. Kansallisarkisto vastaa säilytyspäättösten valmistelusta.
- *Digitoiva taho* on toimija, joka vastaa massadigitoinnin käytännön toteuttamisesta alkaen aineiston logistisen siirron suunnittelusta ja päättyen digitoidun analogisen aineiston hävittämiseen tai sijoittamiseen Kansallisarkiston säilytykseen.

Digitoinnin kokonaisprosessi (kuva 1) jakaantuu kuuteen erilliseen toiminnalliseen prosessiin. Jokaisen prosessin päätepiste toimii syötteenä seuraavalle prosessille.



Kuva 1. Massadigitoinnin prosessikokonaisuus

Massadigitoitava aineisto on valtaosin pysyvästi säilytettäväksi määriteltyä asiakirjallista kulttuuriperintöä, jonka analoginen manifestaatio tullaan hävittämään sen digitaaliseen muotoon muuntamisen jälkeen. Digitoinnin prosessi saa näin useita erilaisia vaatimuksia lähtien analogisen aineiston käsittelytavoista, päätyen siihen, että aineistosta tulee tuottaa todistusvoimainen, autenttinen ja tietosisällöltään täydellinen digitaalinen ilmentymä. Pitkäaikaissäilytykseen on saatettava tallekappale, joka on mahdollisimman laadukas ja josta on tulevaisuudessa tuotettavissa erilaisia käyttökappaleita sen hetken tarpeisiin.

Prosessikokonaisuuden hallitussa läpiviennissä keskeinen osa on tuotannonohjausjärjestelmällä, jonka avulla hallitaan digitointiin tulevia aineistoja, aineistojen siirtoja sekä kaikkia aineistolle tehtäviä toimia digitointipaikassa. Tuotannonohjausjärjestelmällä hallitaan, että aineistoa ei kasaudu jollekin toiminnolle kerralla liikaa ja toisaalta että kaikille toiminnoille on riittävästi





aineistoa. Tuotannonohjauksella siis pyritään välttämään tuotannossa pullonkauloja ja tyhjäkäyntiä.

*Massadigitoinnin prosessi on kuvattu yksityiskohtaisemmin liitteissä 3.1 ja 3.2*

#### 4.1 Prosessin ennakkoehdot

Digitoitavalla aineistolla tulee olla Kansallisarkiston säilytyspäätos, jossa on määrätty pysyvästi säilytettävän asiakirjatiedon analogisten ilmentymien säilyttämisestä tai hävittämisestä digitoinnin jälkeen. Säilytyspäätoksessä arvioidaan digitoitavan asiakirjan säilytysmuoto sekä todistusvoimaisuuden että kulttuurihistoriallisen arvon perusteella. Arviona on, että noin 10 % digitoitavasta aineistosta tullaan säilyttämään myös analogisessa muodossa digitoinnin päätteeksi. Pääsääntönä on, että säilytyspäätoksessä ei enää muuteta Kansallisarkiston aiemmin tekemää arvonmäärittystä (säilytysaikaa) asiakirjatiedolle.

Digitoitavaksi tulevien aineistojen järjestys ja siirtoaikataulu määräytyvät tiekarttasuunnitelman mukaan. Viranomaisten tulee sitoutua tiekarttasuunnitelman mukaisiin aikatauluihin ja ajoittaa omat valmistelevat toimet siten, että aineisto on siirrettävissä digitointiin tiekartan määrittämässä aikataulussa. Sitoutuminen tiekartan aikatauluihin on kriittinen tekijä massadigitoinnin onnistumisen kannalta.

#### 4.2 Viranomaisvalmistelu

Viranomainen valmistelee aineiston logistista siirtoa varten. Viranomaisella on yksilöitynä tiekartan perusteella aineistokokonaisuus, joka digitoidaan. Tiekartassa on määritelty aineiston digitoinnin aikataulu.

Viranomaisvalmistelun tavoitteena on mahdollistaa aineiston luotettava siirto digitointipaikalle. Sen tavoitteena ei ole valmistella aineistoa siihen kuntoon, että se voidaan suoraan digitoida.



Viranomaisvalmistelu vaatii Kansallisarkiston nykyistä siirtomenettelyä vähemmän toimenpiteitä, ja aineiston fyysinen valmistelu voidaan tehdä aiempaa kevyemmin. Viranomaisen tulee yhä muun muassa luetteloida aineisto Kansallisarkiston luettelointijärjestelmään sekä suojata aineisto. Säilytysyksiköiden ei kuitenkaan tarvitse olla arkistokelpoisia vaan riittää, että ne suojaavat aineiston kuljetuksen ajan ja estävät aineistoa sekoittumasta muiden aineistojen kanssa. Viranomaisen huolehtii myös säilytysyksiköiden yksilöinnistä, joka tapahtuu luettelointijärjestelmästä saatavilla viivakoodeilla. Digitointipaikassa suoritetaan aineiston varsinainen valmistelu skannaukseen.

Kansallisarkisto tulee tarkemmin ohjeistamaan viranomaisvalmistelun uudet menettelyt. On tärkeää huomioida, että viranomaisvalmistelun tavoite muuttuu aiemmasta perustavanlaatuisesti: aiemmin aineiston valmistelu on tähdännyt aineiston analogisen muodon säilyvyyden turvaamiseen. Uudessa menettelyssä lähtökohtana on kuitenkin analogisten ilmentymien hävittäminen digitoinnin päätteeksi.

Viranomaisvalmistelun viimeisessä vaiheessa Kansallisarkiston asiantuntija valmistelee yhdessä viranomaisen kanssa aineistosta käsittelysuunnitelman, jossa kuvataan aineiston valmisteluun ja digitointiin liittyviä erikoispiirteitä. Käsittelysuunnitelman tavoitteena on helpottaa prosessin vaiheiden suunnittelua digitointipaikassa ja tuoda siten ennakoitavuutta. Prosessin lopputuloksena on *siirtovalmis aineisto*.

### 4.3 Logistinen siirto

Digitoitavat aineistot ovat hajaantuneena eri viranomaisten arkistotiloihin. Aineistot tulee saada siirrettyä luotettavasti ja koordinoitusti eri arkistotiloista yhteen keskitettyyn digitointipaikkaan. Digitointipaikkaan ei ole tarkoituksenmukaista kerätä laajaa välivarastoa. Tuotannonohjausjärjestelmällä on keskeinen rooli logistiikassa ja siinä ylläpidetään tietoa siitä, miten paljon aineistoa on eri viranomaisten tiloissa *siirtovalmiina*.



Logistisista kuljetuksista vastaa kokonaisuudessaan digitoiva taho, joka suunnittelee ja tilaa kuljetukset viranomaisten tiloissa olevalle aineistolle. Viranomaiset eivät osallistu kuljetusten järjestämiseen. Logistisen siirron aikana aineisto on koko ajan jäljitettävissä viranomaisten tiloista kuljetusvälineeseen ja digitointipaikan välivarastoon tuotannonohjauksessa hallittavan tilatiedon perusteella sekä kuljetuksien seurantaan käytettävän GPS-paikannuksen avulla.

Digitointipaikassa kuljetettu aineisto siirretään rullakoissa digitointipaikan välivarastoon odottamaan digitointia. Prosessin lopputuloksena on *kuljetettu aineisto*.

*Massadigitoinnin logistiikka on kuvattu tarkemmin liitteissä 3.4 ja 3.5*

#### 4.4 Digitoinnin valmistelu

Digitoinnin valmistelulla varmistetaan, että digitointi sujuu tehokkaasti ja turvallisesti. Valmistelu on järkevää keskittää digitointipaikkaan, sillä tarvittavat toimenpiteet edellyttävät muun muassa eri materiaalien käsittelyyn liittyvää erityisosaamista sekä skannauslaitteiston tuntemusta. Puutteellisesti valmisteltu aineisto pysäyttää digitoinnin ja saattaa vahingoittaa laitteistoa tai aiheuttaa pahimmassa tapauksessa informaation häviämistä.

Valmistelutoimenpiteissä pitää ottaa huomioon Kansallisarkiston antama säilytyspäättös sekä asiakirjatiedon tallennusala eli aineistotyyppi. Jos asiakirjatieto säilytetään digitoinnin jälkeen myös analogisessa muodossa, pitää valmistelutoimenpiteiden ja aineistoon lisättävien materiaalien olla arkistokelpoisia. Valmistelutoimet pyritään pitämään mahdollisimman vähäisinä ja selkeinä. Valmistelua pyritään mahdollisuuksien mukaan tehostamaan vaarantamatta kuitenkaan aineistojen tietosisältöä.



Valmistelu vaatii digitointiprosessissa paljon aikaa, sillä iso osa työstä pitää tehdä manuaalisesti. Automatiikkaa ei voida useimmiten hyödyntää, koska se lisää riskejä informaation häviämiseksi ennen digitointia. Viranomaisvalmistelun yhteydessä syntyvän käsittelysuunnitelman avulla voidaan ennakoida tarvittavia resursseja ja toimenpiteitä.

Valmistelun yhteydessä tehdään päätökset, millä laitteella aineisto skannataan. Pääosin aineistot valmistellaan suurtehoskannerille skannattavaksi irtoarkkeina. Sidokset voidaan digitoida purkamisen jälkeen. Jos aineistoa ei ole mahdollista skannata suurtehoskannerilla, valmistellaan aineisto erikoisskanneria varten. Prosessin lopputuloksena on *digitointivalmis aineisto*.

#### 4.5 Digitointi

Digitointi pitää sisällään digitointivalmiin analogisen aineiston muuntamisen sähköiseen muotoon (skannaus kuvaksi), kuvan laadun validoinnin, kuvan sisällön tunnistuksen OCR-menetelmin ja siirtopaketin muodostuksen.

Skannausprosessin tavoitteena on tuottaa yhdellä skannauksella sekä tallekappale että käyttökappale. Skannauksessa jokainen arkistoyksikkö on yksilöity ja arkistoyksikön sisältö skannataan sen alkuperäisessä järjestyksessä.

Validoinnilla tarkoitetaan skannauksen jälkeen tapahtuvaa kuvan laadunvarmistusta. Validoinnissa tarkastetaan skannausvaiheessa syntyneiden digitaalisten ilmentymien määritysten mukaisuus sekä digitaalisen manifestaation eheys.

Sisällön tunnistus toteutetaan OCR-menetelmällä ja se kohdennetaan alkuvaiheissa vain niihin digitaalisiin ilmentymiin, joiden tietosisältö on pääsääntöisesti konekirjoitettua. OCR-tekstintunnistusta hyödynnetään myös osana kuvanlaadun validointia.

On korostettava, että laadunvarmistus on kokonaisuutena laajempi kuin pelkkä validointi. Kokonaisprosessissa ei ole vain yhtä kohtaa, jossa laatua varmistetaan. Laadunvarmistukseen kuuluu jo ennen skannausta tehtävä laitteistojen ja ohjelmistojen toimivuuden todentaminen. Asiakirjan tai siihen sisältyvän tiedon säilyminen, eheys ja autenttisuus pyritään turvaamaan sillä, että koko prosessi aineiston viranomaisvalmistelusta sen lopulliseen hävittämiseen asti on hallittu ja aineistolle prosessin eri vaiheissa tehtävistä toimenpiteistä tallentuu tieto.

Prosessin lopuksi digitaalisista ilmentymistä, OCR-tekstitunnistuksesta ja metatiedoista luodaan siirtopaketti. Lopputuloksena on *lähetysvalmis digitoitu aineisto*.

## Digitaaliset ilmentymät

Taulukko 2. Analogisesta manifestaatiosta tuotettavat digitaaliset ilmentymät

Tiedoston rooli	Tiedostoformaatti	Pakkaus	Resoluutio	Bittisyvyys	Soveltuva ICC-profiili
Tallekappale	TIFF 6.0	LZW (häviötön)	300 ppi	24	sRGB (tai laajempi)
Käyttökappale	JPEG	Jpeg-laatu 60%	300 ppi	24	sRGB
Tunnistettu teksti	ALTO (3.0 tai 3.1) XML <sup>3</sup>	-	-	-	-
Digitointiprosessia kuvaavat metatiedot	XML (MIX-skeema <sup>4</sup> )	-	-	-	-

Jokaisesta tietosisältöä sisältävästä asiakirjan sivusta (etu- ja kääntöpuoli) tuotetaan saman tietosisältöiset talle- ja käyttökappaleet. Kuvien resoluutio, bittisyys ja värillisyyset perustuvat digitaalisen ilmentymän todistusvoimaisuuden varmistamiseen ja analogisten ilmentymien

<sup>3</sup> The Library of Congress » Standards » ALTO. Kongressin kirjaston verkkosivu. Viitattu 19.12.2017. <https://www.loc.gov/standards/alto/>

<sup>4</sup> The Library of Congress » Standards » MIX. Kongressin kirjaston verkkosivu. Viitattu 20.12.2017. <http://www.loc.gov/standards/mix/>



sisältämän informaation häviöttömään säilyttämiseen. Tallekappaleet ovat häviöttömiä tiedostoja, jotta digitaalisen kuvan laatu ja siten sen sisältämä informaatio ja todistusvoimaisuus eivät kärsi tiedostojen mahdollisissa tulevilla migraatioissa. Lisäksi häviölliset tiedostot, kuten JPEG, laskevat OCR-tarkkuutta. Tallekappaleiden yhteyteen tuotetaan MIX-skeeman mukainen XML-tiedosto, joka dokumentoi digitointiprosessin. Jokaisesta konekirjoitettua tietosisältöä sisältävästä asiakirjan sivusta tallennetaan oma XML-tiedostonsa Analyzed Layout and Text Object (ALTO) -muodossa. Tämä mahdollistaa tekstihaut digitaalisiin ilmentymiin.

*Vaatimukset digitoinnissa tuotettaville digitaalisille ilmentymille on kuvattu tarkemmin liitteessä 3.3.*

#### 4.6 Digitaalisten ilmentymien siirto

Digitaalisten ilmentymien siirto toteutetaan automatisoidusti siten, että tuotannonohjausjärjestelmä vastaa siirtopaketin siirtämisestä vastaanottavaan järjestelmään, jonka tehtävänä on tarkastaa siirtopaketti sekä kuitata siirron onnistuminen tai epäonnistuminen tuotannonohjausjärjestelmälle. Vastaanotettavan järjestelmän on suunniteltu olevan sähköisen arkistoinnin palvelujärjestelmä SAPA.

Hyväksytyyn vastaanoton jälkeen prosessi on valmis ja sen lopputuloksena on *digitoidun aineiston siirto valmis*.

#### 4.7 Analogisen aineiston käsittely digitoinnin jälkeen

Analogisen aineiston käsittely on riippuvainen aineistolle annetusta säilytyspäätöksestä eli siitä hävitetäänkö analoginen aineisto digitoinnin jälkeen vai säilytetäänkö se. Skannauksesta tullut analoginen aineisto välivarastoidaan määrätyn varoajaksi, jonka jälkeen se joko tietoturvallisesti hävitetään tai koteloidaan ja sijoitetaan Kansallisarkistoon säilytykseen. Mikäli analoginen aineisto säilytetään digitoinnin jälkeen, vastaa digitoiva taho aineiston koteloinnista digitoinnin jälkeen. Kansallisarkisto vastaa aineiston sijoittamisesta säilytykseen.



## 5 Massadigitoinnin tuotanto

Massadigitoinnin tuotannon suunnittelu on tehty tiekartassa selvitetyn aineistomäärän (212 hyllykilometriä) perusteella. Tavoitteena on, että aineisto on kokonaisuudessaan digitoitu vuoteen 2030 mennessä. Kaksi ensimmäistä vuotta käytetään tuotannon pystyttämiseen, josta tuotanto ajetaan ylös täyteen kapasiteettiin kolmannelta vuodelta eteenpäin. Aineistomäärällä mitattuna tämä tarkoittaa 12 hyllykilometrin digitoimista kahden ensimmäisen vuoden aikana (*tuotannon aloitusvaihe*) ja 200 hyllykilometrin digitoimista yhdeksän seuraavan vuoden aikana (*tuotantovaihe*). Kokonaisuutena tulee digitoitavaksi yhteensä noin **1,7 miljardia arkkiä** ja osan kaksipuolisuus huomioiden lähes **2,55 miljardia kuvaa**, josta syntyy **40 petatavua dataa**. Määrät perustuvat arvioon 8000 asiakirjaa/hyllymetri. Asiakirjojen määrä hyllymetrillä voi analogisen ilmentymän olomuodosta johtuen vaihdella merkittävästi. Arvion mahdollisista epätarkkuuksista huolimatta mittakaava on kansainvälisen selvityksen perusteella huomattavasti suurempi, kuin aiemmissa arkistojen toteuttamissa digitointiprojekteissa.

Tästä seuraavat vaatimukset tuotantokapasiteetille on esitetty taulukoissa 3–4. Luvut perustuvat tiekarttatyön selvityksiin sekä arvioihin eri laitteiden ja henkilöstön tuotantokapasiteetista.

Taulukko 3. Tuotantomäärät tuotannon aloitusvaiheessa

<b>Aineistoa digitoitava vuosittain</b>	<b>6000 hyllymetriä, n. 48 miljoonaa arkkiä</b>
<b>Aineistoa digitoitava työpäivässä (250 työpäivää)</b>	<b>24 hyllymetriä, n. 192 000 arkkiä</b>
- josta suurteholinjaston aineistoa	22 hyllymetriä, n. 176 000 arkkiä
- josta suurteholinjastolle sopimatonta erikoisaineistoa	2 hyllymetriä, n. 16 000 arkkiä

Taulukko 4. Tuotantomäärät täyden tuotannon ajalta

<b>Aineistoa digitoitava vuosittain</b>	<b>22250 hyllymetriä, n. 178 miljoonaa arkkia</b>
<b>Aineistoa digitoitava työpäivässä (250 työpäivää)</b>	<b>89 hyllymetriä, n. 712 000 arkkia</b>
- josta suurteholinjaston aineistoa	80 hyllymetriä, n. 640 000 arkkia
- josta suurteholinjastolle sopimatonta erikoisaineistoa	9 hyllymetriä, n. 72 000 arkkia

### Massadigitoinnin skannauslaitteistot

Massadigitointia varten suunnitellulla skannausinfrastruktuurilla mahdollistetaan paperiaineistojen skannaus. Taulukossa 5 on esitetty suunniteltu massadigitoinnin edellyttämä laitteisto.

Taulukko 5. Massadigitoinnissa käytettävä skannauslaitteisto

Laite	Teknologia	Käyttötarkoitus
<b>Suurtehoskanneri</b>	Avorata-skanneri, jossa analoginen aineisto ei kulje kahden telan läpi. Mahdollistaa yhtäaikaaisesti asiakirjan molempien puolien skannaamisen	Suurtehoskannerille valmistellut irtoarkit. Pääsääntöinen skannaustapa.
<b>Suurkokoskanneri</b>	Läpisyöttöskanneri suurikokoiselle aineistolle, joka mahdollistaa maksimissaan ~140cm analogisen aineiston skannaamisen. Mahdollistaa yhtäaikaisesti asiakirjan molempien puolien skannaamisen	Erikoisaineistot. Paperiarkki, joka on suurempi kuin A3.
<b>Mastoskanneri</b>	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.
<b>Dokumenttiskanneri</b>	Pöytämallinen läpisyöttöskanneri. Mahdollistaa yhtäaikaisesti asiakirjan molempien puolien skannaamisen.	Erikoisaineistot. Suurtehoskannerin vaihtoehtoinen tapa toteuttaa alle A3-koon irtoarkkien skannaus. Käytetään harkinnan mukaan.





Hyvin suunnitellulla sekä ohjatulla logistiikalla ja aineiston valmistelulla mahdollistetaan tehokkain ja tuloksellisin skannausvaihe. Aineiston valmistelussa digitointia varten aineisto lajitellaan ja ohjataan skannattavaksi oikealle skannerille. Suurtehoskannerit ovat itsenäisiä linjastoja, joissa yksi skanneri muodostaa yhden tuotantolinjaston. Suurtehoskannerilla on tarkoitus skannata jatkuvasti ja mahdollisimman tehokkaasti sille soveltuvaa aineistoa. Aineisto, joka ominaisuuksiltaan edellyttäisi suurtehoskannerin hidastamista, ohjataan skannattavaksi toiselle laitteelle. Suurtehoskannereille soveltumaton aineisto skannataan erikoisaineistolinjalla, joka koostuu dokumenttiskannereista, suurkokoskannereista sekä mastoskannereista

### Muut laitteistot

Skannauslaitteiston lisäksi prosessi edellyttää myös muita laitteisto- ja työkaluhankintoja. Aineistojen valmisteluun tarvitaan erilaisia välineitä, kuten leikkureita sidosten purkamiseen, lämpölusikoita tai vastaavia asiakirjojen taitosten suoristamiseen, imureita sekä metallinpaljastin niittien ja muiden metallisten liittimien tunnistamiseen. Digitointipaikalla pitää olla myös esimerkiksi vetokaappi, jos homeista aineistoa valmistellaan digitointiin. Lisäksi käytettävissä pitää olla asianmukaiset henkilösuojaimet. Logistiikka edellyttää usean sadan rullakon käyttöä aineiston säilyttämiseen ja liikutteluun tuotantoprosessin aikana. Analogisten aineistojen hävittäminen saattaa aiheuttaa myös tarpeita silppureiden ja paalaimen hankinnalle, mikäli tätä ei ulkoisteta. Myös tietotekninen infrastruktuuri on mittava palvelinratkaisusta yksittäisiin päätelaitteisiin.

Oman kokonaisuutensa muodostavat tuotannossa tarvittavat järjestelmät ja ohjelmistot. Koko tuotantoprosessin hallinta toteutetaan tuotannonohjausjärjestelmällä. Muita merkittäviä ohjelmistotarpeita ovat itse skannausohjelmistot sekä sisällöntunnistusohjelmisto.



## Prosessin henkilöstötarve

Prosessin läpiviemiseksi on tunnistettu taulukossa 6 esitetyjä rooleja.

Taulukko 6. Massadigitoinnin tuotannossa tunnistetut roolit

Rooli	Tehtävä
<b>Varastotyöntekijä</b>	Aineiston vastaanotto ja liikuttaminen prosessin eri vaiheissa. Muun muassa skannatun materiaalin purkaminen suurtehoskannereista ja sen varastointi välivarastoon.
<b>Valmistelija</b>	Valmistelee aineistoa eri skannausprosesseihin.
<b>Suurtehoskannerin operaattori</b>	Skannaus suurtehoskannerilla. Lastaa skanneria ja skannaa niput.
<b>Erikoislinjaston operaattori</b>	Skannaus erikoislinjastolla. Lastaa skanneria ja skannaa aineistoa.
<b>Validoija</b>	Validoi eli tarkastaa skannaustyön laatua silmämääräisesti.
<b>Työnjohto</b>	Normaalit työnjohdolliset tehtävät ja päivittäispäätösten tekeminen.
<b>Tietopalveluhenkilöstö</b>	Vastaa digitointipaikalla olevaan aineistoon tulevien aineistopyyntöjen hoitamisesta. Tyypillisimmin aineistopyyntö kohdistuu aineistoon, joka on välivarastossa odottamassa valmistelua.

Edellä mainittujen roolien lisäksi työpanosta tarvitaan koko tuotannon suunnittelussa, johtamisessa ja logistiikan hallinnassa. Lisäksi tarvitaan teknistä henkilöstöä vastaamaan skannauslaitteistojen, tietoteknisen infrastruktuurin ja ohjelmistojen jatkuvasta ylläpidosta.

## Digitointipaikan tilat

Massadigitoinnin aloittaminen edellyttää riittäviä ja käyttötarkoitukseen soveltuvia tiloja. Tilojen tarve kasvaa asteittain digitoinnin volyymien kasvaessa. Tiloissa tulee erityisesti huomioida saapuvan aineiston ja hävittämistä odottavan aineiston välivarastoinnin edellyttämät tilatarpeet, aineiston helppo liikuteltavuus tiloissa sekä ison henkilöstömäärän edellyttämien sosiaalitulojen tarpeet. Lisäksi on muita erityistarpeita esim. sähkönjakeluun ja ilmastointiin liittyen.

Massadigitoinnin suunnitteluprojektissa on selvitetty digitoinnin käynnistämistä Kansallisarkiston Mikkelin keskusarkiston tiloissa. Digitoinnin aloittamista Mikkelin keskusarkistossa puoltaa se, että keskusarkistossa olevat työskentelytilat voidaan käytännössä kokonaan ottaa digitoinnin



käyttöön. Näin ollen tiloista ei tule lisäkustannuksia digitoinnille. Lisäksi keskusarkiston käyttöä puoltaa siellä vapaana olevan makasiinitilan hyödyntäminen digitoinnin välivarastointiin. Digitoinnin sijoittamiselle Mikkeliin on myös muita etuja, kuten Kansalliskirjaston digitoitotoiminnon sijainti Mikkelissä sekä alueen vahva digitaaliseen tiedonhallintaan liittyvä osaamiskeskittymä.

Projektissa toteutettiin digitoinnin tilasuunnittelu keskusarkiston tiloihin. Suunnittelutyön perusteella keskusarkisto voidaan vähäisillä muutostöillä ottaa massadigitoinnin tuotannon käynnistysvaiheen tiloiksi. Tuotannon aloitusvaiheen jälkeen digitoinnin täysi tuotanto edellyttää kuitenkin suurempia tiloja kuin keskusarkistossa on tähän käyttötarkoitukseen olemassa.

## 6 Kustannukset ja hyödyt

### 6.1 Digitoinnin kustannukset

Massadigitoinnin kustannusarvion lähtökohtana on tiekartan mukainen aineistomäärä 212 hyllykilometriä sekä periaatepäätöksen linjaus digitoinnin toteuttamisesta vuoteen 2030 mennessä. Kustannukset perustuvat luvussa 5. kuvatun tuotannon toteuttamiseen.

Massadigitoinnin kustannuksista lähes 80 % tulee tuotannon henkilötyöstä. Lopullisten kustannusten kannalta on erittäin suuri merkitys sillä, miten paljon työtä aineiston valmistelu ja itse skannaus tulevat vaatimaan. Työmäärään vaikuttavat käytössä olevat menetelmät ja teknologiat eli miten valmistelun ja skannauksen toteuttamista voidaan tehostaa. Merkittävä vaikutus on ennen muuta digitoitavalla aineistolla, koska vaikeasti käsiteltävän aineiston valmistelun ja skannauksen työmäärä voi olla yli kymmenkertainen helposti valmisteltavaan aineistoon nähden. Alkuvaiheessa kustannuksia synnyttävät myös laitteistojen, tietojenkäsittelykapasiteetin sekä ohjelmistojen edellyttämät investointikulut.



Kustannusten arvioinnin lähtökohtana on digitoinnin toteuttaminen käyttäen nykyisiä kaupallisesti saatavilla olevia teknologioita (esimerkiksi skannauslaitteisto ja OCR-tunnistus). Kustannuksia ei ole lähdetty arvioimaan mahdollisesti tulevaisuudessa käyttöön tulevien ratkaisujen pohjalta. Tästä huolimatta kustannusarvioon sisältyy epävarmuustekijöitä, joista merkittävin liittyy digitoitavaan aineistoon: Aineiston määrä ja sen ominaisuudet voivat poiketa merkittävästikin aiemmin arvioidusta, millä voi olla huomattavia vaikutuksia aineiston valmistelun ja skannauksen työmäärään. Toisaalta, jos aineistosta aiempaa suurempi osuus osoittautuu helpommin digitoitavaksi, on työmäärä ja siten kustannukset alhaisempia.

## 6.2 Digitoinnin hyödyt

Massadigitoinnin keskeisenä hyötynä on aineiston saatavuuden ja käytettävyyden merkittävä parantuminen digitoinnin myötä. Aineistot ovat digitaalisessa muodossa saatavilla ajasta ja paikasta riippumatta ja tietoa voidaan hakea käyttäen moderneja tiedonhaun menetelmiä, kuten kokotekstihakua sekä erilaisia ontologioita. Julkiset aineistot voidaan saattaa entistä helpommin saataville suoraan verkossa, mikä mahdollistaa tiedon haun ja palveluiden tuottamisen itsepalveluperiaatteella.

Taloudellisia hyötyjä arvioitaessa on lähdetty niistä hyödyistä, jotka voidaan osoittaa suoraan valtionhallinnon budjettitalouteen. Siten esimerkiksi kansalaisten parantuneeseen tiedonsaantiin tai tutkimuksen käyttämän tiedon saatavuuteen ja käytettävyyteen liittyviä hyötyjä ei ole arvioitu osaksi taloudellisia hyötyjä. Nämä hyödyt ovat kuitenkin merkittäviä ja niiden merkitys korostuu, mitä enemmän aineistoja on saatavilla digitaalisessa muodossa. Yhtenä esimerkkinä voidaan mainita eri aineistojen yhdisteltävyys tutkimuksen tarpeisiin tai mahdollisuus hyödyntää digitaalisessa muodossa aiempaa pidempiä aikasarjoja.



### 6.3 Kustannus- ja hyötyanalyysi

Kustannus- ja hyötyanalyysin lähtökohtana on verrata keskenään kahta säilyttämisen skenaarioita. *Analogisen säilyttämisen skenaariossa* jatketaan arkistoinnin nykytilaa, jossa viranomaiset siirtävät vähitellen Kansallisarkistoon 40 vuotta tai sitä vanhempia analogisia asiakirjoja. Tässä skenaariossa Kansallisarkiston tulee käynnistää lähivuosina seuraavan keskusarkiston suunnittelu nykyisten arkistotilojen täyttymisen vuoksi. Analogisen säilyttämisen skenaariossa viranomaisilla säilyy vielä useita kymmeniä vuosia paperiarkistoja ja niiden hallinnointiin sitoutuu sekä tilakustannuksia että henkilötyökustannuksia.

*Digitaalisen säilyttämisen skenaariossa* viranomaisten siirtymistä täysin digitaaliseen tiedonhallintaan nopeutetaan merkittävästi toteuttamalla analogisten aineiston massadigitointi vuoteen 2030 mennessä. Näin viranomaiset vapautuvat paperiarkistojen hoidosta huomattavasti nopeutetussa aikataulussa. Lisäksi operatiivisessa toiminnassa tarvittavat analogiset asiakirjat saadaan digitoinnin myötä samojen tiedonhallinnan prosessien ja menettelyjen piiriin kuin digitaaliset aineistot. Digitaalisesti toimivassa hallinnossa ei ole tarkoituksenmukaista käsitellä ja hallita laajoja paperiarkistoja. Kahden päällekkäisen prosessin toteuttaminen on toiminnallisesti tehotonta ja se aiheuttaa tarpeettomia kustannuksia.

Kustannus- ja hyötyanalyysin pohjana on arvio valtionhallinnon paperiarkistoinnin nykytilan kustannuksista sekä arvio kustannusten kehittymisestä vuoteen 2040 asti. Kustannus- ja hyötyanalyysissa verrataan mitä kustannussäästöjä digitaalisen säilyttämisen skenaariolla saavutetaan verrattuna analogisen säilyttämisen skenaarioon:

- **Valtionhallinnon organisaatioiden kustannussäästöt arkistotiloissa.** Digitoinnin myötä valtionhallinnon organisaatioiden arkistotilat tyhjenevät asteittain. Viranomaisten arkistotilojen kustannukset ilman Kansallisarkiston, Museoviraston sekä eräiden turvallisuusviranomaisten tiloja ovat noin 7 miljoonaa euroa vuodessa. Senaattikiinteistöjen arvion mukaan arkistotilojen tilakustannuksista on mahdollista säästää noin 70 % sen jälkeen kun digitointi on toteutettu ja tilat ovat vapautuneet.



Säästöjen realisoituminen edellyttää, että digitointi koordinoidaan siten, että arkistotiloja voidaan tyhjentää yhtenäisinä tilakokonaisuuksina.

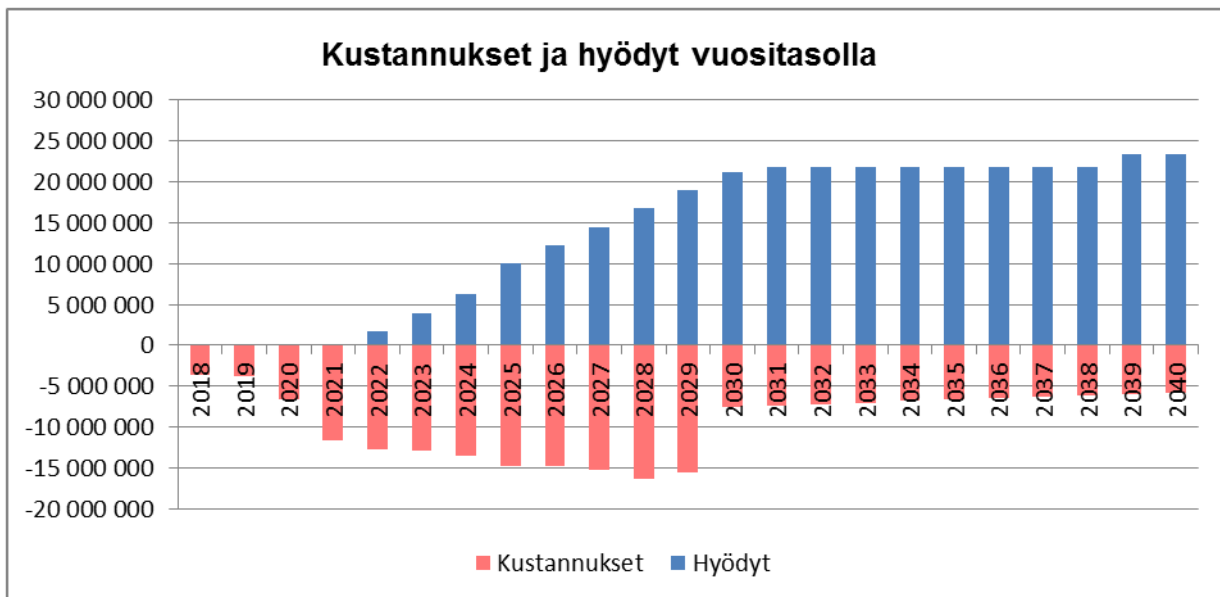
- **Asiakirjahallinnon ja arkistotoimen tehtäviin käytettävien henkilötyövuosien väheneminen valtionhallinnon organisaatioissa.** Digitoinnin myötä aineiston hallintaan ja käyttöön liittyviä työtehtävät muuttuvat valtionhallinnon virastoissa. Kaikki manuaalisen arkistonhoitoon liittyvät työtehtävät poistuvat (arkistointi, noudot arkistosta ja arkiston järjestäminen). Merkittäviä säästöjä saadaan aineiston käyttöön liittyvistä tehtävistä (tietopalvelu) ja tiedon etsimiseen ja löytämiseen käytettävästä työstä. Valtionhallinnossa asiakirjahallinnon ja arkistotoimen tehtäviin liittyvät henkilöstökustannukset ovat vuosittain noin 50 miljoonaa euroa. Paperiarkistoinnista ja siihen liittyvistä prosesseista luopumalla voidaan säästää asiakirjahallinnon ja arkistotoimen henkilötyövuosista arviolta 30 %. Säästöjen saavuttaminen edellyttää, että viranomaiset muuttavat omia toimintatapojaan ja prosessejaan digitaaliseen toimintaympäristöön sopiviksi. Säästöt tulevat olemaan suurimpia niiden organisaatioiden kohdalla, jotka huolehtivat prosessien uudistamisesta ja pystyvät kehittämään täysin uusia toimintatapoja.
- **Vältytään Kansallisarkiston tarvitseman arkistotilan lisärakentamiselta.** Jos arkistointia jatketaan paperimuodossa, tulee Kansallisarkiston käynnistää uudelleen paperiasiakirjojen vastaanotto. Paperiaineistojen jatkamisen skenaariossa Kansallisarkiston tulisi ottaa käyttöön seuraava keskusarkisto vuonna 2025 ja järjestyksessä kolmas keskusarkisto vuonna 2039. Uusien rakennusvaiheiden yhteenlasketut kustannukset nykyisen hintatason mukaan olisivat 32,8 miljoonaa euroa.

Digitaalisen säilyttämisen skenaarion kustannukset muodostuvat massadigitoinnin toteuttamisesta seuraavan 11 vuoden aikana sekä digitaalisen säilyttämisen kustannuksista, jotka ovat jatkuvia ja kasvavat digitaalisen aineistomäärän kasvun mukaisesti. Digitaalisen

säilyttämisen kustannus perustuu aineiston tallekappaleiden säilyttämiselle Kansallisen digitaalisen kirjaston pitkäaikaissäilytyspalvelussa (KDK-PAS).<sup>5</sup>

Massadigitoinnin päätyttyä vuonna 2030 analogisen säilyttämisen skenaarion säästö verrattuna digitaalisen säilyttämisen skenaarioon on arviolta 14 miljoonaa euroa vuosittain. Kustannussäästön suuruuteen vaikuttaa toisaalta se, miten valtionhallinnon organisaatioiden tila- ja henkilöstösäästöt toteutuvat ja toisaalta digitaalisen aineiston säilytyskustannukset, jotka ovat riippuvaisia sekä säilytettävästä datamäärästä että tallennustilan yksikköhinnasta.

Taulukko 7. Massadigitoinnin kustannukset ja hyödyt vuositasolla



<sup>5</sup> Tallennuskustannusten arvio vuoteen 2030 asti perustuu CSC Oy:n laatimaan budjetäriseen arvioon, jonka avulla toiminnan kokonaiskustannuksia voidaan hahmotella. Tallennusteknologian mahdollista halpenemista on huomioitu siten, että hankintakustannukset alenesivat 2019–2029 noin 10 prosenttia. Kustannusten arviointi vaikeutuu merkittävästi vuosista 2021–2022 eteenpäin, koska tallennusteknologioissa on parhaillaan käynnissä teknologinen murros.

## 7 Massadigitoinnin riskit

Seuraavassa on esitetty ne riskit, joita on tunnistettu liittyvän tässä raportissa esitettyyn massadigitoinnin toteuttamisen toimintamalliin. Projektissa ei ole lähdetty arvioimaan tarkemmin eri toteutusvaihtoehtojen riskejä, vaan arviossa on pitäydytty yleisemmällä tasolla.

Taulukko 8. Massadigitoinnin tunnistetut riskit

#	Kuvaus riskistä	Vaikutukset
1	<b>Johtaminen ja yhteistyö:</b> Yhteistyö eri tahojen (Kansallisarkisto, opetus- ja kulttuuriministeriö, valtiovarainministeriö ja viranomaiset) kesken ei toimi. Ministeriöiden ohjausroolit eivät ole selkeitä. Toimijat eivät sitoudu toimintamalliin riittävästi.	Massadigitoinnin tavoitteiden muuttuminen kesken, massadigitoinnin jatkuvuuden epävarmuus ja mahdollinen rahoituksen katkeaminen.
2	<b>Toimintamallin hallinnointi epäonnistuu:</b> Uuden toiminnon ja toimintamallin johtaminen ja hallinnointi ei toimi. Ongelmat Kansallisarkistolla toimintamallin ohjaamisessa ja jalkauttamisessa viranomaisille	Ongelmat toimintamallissa, asetettuja tavoitteita ei saavuteta. Epäselvyydet viranomaisissa toimintamallista.
3	<b>Liittyvien järjestelmien kyvykkyys:</b> Massadigitointiin liittyy useita eri tietojärjestelmiä (ks. tarkemmin luku 8.). Riskinä on, että näiden tietojärjestelmien toteuttaminen viivästyy tai järjestelmien kyvykkyys ei ole massadigitoinnin kannalta tarvittavalla tasolla.	Merkittäväillä poikkeamilla voi olla vaikutusta koko digitoinnin aloittamisen aikatauluun. Vähäisempien poikkeamien vaikutuksesta pitää toteuttaa korvaavia ratkaisuja, jotka heijastuvat prosessin sujuvuuteen.
4	<b>Digitoitavan aineiston tarkkuus:</b> Digitoitavan aineiston määrä osoittautuu arvioitua suuremmaksi tai digitoitavan aineiston laatu osoittautuu arvioitua haastavammaksi. Aineisto voi osoittautua myös arvioitua helpommaksi käsitellä.	Digitoitavan aineiston määrä kasvaa ja digitoinnin kustannukset kasvavat.  Vaikutukset digitoinnin aikatauluun
5	<b>Tuotannon käynnistämisen haasteet:</b> Tuotannon toteuttamiseen ei ole olemassa suoraan sovellettavia valmiita käytäntöjä. Ei ole olemassa toimijoita, jotka olisivat toteuttaneet mittakaavalta	Vaikutukset digitoinnin aikatauluun ja tuotannon kustannuksiin.





	ja laatuvaatimuksiltaan vastaavaa hanketta. Riskinä on, että tuotannon ylösajo toteutuu suunniteltua hitaammin.	
6	<b>Tarvittavan osaamisen haasteet:</b> Digitointia toteuttavalla taholla ei ole riittävää osaamista digitointiin.  Kansallisarkistolla ei ole riittäviä resursseja viranomaisten ohjaamiseen ja säilytyspäästösten laadintaan.	Vaikutukset digitoinnin tuotannon nopeuteen ja pahimmassa tapauksessa digitoinnin laatuun.  Vaikutukset näkyvät viranomaisten valmistelutoimien puutteissa tai hitautena säilytyspäästöprosessissa.
7	<b>Viranomaisvalmistelun ongelmat:</b> Viranomaisilla ei ole resursseja tehdä aineiston valmistelevia tehtäviä tiekartan mukaisessa aikataulussa.	Vaikuttaa digitoitavan aineiston saatavuuteen. Aineistoa ei tule digitointiin siinä tahdissa kuin pitäisi. Merkittävät vaikutukset koko digitoinnin prosessiin.
8	<b>Ongelmat prosessissa:</b> Prosessia ei saada kokonaisuudessaan toimimaan luotettavasti ja aukottomasti.	Ongelmat prosessissa johtavat prosessin muuttamiseen ja mahdollisesti uusien vaiheiden lisäämiseen ja esimerkiksi laadunvarmistuksen tehtävien kasvamiseen.

## 8 Massadigitoinnin riippuvuudet ja jatkotoimenpiteet

Massadigitoinnin tuotannon käynnistäminen edellyttää monia erilaisia valmiuksia. Osa näistä valmiuksista liittyy suoraan raportissa esitetyn toimintamallin, osa valmiuksista taas ei kuulu suoraan massadigitoinnin piiriin, vaan kyse on ulkoisista riippuvuuksista mm. toisista tietojärjestelmistä.

Toimintamalliin liittyviä valmiuksia sekä ulkoisia riippuvuuksia tarkastellaan tässä *arkkitehtuurikyvykkyyksien* näkökulmasta. Kyvykkyys tarkoittaa kykyä toimia tarkoituksenmukaisella tavalla tietyllä osa-alueella, ja hyödyntää osaamistaan sekä resurssejaan, jotta tavoitteet saavutetaan.



Organisaatioiden kyvykkyysien toteuttamiseen tarvitaan yleensä yhdistelmiä kolmesta osakokonaisuudesta, joita ovat *toimintamallit ja prosessit, henkilöstö ja osaaminen* sekä *tiedot ja järjestelmät*. Seuraavassa on esitetty *henkilöstön ja osaamisen* sekä *tietojen ja järjestelmien* näkökulmista se mitä kyvykkyksiä näiltä osa-alueilta edellytetään, jotta raportissa kuvattu toimintamalli ja prosessi voidaan toteuttaa. Massadigitoinnin ulkoisten riippuvuuksien eteneminen edellyttää systemaattista seurantaa ja selkeää riskimatriisia.

## 8.1 Henkilöstö ja osaaminen

Kansallisarkisto:

- Kyvykkyys ohjata viranomaisia suorittamaan digitoinnin viranomaisvalmistelun tehtäviä sekä laatimaan aineiston käsittelysuunnitelmia.
- Kyvykkyys ylläpitää digitoinnin tiekarttaa yhteistyössä viranomaisten kanssa. Tiekartalla määritetään missä järjestyksessä ja millä aikataululla viranomaisten aineistot digitoidaan.
- Kyvykkyys laatia tiukalla aikataululla ja laajoissa määrin säilytyspäätöksiä digitoitavista aineistoista.
- Kansallisarkistolla on ennestään kokemusta viranomaisten toiminnan ohjauksesta, siirtosuunnittelusta sekä säilytyspäätöksistä. Keskeiset haasteet kyvykkyysien kehittämiseen tulevat vaatimuksesta muokata prosesseja vastaamaan uusia tarpeita sekä huomattavasti kasvavista aineistomääristä, joiden ohjaus tulee edellyttämään uudenlaisia menetelmiä.

Digitoiva taho:

- Kyvykkyys toteuttaa digitointiprosessi Kansallisarkiston määritysten mukaan. Erityisesti digitoivan tahon tulee pystyä tuottamaan digitointiprosessissa Kansallisarkiston määritysten mukainen digitaalinen ilmentymä sekä varmistumaan määritetyillä menetelmillä digitoinnin laadusta.



- Kyvykkyys toteuttaa massadigitointi tiekartan mukaisin aineistomäärin ja tiekartan mukaisessa laajuudessa eli toisin sanoen kyvykkyys hyvin laajamittaisen digitoinnin toteuttamiseen.
- Kyvykkyys ohjata massadigitoinnin tuotantoa ja aineistojen logistisia siirtoja viranomaisilta digitointitiloihin.
- Digitoivalle taholle osoitetut vaatimukset koskien toisaalta digitoinnin laatuvaatimuksia ja toisaalta aineistomääriä ovat kansainvälisestikin katsoen ainutlaatuisia, joten millään toimijalla ei ole suoraan valmiuksia massadigitoinnin edellyttämiin vaatimuksiin.

#### Viranomaiset:

- Kyvykkyys valmistella aineisto uusien siirto-ohjeiden mukaan odottamaan siirtoa digitointiin. Viranomaisten tulee tehdä valmistelevat tehtävät tiekarttasuunnitelman edellyttämässä aikataulussa. Aineiston valmistelu tulee kyetä toteuttamaan uusien lähtökohtien mukaan, koska tavoitteena valmistelussa ei ole enää alkuperäisen analogisen aineiston säilyvyyden turvaaminen, vaan siirron ja digitoinnin mahdollistaminen.
- Kyvykkyys hyödyntää digitaalista aineistoa osana prosesseja ja muuttaa omia toimintatapoja digitaaliseen tiedonhallintaan.
- Viranomaisten kannalta massadigitoinnin toteuttaminen on merkittävä muutos verrattuna nykytilaan. Siirtyminen nopealla aikataululla paperiarkistoinnista täysin digitaaliseen tiedonhallintoon voi olla joissakin organisaatioissa valtava muutos, johon tulee varautua etukäteen. Monessa viranomaisessa perinteisen paperiarkistoinnin tehtävien korvautuminen digitaalisen tiedonhallinnan tehtäville voi edellyttää osaamisen päivittämistä.



## 8.2 Tietojärjestelmät

Seuraavassa on tunnistettu massadigitoinnin toteuttamisen kannalta keskeiset tietojärjestelmät. Tietojärjestelmien toteuttamiin palveluihin kohdentuu erilaisia kyvykkyyksivaatimuksia, joita ei ole tässä lähdetty tarkemmin avaamaan. Lähtökohtana on osoittaa riippuvuudet keskeisiin tietojärjestelmiin.

- *Tuotannonohjausjärjestelmä* on massadigitoinnin hallinnointijärjestelmä, jonka rajapintoja sekä tietoja käytetään massadigitoinnin kokonaisprosessin ohjaamiseen sekä aineistolle tehtyjen toimenpiteiden dokumentointiin. Järjestelmän tulee olla tiiviisti integroitu osaksi Kansallisarkiston järjestelmiä. Järjestelmän toteutus on aloittamatta.
- *Arkistoaineiston metatietojärjestelmä (AHAA)* on tietovaranto, jossa sijaitsee arkistoaineiston kuvailu- ja luettelointitietojen ydintieto. Käytetään massadigitoinnissa viranomaisten aineistojen luettelointiin ja aineiston tietojen ydintietovarantona, mistä tiedot välitetään tuotannonohjausjärjestelmään. Järjestelmä otetaan käyttöön Kansallisarkistossa vuoden 2018 aikana. Massadigitoinnin kyvykkyys voi edellyttää järjestelmän rajapintojen kehittämistä.
- *Aineistohallintajärjestelmä (AHJ)* on tietovaranto, jossa sijaitsee analogisen aineiston sijainnin ydintieto. Käytetään tuotannonohjausjärjestelmän taustalla ja erityisesti digitoinnin jälkeen analogisessa muodossa säilytettävän aineiston sijoittamisessa. Järjestelmä otetaan käyttöön Kansallisarkistossa vuoden 2018 aikana.
- *Vastaanottava järjestelmä* vastaanottaa digitoinnissa syntyneen siirtopaketin, rikastaa ja jakelee sitä sekä käyttöön että säilyttämiseen. Vastaanottavan järjestelmän tehtävät on suunniteltu hoitavan SAPA-palvelukokonaisuus. Järjestelmän toteutus on aloittamatta.
- *Pitkäaikaissäilytysjärjestelmä* vastaa digitoinnissa tuotettavien tallekappaleiden pitkäaikaisesta säilymisestä. Järjestelmä on olemassa KDK-PAS-palveluna. Massadigitoinnin edellyttämä kyvykkyys voi edellyttää kehittämistoimia järjestelmälle.



- *Sisällön analysointi* hyödyntää OCR-menetelmin kuvista tunnistettua tekstiä ja tuottaa tekstistä analysoinnin keinoin automaattisesti aineiston käyttöä helpottavia arkistokuvailun metatietoja. Järjestelmän toteutus on aloittamatta.

### 8.3 Ehdotetut jatkotoimet

Massadigitoinnin valmiuksien ja edellytysten parantamiseksi projekti pitää vähintään seuraavien jatkotoimenpiteiden toteuttamista välttämättömänä. Toimenpiteiden lykkäytymisellä on vaikutuksia massadigitoinnin tuotannon käynnistämisen aikatauluun.

- Kaikkien toimijoiden sitoutuminen massadigitoinnin tavoitteisiin: viranomaiset eivät käynnistä heiltä edellytettäviä valmistelevia toimia ellei massadigitoinnin toteuttamisesta ole selkeitä ja sitovia päätöksiä. Tämä edellyttää toteutusta ohjaavilta ministeriöiltä selkeitä kannanottoja, sitoutumista sekä päätöksiä rahoituksesta.
- Viranomaisten tulee sitoutua aineistojen siirroissa tiekartan aikatauluihin: Kansallisarkiston tulee täsmentää yhteistyössä viranomaisten kanssa tiekarttasuunnitelma. Viranomaiset eivät kuitenkaan voi sitoutua tiekarttaan ellei viesti massadigitoinnin käynnistämisestä ole riittävän vahva ohjaavilta ministeriöiltä.
- Viranomaisten siirtovalmiuksien kehittäminen: massadigitoinnin kokonaisprosessin kriittinen piste on aineistojen siirto viranomaisilta digitointipaikalle. Viranomaiset tarvitsevat selkeät ohjeet, mitä toimia aineistoille on tehtävä ja miten aineistoja voidaan valmistella digitointia varten. Kansallisarkiston tulee laatia viranomaisvalmistelun ohjeistus.
- Toimiva sähköisten aineistojen arkistoinnin SAPA-palvelukokonaisuus, joka mahdollistaa digitaalisen muotoon muutetun aineiston käytön. SAPA-palvelukokonaisuus tarvitaan, jotta digitaaliseen muotoon muutetut asiakirjat ovat viranomaisen käytettävissä osana tehtävien hoitoa ja jotta aineistoa voidaan tarjota helppokäyttöisin liittymin myös muille



tahoille käyttöön. Analogista aineistoa ei voida digitoinnin jälkeen hävittää, jos sitä ei saada yhteisesti ja luotettavasti käyttöön.

- *Proof of conceptin* toteuttaminen suurtehoskannerista sekä tekstintunnistuksen OCR-menetelmistä. Kansallisarkiston tulee testata valittuja teknologioita, jotta valmistelun, digitoinnin ja skannaukset toimenpiteistä on tarkempi käsitys.
- Tuotannonohjausjärjestelmän määrittelyn ja toteutuksen käynnistäminen sekä rajapintojen toteuttaminen Kansallisarkiston AHAA- ja AHJ-järjestelmiin. Massadigitointia ei voida toteuttaa luotettavasti ilman käyttötarkoitukseen soveltuvaa tuotannonohjausjärjestelmää.