

# DIGIKSI

Digitointi-  
projektin  
suunnittelijan  
ja toteuttajan  
opas

Emmi Liikanen ja  
Henna Ristolainen (toim.)



Opas on tuotettu Digitoinnin taitajaksi -hankkeessa. Hankkeen ovat toteuttaneet Kaakkois-Suomen ammattikorkeakoulu ja Mikkelin kehitysyhtiö Miksei Oy 1.1.2022-31.12.2022. Hanketta rahoitti Etelä-Savon ELY-keskus Euroopan sosiaalirahastosta. Hanke on osa Digitaliaa - digitaalisen tiedonhallinnan tutkimus- ja kehittämiskeskusta - ja Memory Campusta.

Hankkeen järjestämissä koulutuksissa tuettiin digikyvykkyyttä sekä lisättiin ymmärrystä digitointi-prosesseista ja digitointiprojektien läpiviennistä. Hankkeessa kerättiin tietoa digitointiprojekteista ja tietoa käytettiin sekä koulutussisällöissä että tässä oppaassa.



## **XAMK KEHITTÄÄ 207 KAAKKOIS-SUOMEN AMMATTIKORKEAKOULU MIKKELI 2022**

© Tekijät ja Kaakkois-Suomen ammattikorkeakoulu

Kuvitus: Kuvat Varkauden museoista, Mikkelin kaupungilta ja Kansalliskirjastolta tai erikseen mainittu.

Kannen kuva: Varkauden museoiden Tuomas Toratti digitoimassa.

Taitto: Mainostoimisto Grotteski Oy

ISBN: 978-952-344-501-7 (PDF)

ISSN: 2489-3102 (verkkójulkaisu)

julkaisut@xamk.fi



# Kirjoittajat

Sari Järn, TKI-asiantuntija, Kaakkois-Suomen ammattikorkeakoulu

Anssi Jääskeläinen, tutkimuspäällikkö, Kaakkois-Suomen ammattikorkeakoulu

Ville Kajanne, ylitarkastaja, Kansallisarkisto

Minna Kaukonen, suunnittelupäällikkö, Kansalliskirjasto

Emmi Liikanen, projektipäällikkö, Mikkelin kehitysyritys Miksei Oy

Kari Peiponen, tietohallintopäällikkö, Kansallisgalleria

Maiju Pohjola, suunnittelija, Kansallisarkisto

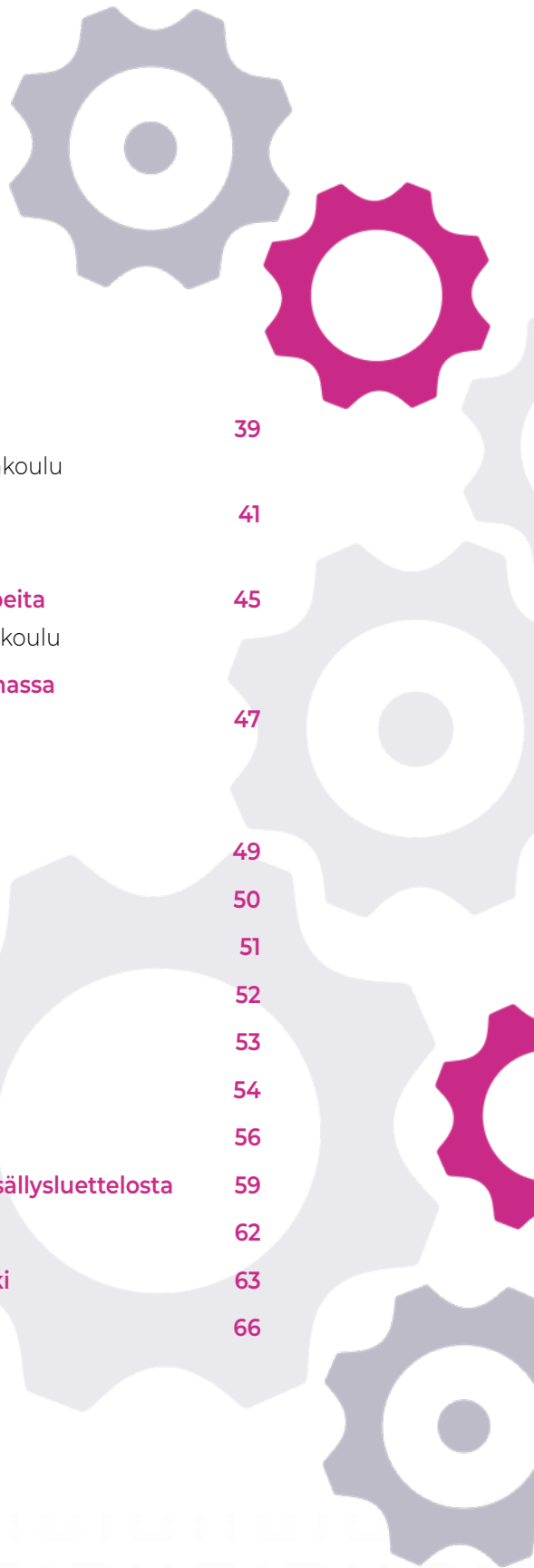
Henna Ristolainen, projektipäällikkö, Kaakkois-Suomen ammattikorkeakoulu

Kati Saltiola, ohjelmakoordinaattori, Kaakkois-Suomen ammattikorkeakoulu

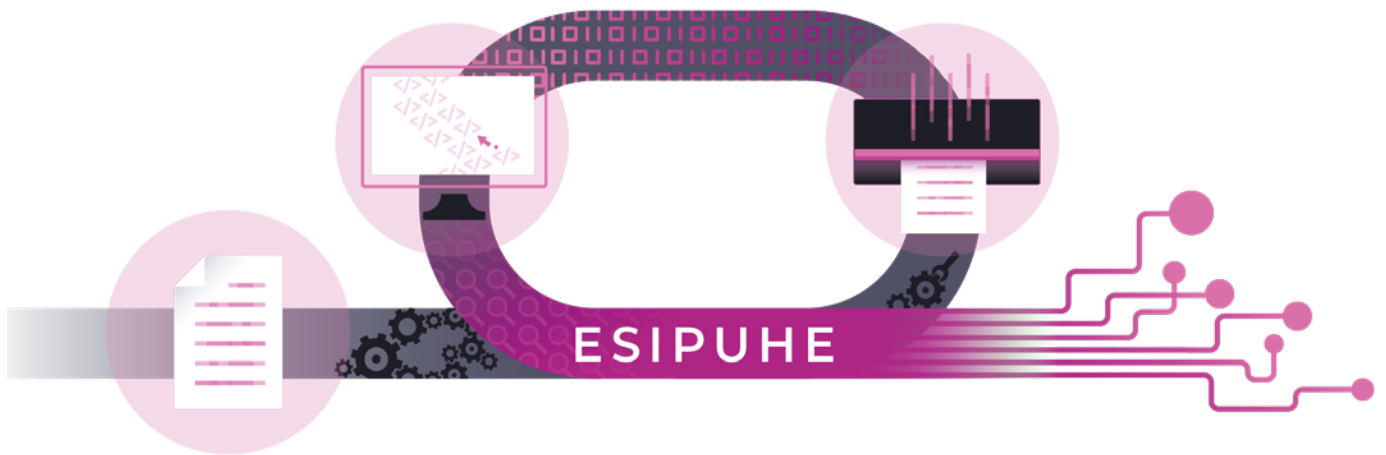
Stina Westman, TKI-yksikön johtaja, Kaakkois-Suomen ammattikorkeakoulu

# Sisällys

<b>Kirjoittajat</b>	<b>3</b>
<b>Kohti onnistunutta digitointiprojektia</b>	<b>6</b>
<b>OSA 1 – OPAS</b>	<b>7</b>
<b>Digitoinnin hyödyt yhteiskunnalle, organisaatiolle ja yksilölle</b> Sari Järn, Kaakkois-Suomen ammattikorkeakoulu	<b>8</b>
<b>Digitointiprosessin vaiheet</b> Sari Järn, Kaakkois-Suomen ammattikorkeakoulu	<b>11</b>
<b>Alkuun digitointiprojektissa</b> Sari Järn, Kaakkois-Suomen ammattikorkeakoulu	<b>15</b>
<b>Digitointiprojektin hallinta</b> Sari Järn ja Henna Ristolainen, Kaakkois-Suomen ammattikorkeakoulu Emmi Liikanen, Mikkelin kehitysyritys Miksei Oy	<b>18</b>
<b>Digitoinnin laatu</b> Sari Järn ja Henna Ristolainen, Kaakkois-Suomen ammattikorkeakoulu	<b>28</b>
<b>OSA 2 – NÄKÖKULMIA</b>	<b>32</b>
<b>Digitointi muuttuu ja kehittyy</b> Henna Ristolainen, Kaakkois-Suomen ammattikorkeakoulu	<b>33</b>
<b>Kansalliskirjaston digitointiohjelma digitoinnin lähtökohtana</b> Minna Kaukonen, Kansalliskirjasto	<b>35</b>
<b>Tekoälyvälineiden käyttöönoton tavoitteet ja sudenkuopat massadigitoinnissa</b> Ville Kajanne ja Maiju Pohjola, Kansallisarkisto	<b>37</b>



<b>Digitoinnissa keinoäly tunnistaa tyhjät sivut</b>	39
Anssi Jääskeläinen, Kaakkois-Suomen ammattikorkeakoulu	
<b>Kokemuksia tiedostojen PAS-paketoinnista</b>	41
Kari Peiponen, Kansallisgalleria	
<b>Digitointikoulutus kehittyä palvelemaan digitointitarpeita</b>	45
Henna Ristolainen, Kaakkois-Suomen ammattikorkeakoulu	
<b>Digitointiprojektien haasteet – case-esimerkit kuvaamassa digitointityön monimuotoisuutta</b>	47
Emmi Liikanen, Mikkelin kehitysyritys Miksei Oy	
<b>LIITTEET</b>	49
<b>Case: Lahden kaupunki</b>	50
<b>Case: Miehikkälän museot</b>	51
<b>Case: Mikkelin kaupunki</b>	52
<b>Case: Päivälehdien arkisto</b>	53
<b>Case: Turun kaupunki</b>	54
<b>Digitointiprosessin vaiheita</b>	56
<b>Esimerkki digitointiprojektin projektisuunnitelman sisällysluettelosta</b>	59
<b>Hankintasuunnitelma</b>	62
<b>Esimerkki hankintasuunnitelmasta: Mikkelin kaupunki</b>	63
<b>Käsitteet ja termit</b>	66



# Kohti onnistunutta digitointiprojektia

STINA WESTMAN

Digitointiprojektien tavoitteet ovat usein hyvin käytännönläheisiä. Kalliista fyysisestä säilytys-tilasta halutaan luopua, aineistojen säilyvyyttä halutaan parantaa tai arkistot halutaan avata käytettäväksi verkon yli. Digitointiin liittyy usein myös laajempaa tiedonhallinnan prosessien ja käytäntöjen uudistamista organisaatioissa. Joskus on jo syntynyt uusi digitaalinen työnkulku, johon vanhat analogiset aineistot halutaan tuoda mukaan löydettäväksi ja käytettäväksi. Näiden lähtökohtien ymmärtäminen auttaa määrittelemään, mitä digitoinnilla itse asiassa tavoitellaan.

Digitoinnista puhuttaessa keskustelu keskittyy usein tekniikkaan, esimerkiksi tiedostomuotoihin, laitteistojen ominaisuuksiin tai digitoidun aineiston laatuun. Digitointi ei kuitenkaan ole samaa kuin skannaaminen, eikä sitä kannata ajatella ainoastaan teknisenä toimenpiteenä. Aineiston muuntaminen digitaaliseen muotoon tapahtuu ihmisen ja koneen vuorovaikutuksessa. Teknologian käyttömahdollisuudet kehittyvät, ja toisaalta ihmisen osaaminen ja valinnat vaikuttavat siihen, miten teknologiaa hyödynnetään kussakin projektissa. Digitointia suunnitteleva paitsi suunnittelee digitoinnissa tarvittavan teknisen ympäristön myös saattaa sopeuttaa tavoitteitaan teknologian tarjoamien mahdollisuuksien mukaan. Digitoinnissa syntyvän digitaalisen aineiston tulisi paitsi säilyttää tietosisältö myös tarjota halutunlainen ilmentymä alkuperäisestä aineistosta. Digitoinnin suunnittelussa on siis päätettävä, miten vaurioituneita alkuperäisiä käsitellään

ja miten päätellään, mikä on aineiston alkuperäinen ilmiä. Harvoin on mahdollista digitoida kaikkea aineistoa, jolloin on päätettävä, mikä on tärkeintä digitoida ja miten.

Digitoinnin valmistelu ja aineiston kuvailu vievät aikaa. On keskeistä, että digitointiin ryhtyvä taho ymmärtää, mitä digitoinnilla tavoitellaan ja keitä se palvelee, nyt ja tulevaisuudessa. Osana digitointiprojektin suunnittelua tehtävät valinnat ja priorisoinnit ovatkin merkityksellisiä sekä digitoinnin toteutuksen että lopputuloksen jatkoehdottomuuden kannalta. Voi jopa väittää, että digitointia on mahdotonta toteuttaa täydellisesti, sillä digitoinnin määritykset eivät koskaan ole lopullisia. Tekniikan kehittyessä pystytään kuitenkin vastaamaan uusiin käyttötarpeisiin.

Tämä opas on suunnattu kaikille digitointiprojektia suunnitteleville ja valmisteleville. Opas käsittelee erityisesti julkisen sektorin digitointia, mutta soveltuu myös yksityisen ja kolmannen sektorin käyttöön. Asiantuntijatekstit tuovat esille digitointiprosessin osa-alueita, käytännön esimerkkejä erilaisista digitointiprojekteista ja apuvälineitä laadukkaaseen digitointiin. Toivon, että opas avaa lukijoilleen moninaisia näkökulmia digitointiin ja tukee oman digitointiprojektin hallittua ja tavoitteellista läpivientiä.





# Osa 1

—

# Opas



# Digitoinnin hyödyt yhteiskunnalle, organisaatiolle ja yksilölle

SARI JÄRN

*Digitointi voi muuttaa organisaation tapaa toimia, mutta digitointi ja digitalisaatio on erotettava käsitteinä toisistaan. Niitä yhdistää se, ettei kumpikaan ole pelkkää tietotekniikkaa. Digitointi ei ole yksittäinen tekninen suoritus. Digitalisaation voi puolestaan jakaa erilaisiin periaatteisiin, joiden tavoitteena on muun muassa tukea julkisten palveluiden ensisijaista digitalisuutta.*

Tiedon hallinta helpottuu digitoinnin myötä, mikä onkin digitalisaation ydintä. Digitalisaatio perustuu tiedon jakamiseen. Julkisia palveluja kehitettäessä on pitänyt sopia yhteisiä digitalisoinnin periaatteita.

Digitointiin liittyviä ohjeita, määräyksiä sekä digitoinnin periaatteita on ollut luomassa moni sivistyshallinnon alla toimiva organisaatio. Näille organisaatioille on syntynyt digitoinnin kehittämisen myötä erityisosaamista. Kansallisarkistolla se liittyy asiakirjojen digitointiin ja massadigitoinnin toimintoihin. Kansalliskirjasto on keskittynyt sanomalehtiin ja muihin vapaakappaleaineistoihin. Kansallinen audiovisuaalinen instituutti (KAVI) hallitsee elokuvien digitoinnin, ja Museo- virasto on panostanut valokuvien sekä esineiden ja muiden kolmiulotteisten kohteiden digitointiin sekä mallintamiseen.

Digitoinnin ja digitalisaation rinnalle nousee digitaalinen transformaatio. Kun jokin analoginen aineisto digitoidaan, sen jälkeen voidaan digitalisaation avulla luoda digitaalisesta aineistosta

uusia palveluja. Jos luodusta palvelusta kehitetään tämän jälkeen palvelu, joka muuttaa tapaamme toimia, puhutaan digitaalisesta transformaatiosta.

Esimerkiksi, kun kirja digitoidaan, sen voi ostaa sähköisenä. Sähköisten kirjojen, varsinkin äänikirjojen, palveluja tarjoaa moni toimija. Yksittäisen kirjan sijaan voikin hankkia kokonaisen palvelupaketin, joka sisältää monia kirjoja.

## Digitoinnin edistäminen on yhteinen asia

Digitointi parantaa aineistojen saavutettavuutta ja käytettävyyttä. Aineistot ovat saatavilla ajasta ja paikasta riippumatta, tiedonhaku helpottuu sekä mahdollisuudet itsepalvelun tarjoamiseen paranevat. Tulevaisuudessa asiakirjat eivät vie tilaa fyysisissä arkistoissa.

Digitoinnin hyödyt yhteiskunnalle ovat merkittävät. Suomi on sitoutunut edistämään yhteistä eurooppalaista kulttuuriperinnön data-avaruutta noudattamalla digitaalisen sisällön ja metadatan tuottamisessa sekä jakamisessa yhteisiä stan-

dardeja ja toimintamalleja sekä digitoimalla kulttuuriperintöä. Tämä lupaus sisältyy Suomen digitaaliseen kompassiin.

Kulttuuriperintöämme suojellaan sekä kansallisella lainsäädännöllä että kansainvälisillä sopimuksilla. Perintö voi olla uhattuna hyvin erityyppisissä tilanteissa. Suomi on ratifioinut keskeiset kansainväliset kulttuuriperinnön suojelua koskevat sopimukset, ja kulttuuriomaisuuden suojelu on tunnistettu myös osana henkisen kriisinkeskeytyksen ylläpitämistä.

Digitointi on osa tätä kokonaisuutta. Kulttuuriperintöaineistoja digitalisoimalla voidaan parantaa niiden saavutettavuutta sekä suojata hauraita tai paljon käytettäviä aineistoja kulumiselta. Opetus- ja kulttuuriministeriö tukee digitaalisten kulttuuriperintöaineistojen saatavuutta, käyttöä ja säilyttämistä monin eri tavoin. Kansalliskirjaston ylläpitämässä Finna-palvelussa on 2,5 miljoonaa digitaalista objektiota sekä erilaisia palveluja. Tieteen tietotekniikan keskuksen CSC:n ylläpitämä Kulttuuri-perintö-PAS-palvelu varmistaa digitaalisten kulttuuriperintö-



aineistojen säilyvyyden pitkälle tulevaisuuteen.

Myös kulttuuriperintölaitokset digitoivat kokoelmiaan ja tarjoavat asiakkailleen digitaalisia palveluja.

Kulttuuriperintöaineistojen lisäksi digitoinnin hyödyt näkyvät kaikkialla yhteiskunnassa, kun aineistot ovat saatavilla ajasta ja paikasta riippumatta. Digitaalisatiiokehitys vaatii kuitenkin laajaa yhteistyötä valtionhallinnon, kuntasektorin, yritysten sekä myös yksityisten toimijoiden kesken.

Kaikkea ei kuitenkaan ole mahdollista eikä järkevää digitoida, eikä kaikilla organisaatioilla ole resursseja tai osaamista digitointiin ja sähköiseen arkistohallintaan. Digitointiin liittyvä osaamista on lisättävä, jotta siihen ohjatut resurssit käytetään tehokkaasti.

Yritysten näkökulmasta digitointi lisää asioiden ja ilmiöiden välistä yhdisteltävyyttä sekä luo edellytyksiä uudelle liike-toiminnalle ja liiketoimintamalleille. Digitoinnin avulla voidaan uudistaa toimintaprosesseja. Tietoturva on toteutettavissa paremmin kuin analogisessa aineistossa. Tietopalvelu ja aineiston jakelu tehostuu, kun sisällöt on kuvailtu ja indeksoitu. Osalle organisaatioista analoginen säilyttäminen on yhä paras tapa ja digitointia tehdään vain tarvittaessa.

”

*”Digitointi helpottaa asiointia, kun asiakas voi saada tarvitsemansa asiakirjat käyttöönsä ajasta ja paikasta riippumatta, eikä tarvitse jonotella virastoihin. Asiakkaat odottavat usein saavansa aineistot käyttöönsä digitaalisessa muodossa.”*

*”Aineiston metatiedot ja digitoinnissa käytetty sisällöntunnistus mahdollistavat helposti erilaisia hakutoimintoja, joita ei voi fyysisille dokumenteille tehdä. Digitoimista aineistosta voi saada enemmän irti.”*

”

*”Erityisesti tutkijat ja sukututkijat pääsevät samalla käsiksi aineistoihin, joita eivät tieneet olevan olemassakaan, ja saavat tietoa, josta hyötyvät.”*

*”Digitointi mahdollistaa aineistojen laajan näkyvyyden. Kuka tahansa voi tutustua aineistoihin verkkoyhteyden välityksellä kotoaan, ja aineiston käyttö on mahdollista useammalle asiakkaalle samaan aikaan. Tämä on ekologista, sillä fyysisen arkistotilan tarve vähenee ja matkustamisen vaiva säästyy.”*

”

*”Tietopalvelu tehostuu, kun aineisto on digitaalisessa muodossa. Asiakkaalle tämä näkyy aiempaa nopeampana palveluna. Digitointi vähentää aineistojen kulumista, kun alkuperäiseen aineistoon ei tarvitse kajota.”*

*”Julkishallinnon organisaatioilla on veltollisuus saattaa kokoelmia digitaaliseen muotoon ja tehdä se laadukkaasti ja turvallisesti sekä riskit ja lainsäädäntö huomioiden.”*

”

*”Digitointi kehittää avointa pääsyä aineistoihin ja mahdollistaa esimerkiksi niiden helpon jakamisen eri toimijoiden välillä. Digitointi lisää omalta osaltaan tiedon läpinäkyvyyttä ja edistää aineistojen säilymistä tuleville sukupolville.”*

# Esimerkkejä digitoinnin hyödyistä

## Toimintaympäristön muutos

- Digitaalisina tarjolla olevat aineistot muuttavat toimintaprosesseja ja vaikuttavat toimintatapoihin.
- Ne tehostavat toimintaa, lisäävät tietoturvaa, helpottavat aineistojen hallintaa ja säästävät arkistotilaa.

## Säilyvyyden hyödyt

- Digitoinnilla voidaan varmistaa dokumentin säilyminen myös muussa kuin fyysisessä muodossa.
- Digitointi vähentää analogisten aineistojen kulumista, kun alkuperäiseen aineistoon ei kohdistu tietopyyntöjä.

## Saavutettavuus ja tiedonhaun hyödyt

- Digitointi parantaa aineistojen saavutettavuutta ja käytettävyyttä.
- Digitaaliset aineistot ovat saatavilla ajasta, paikasta ja käyttäjämäärästä riippumatta.
- Aineistoista tehtävien tiedonhakujen teko helpottuu, jakelu nopeutuu sekä mahdollisuudet itsepalvelun tarjoamiseen paranevat.
- Digitoidut aineistot saatetaan laajaan käyttöön, ja käyttäjillä on entistä paremmat mahdollisuudet selata, hakea ja jakaa aineistoja.

## Taloudelliset hyödyt

- Taloudellisia hyötyjä saadaan, mikäli arkistotiloja vapautuu tai kun vältetään uusien arkistotilojen rakentamiselta.
- Digitoimalla luodaan uudenlaisia mahdollisuuksia kokoelmien hyödyntämiseen sekä ansaintalogiikoiden rakentamiseen.

## Ekologiset hyödyt

- Fyysisten arkistotilojen tarve vähenee samoin kuin matkustelu arkistotiloihin.

### Lähteet:

Hänninen, T. 2019. Digitointiopas. Kaakkois-Suomen ammattikorkeakoulu. WWW-dokumentti. Saatavissa: <https://digitalia.xamk.fi/digiopas/sanasto> [viitattu 1.11.2022].

Kansallisarkisto. 2022. Digitointi. WWW-dokumentti. Saatavissa: <https://kansallisarkisto.fi/fi/viranomaisille/digitointi> [viitattu 15.11.2022].

Kasvi, J.J.J. 2019. Digi digi digi. TIEKE Tietoyhteiskunnan kehittämiskeskus ry. WWW-dokumentti. Saatavissa: <https://tieke.fi/digi-digi-digi/> [viitattu 15.11.2022].

Kiron, D. & Unruh, G. 2017. Digital Transformation on Purpose. MIT Sloan Management Review. WWW-dokumentti. <https://sloanreview.mit.edu/article/digital-transformation-on-purpose/> [viitattu 16.11.2022].

Opetus- ja kulttuuriministeriö. 2022. Sivistyshallinto 2030. Opetus- ja kulttuuriministeriön konsernin kehittämishankkeen loppuraportti. Opetus- ja kulttuuriministeriön julkaisuja 2022:35. PDF-dokumentti. Saatavissa: <http://urn.fi/URN:ISBN:978-952-263-816-8> [viitattu 15.11.2022].

Pölkki, T. 2022. Digitaalisuus ilman muutosta on turhaa. Blogi. Päivitetty 11.3.2022. Saatavissa: <https://blogit.jamk.fi/turbinaattori/2022/03/11/digitaalisuus-ilman-muutosta-onturhaa/> [viitattu 16.11.2022].

Valtioneuvosto. 2022. Valtioneuvoston selonteko: Suomen digitaalinen kompassi. PDF-dokumentti. Saatavissa: <http://urn.fi/URN:ISBN:978-952-383-906-9> [viitattu 16.11.2022].

Valtiovarainministeriö. 2017. Digitalisoinnin periaatteet. Muistio 3.4.2017. PDF-dokumentti. Saatavissa: <https://vm.fi/documents/10623/1464506/Digitalisoinnin+periaatteet/63c2a2fa-b7b5-45e4-8516-bd804490fecf/Digitalisoinnin+periaatteet.pdf?t=1491289673000> [viitattu 15.11.2022].

# Digitointiprosessin vaiheet

SARI JÄRN

## Laatu varmistetaan suunnittelulla

Digitointiprosessi alkaa suunnittelusta. Se on tärkeää prosessin laadun ja onnistumisen kannalta. Suunnitteluvaiheessa perehdytään voimassa olevaan lainsäädäntöön, määräyksiin ja ohjeisiin, jotta prosessissa toimitaan lain ja hyvien toimintaperiaatteiden mukaisesti. Lainsäädännön lisäksi käydään läpi digitoinnin lähtökohdat, tavoitteet sekä aineiston elinkaaren hallinta.

Digitointiin liittyvät vaatimukset ovat käyttötarkoituksesta riippuvia. Aineisto, joka on tarkoitus siirtää pitkäaikaissäilytykseen, vaatii toisenlaisen prosessin kuin pelkkään sisäiseen käyttöön tai yksittäiseen tiedontarpeeseen tähtäävä digitointi. Aineiston valmisteluun ja kuvailuun liittyvät työvaiheet kirjoitetaan auki.

Prosessiin kuuluu myös tallentaminen, laadunvarmistus ja validointi. Digitointia suunniteltaessa on otettava alusta lähtien huomioon saavutettavuus ja käytettävyys sekä kohderyhmän tarpeet.

**Katso liite sivulta 56:  
Esimerkki digitointi-  
prosessin vaiheista.**



## Kaikkea ei tarvitse digitoida

Digitointia suunniteltaessa aineiston valinta on keskeisessä asemassa. Aineiston määrä voi vaihdella hyllykilometreistä laatikolliseen. Aineiston valintaa voi ohjata kokoelmapolitiikka, Kansallisarkiston laatimat massadigitoinnin tiekartat, tarve digitoida tietty aineisto tai asiakkaan yksilöllinen tarve.

Resurssit ovat usein rajalliset, joten on tarkkaan määriteltävä, mitä aineistoa digitoidaan. Lähestymistapana voi olla aineiston arvo kulttuuriperintö- ja tutkimustarkoituksiin, jolloin tavoitteena on turvata säilyvyys. Digitoitavaksi voidaan valita aineisto, johon kohdistuu paljon tietopyyntöjä.

Aineistoja voidaan säilyttää analogisina, ja digitoida vain pyynnöstä. Tällöin kyse ei ole niinkään rajatusta projektista vaan palvelumallin prosessista. Mallin tuottaminen vaatii kuitenkin pitkälti samojen kysymysten pohdintaa kuin rajattu digitointiprojekti.

Aineiston valinnassa kannattaa pohtia, onko aineisto tuhoutumassa ja onko se kysyttyä. Halutaanko aineistoa tuoda kokoelmasta esiin? Liittyykö aineistoon rajoittavia tekijöitä, kuten lainsäädäntöä?

Suunnitelmaan kirjoitetaan auki digitoitavan aineiston valintaperiaatteet. Nämä periaatteet

ohjaavat digitointia. Kun periaatteet on kirjattu, voidaan aloittaa digitointijärjestyksen suunnittelu ja priorisointi. Näin syntyy digitoinnin tiekartta: mitä digitoidaan ja missä järjestyksessä?

Suunnittelussa on syytä kartoittaa digitoiva aineisto, jotta aineistolle voidaan valita sille sopivat digitointimenetelmät. Kartoituksessa selviää, millaista käsittelyä aineisto vaatii. Suunnittelun aikana selvitetään, suoritetaanko digitointi itse vai hankitaanko työ ulkopuoliselta toimijalta.

## Digitoidaanko itse vai ulkoistetaanko?

Kun digitoitavat aineistot ja digitointimenetelmät alkavat hahmottua, on syytä pohtia seuraavia kysymyksiä: Kuka tekee? Ulkoistetaanko vai tehdäänkö itse? Missä digitointi tapahtuu? Ketkä ovat vastuussa prosesseista? Mitkä ovat prosessin mahdolliset kriittiset kohdat?

Prosessin testaukseen ja aikataulutukseen on syytä käyttää riittävästi aikaa, jotta digitointityö etenee suunnitellulla tavalla.

Suunnitelmaan tarvitaan myös tieto digitoinnin jälkeisestä säilytystarpeesta. On käytävä läpi tallennus- ja säilytysjärjestelmien vaatimukset. Suunnitelmasta on syytä käydä ilmi, missä järjestelmissä digitoitu aineisto liikkuu.

Metatiedot ovat avainasemassa aineiston käytettävyyden kan-

nalta. Digitoinnin lopputulokseen vaikuttaa siis metatietojen ja tiedostomuotojen määrittely. Digitoinnin yhteydessä kerättävän metatiedon merkitys tulee ymmärtää ja tunnistaa. Jos digitointi ulkoistetaan, on huolehdittava, että digitoidusta aineistosta saadaan tiedostojen lisäksi vaaditut metatiedot.

## Aineiston valmistelu digitointia varten

Prosessin suunnittelussa määriteltiin, mitä aineistoa digitoidaan. Tiekarttaan kirjoitettiin aineiston digitointijärjestys. Aineiston käsittely aloitetaan sisällön seulonnalla ja järjestelyllä. Jos kyse on asiakirjoista, tarkistetaan niiden säilytysajat. Määräaikaisesti säilytettäviä asiakirjoja ei välttämättä kannata digitoida, ellei niihin kohdistu paljon tietopyyntöjä.

Digitoinnin valmistelulla varmistetaan, että digitointi sujuu tehokkaasti analogisen aineiston sisältämää informaatiota vaarantamatta. Huolellisella valmistelulla pyritään varmistamaan, että digitointiprosessi etenee ilman turhia keskeytyksiä, digitointilaitteisto ei vahingoitu ja kaikki digitoitava informaatio pysyy tallessa. Puutteellisesti valmistettu aineisto pysäyttää prosessin ja saattaa vahingoittaa asiakirjoja tai skannereita – tai pahimmillaan aiheuttaa asiakirjan koko informaation puuttumisen.

Valmisteluprosessi on usein työvoimaintensiivinen ja aikaa vievä prosessi, jossa automaatiikkaa ei voida juurikaan hyödyntää. Manuaalisella valmistelulla voidaan kuitenkin minimoida riskit informaation häviämislle ennen digitointia. On myös tehtävä suunnitelma erittäin huonokuntoisen, huonokuntoisen tai erikoisaineiston erottelulle, koska ne vaativat erillisiä valmistelu- ja/tai digitointimenetelmiä.

Valmisteluprosessia ohjaa pari tärkeää peruseriaatetta: valmistelutoimien tulisi olla mahdollisimman vähäisiä ja selkeitä. Jos aineisto säilytetään vielä digitoinnin jälkeen myös analogisessa muodossa, on huolehdittava, että valmistelutoimenpiteet ja aineistoon mahdollisesti lisättävät materiaalit ovat arkistointikelpoisia.

Aineistosta otetaan pois sinne kuulumaton materiaali. Järjestämisen tarkoituksena on nopeuttaa varsinaista digitointia. Järjestämiseen kuluva aika on täysin riippuvainen siitä, millainen aineisto on kyseessä. Esikäsittelyyn saattaa kulua huomattavasti enemmän aikaa kuin siihen on varattu. Siksi olisi syytä tehdä suunnitteluvaiheessa testejä, jotta aikataulut pitävät.

Valmisteluvaihe voidaan myös ulkoistaa, mutta useimmiten valmistelu suoritetaan organisaation omana työnä.

Valmisteluun kuuluu usein myös aineiston kuvailu. Kuvailutaso on määritelty metatietomäärittelyssä. Digitoinnin yhteydessä syntyy lisää kuvailutietoja,

ja on huolehdittava, että tiedot siirtyvät tiedostojen mukana.

Suunnittelussa on alustavasti valittu digitointitapa, ja sitä noudatetaan, mikäli aineisto vastaa suunniteltua. Joskus valmisteluvaiheessa huomataan, että digitointitapaa joudutaan muuttamaan.

## Digitointi vaatii logistiikkaa

Digitointiprosessissa tarvitaan logistiikkaa, oli kyse sitten pienestä tai isosta prosessista. Logistiikan prosessiin kuuluu aineiston valmistelu siirtovalmiuteen, kuljetuksen valmistelu, nouto, kuljetus, aineiston vastaanotto sekä aineiston palauttaminen takaisin arkistoon tai varastoon. Vaihtoehtona on myös, että digitoitua aineistoa ei palauteta takaisin, kuten massadigitoinnin prosessissa tapahtuu.

Aineiston siirtämisessä on aina huomioitava, onko aineisto altis olosuhdemuutoksille. On pohdittava, miten mahdolliset haitalliset vaikutukset estetään. Haitallisia vaikutuksia voivat olla vaikkapa lämpötilan vaihtelut tai kuljetuksessa tapahtuvat ei-toivotut asiat.

Mikäli digitointi ulkoistetaan, on otettava huomioon kuljetusvälineet, joilla aineistoa siirretään arkistosta noutopaikalle, kuten lastausalueelle. Digitointipalvelua kilpailutettaessa pitää muistaa ottaa huomioon aineiston kuljetus suorittavalle toimijalle sekä sen palautus.

Siirrosta sopiminen ja kuljetuksen suunnittelu on tärkeää

prosessin kannalta, varsinkin jos palveluntuottajan kanssa on sovittu säännöllisistä siirtoaikatauluista. On tärkeää aineiston eheyden ja löydettävyyden kannalta, että siirrettävä aineisto on merkitty sovitusti. Tähän saattaa liittyä tunnistetietojen merkitä ja aineiston suojaamista siirtoa varten. Pienemmässä mittakaavassa aineiston valmistelun, digitoinnin ja tarkistuksen tekee yksi ja sama ihminen. Isommissa prosesseissa jokaisessa vaiheessa on eri tekijät.

## Digitoidaan laadukkaasti

Digitoinnilla on yleensä kaksi päätavoitetta: entistä parempi saavutettavuus ja aineiston säilyminen. Usein digitoinnilla halutaan varmistaa kumpikin tavoite. Digitointi on prosessi, jolle on määritelty tarkat vaatimukset, jotta lopputulos olisi laadukas ja tarvittaessa pitkäaikaissäilytykseen kelpaava, arkistokelpoinen sekä asiakkaille helposti käytettävissä.

Digitoinnin suunnittelussa on huolehdittava siitä, että työtilat sopivat digitointiin. Tilan on oltava pölytön. Se varmistetaan säännöllisellä siivouksella. Laitteiden testauksesta ja kalibroinnista on syytä olla hyvät ohjeet, ja laadunvarmistuksen yhteydessä on seurattava, että ohjeita noudatetaan.

Digitointi skannaamalla tai valokuvaamalla on yleensä prosessin nopein vaihe. Tämä toki riippuu jälleen siitä, millaisesta aineistosta on kyse. Massadigitoinnissa skannaus sujuu nopeasti. Suuren aineiston

kuvaaminen kameralla vie luonnollisesti enemmän aikaa.

Laadunvarmistuksen tarkoituksena on huolehtia siitä, että digitoinnin lopputuloksena syntyy suunnitelman mukaista, tasalaatuista digitoitua aineistoa. Digitoinnin jälkeen tarkistetaan kuvailutiedot ja varmistetaan, että aineisto on käytettävissä.

Ennen digitointiprosessia on syytä tehdä laadun- tarkkailusuunnitelma sekä määritellä ne kriteerit, joihin laaduntarkkailussa kiinnitetään erityistä huomiota. Laadua varmistetaan koko prosessin ajan. Laadunvarmistuksessa voidaan hyödyntää erilaisia mittatauluja, referenssiarvoja tai analysointiohjelmistoja.

Jokaisessa prosessissa on parantamisen varaa. Koko prosessin tarkalla dokumentoinnilla saadaan selville mahdolliset kehityskohteet.

*Aineiston valinnassa kannattaa aloittaa tärkeän ja tarvittavan aineiston määrittelystä sekä siitä, mitä digitoinnilla halutaan saavuttaa.*

*Digitointia suunniteltaessa on otettava alusta lähtien huomioon saavutettavuus ja käytettävyys.*

*Kuvailussa on arvioitava, millaisin metatiedoin aineistoa halutaan käyttää sähköisessä säilytysjärjestelmässä.*

#### Lähteet:

Isomursu, A. 2011. Kuvansa kullakin. Valokuva-arkiston hallinnan kysymyksiä. Helsinki: Suomen valokuvataiteen museo.

Kansallisarkisto. 2017. Massadigitoinnin suunnitteluprojekti. WWW-dokumentti. Saatavissa: <https://kansallisarkisto.fi/fi/viranomaisille/siirrot/massadigitointi/massadigitoinninsuunnitteluprojekti-4> [viitattu 16.11.2022].



Kansallisarkisto. 2022. Digitointi. WWW-dokumentti. Saatavissa: <https://kansallisarkisto.fi/fi/viranomaisille/digitointi> [viitattu 15.11.2022].

Lilja, J., Hakkarainen, J-P., Karppinen, P. & Lampinen, P. 2021. Kansalliskirjaston Digitointiohjelma 2021–2024. Kansalliskirjasto. PDF-dokumentti. Saatavissa: <https://urn.fi/URN:ISBN:978-951-51-6993-8> [viitattu 15.11.2022].

Museovirasto. 2022a. Museoalan kehittämisen ohjeita ja oppaita. WWW-dokumentti. Saatavissa: <https://www.museovirasto.fi/fi/palvelut-ja-ohjeet/museoalan-kehittaminen> [viitattu 15.11.2022].

Museovirasto. 2022b. Digitoinnin laadunhallinta. WWW-dokumentti. Saatavissa: <https://www.museovirasto.fi/fi/palvelut-ja-ohjeet/museoalan-kehittaminen/digitoinnin-laadunhallinta> [viitattu 15.11.2022].

# Alkuun digitointiprojektissa

SARI JÄRN

*Digitointiprojektin suunnittelu alkaa kysymyksillä. Tärkeimpiä kysymyksiä ovat, mitä hyötyä digitoinnilla tavoitellaan ja keitä muutos palvelee.*

Aineiston käyttötarve määrittää tarvittavan projektin laajuuden ja laadun. Kun digitointiprojektin tarpeet, hyödyt ja vaikutukset on määriteltä, luonnostellaan digitoinnin tiekartta, johon kirjataan alustavasti digitoitavan aineiston sisältö sekä mahdollinen digitointijärjestys. Projektin nimeäminen auttaa hahmottamaan, millaisesta projektista on kyse.

Digitointia suunniteltaessa on otettava huomioon aineiston digitointiin, säilyttämiseen ja julkaisemiseen liittyvää lainsäädäntöä. Kun digitointi tehdään ohjeiden mukaisesti sekä digitoitu aineisto säilytetään ja julkaistaan lainsäädännön ja määräysten mukaan, aineiston saatavuus ja säilyminen varmistetaan.

Projektin suunnittelussa kannattaa käyttää apuna erilaisia projektinhallinnan välineitä esimerkiksi Trello, Microsoft Excel, Google Sheets, LibreOffice Calc, Microsoft Project, Microsoft Teamsin Tasks-sovellus. Digitoinnin jakaminen työpaketteihin helpottaa kokonaisuuden hahmottamista. Digitoitavan aineiston kokonaisuus vaikuttaa siihen, millainen projektiorganisaatio tarvitaan. Tehtäväkokonaisuudet on hyvä aikatauluttaa, jotta projektin resurssit saadaan täysimääräisesti käyttöön.

Digitointiprosessiin kuuluu säilyttämisen lisäksi käyttöön saattaminen. Digitoinnin päämäärän pitää olla kirkkana mielessä.

## Näillä pohdinnoilla pääsee alkuun!

### Valmistelu (tarve ja määrittely)

- Miksi digitointia suunnitellaan? Mitä tarkoitusta se palvelee?
- Mikä rooli digitoinnilla on organisaatiossa?
- Onko kyse yksittäisestä projektista vai jatkuuko digitointi projektin jälkeen?
- Kuka omistaa aineiston? Viranomainen, yritys, säätiö, yhdistys?
- Onko kyse pitkäaikaisäilytyksestä ja/tai verkkojulkaisemisesta?
- Ketkä aineistoa käyttävät? Organisaation sisäinen käyttäjä, loppukäyttäjä, maksava asiakas?
- Miten aineisto kuvaillaan, että se palvelee käyttäjiä?

### Projektisuunnitelma

- Mitä digitointi maksaa? (resurssit, kustannukset ja rahoitusväylät, henkilöstö, laitteisto, budjetti)
- Miten työtä jaetaan omassa organisaatiossa ja mahdollisissa yhteistyöorganisaatioissa? (palvelumallit, vastuunjako)
- Missä sähköinen aineisto säilytetään ja julkaistaan?

### Digitointi

- Millaista aineistoa aiotaan digitoida?
- Missä digitoidaan?
- Miten aineistosirrot toteutetaan?
- Tarjotaanko omaa digitointipalvelua muille?

### Hankinnat

- Millaisia digitointipalveluja on tarjolla? (markkinakartoitus)
- Mitä digitoinnin kilpailutuksessa ja sopimuksissa tulee huomioida? (hankinnat, markkinavuoropuhelu)
- Tarvitaanko sopimuksia ulkopuolisen toimijan kanssa?

### Arviointi

- Mitä jos digitoinnin aikana tapahtuu muutoksia? (muutoksenhallinta)

## Mieti paras toteuttamisvaihtoehto digitoinnille!

## Selvitä aina

sisältääkö digitoitava aineisto

- arkaluontoisia henkilötietoja
- salassa pidettäväksi määrättyjä tietoja
- tekijänoikeuden alaista aineistoa
- tiedonhallintalain alaista aineistoa

## Tutustu!

Kulttuuriperintö-PAS-palvelu  
<https://digitalpreservation.fi/>

Digitaalinen kulttuuriperintömme  
<https://www.digime.fi/>

KAM-kuvailuryhmä  
<https://www.kiwi.fi/pages/viewpage.action?pagelId=85464271>

Sisällönkuvailun automatisoinnin haasteita ja ratkaisuja kulttuuriperintöorganisaatioissa  
<https://urn.fi/URN:ISBN:978-951-51-7233-4>

Suomen tiedekustantajien liiton digitointiohje  
<https://tiedekustantajat.fi/digitointiohjeita/>

Digitointi Musiikkiarkistossa  
<https://musiikkiarkisto.fi/toiminta/digitointi.php>

IASA (The International Association of Sound and Audiovisual Archives)  
<https://www.iasa-web.org/iasa-special-and-technical-publications>

Finto AI, automaattisen sisällönkuvailun palvelu  
<https://ai.finto.fi/>

## Lainsäädäntöä, määräyksiä ja ohjeita

EU:n yleinen tietosuoja-asetus: <https://eur-lex.europa.eu/legal-content/FI/TXT/?uri=celex:32016R0679>

Tietosuoja laki  
<https://www.finlex.fi/fi/laki/ajantasa/2018/20181050>

Tekijänoikeuslaki  
<https://www.finlex.fi/fi/laki/ajantasa/1961/19610404>

Tekijänoikeusasetus  
<https://www.finlex.fi/fi/laki/ajantasa/1995/19950574>

Tiedonhallintalaki  
<https://vm.fi/tiedonhallintalaki>

### Formaatteja

<https://www.kiwi.fi/display/formaattija-konvertointipalvelu/Formaatit>

### Standardeja

<https://sales.sfs.fi/fi>  
<https://www.loc.gov/librarians/standards>

### Sanastoja

<https://finto.fi/fi/>

### Valtion viranomainen/kunta

<https://kansallisarkisto.fi/fi/viranomaisille>

### Museo

<https://www.museoliitto.fi/museoala/lainsaadanto>  
<https://www.museovirasto.fi/fi/palvelut-ja-ohjeet/museoalan-kehittaminen>

### Yksityisarkisto

<http://www.yksityisetkeskusarkistot.fi/arkistointiopas/>

### Yhdistys

<https://kotiseutuliiitto.fi/tietopankki/yhdistystoiminta/yhdistyksen-arkistointiohje/>



## Digitoinnin toteuttamisvaihtoehtoja

1

### Digitoidaan itse omilla laitteilla omassa toimitiloissa

Vaatimukset: laitteet, ohjelmistot, työvoima, tiedonhallintajärjestelmä, tallennustila.

2

### Digitoidaan itse omilla laitteilla omassa toimitiloissa ja ulkoistetaan osa digitoinnista

Vaatimukset: laitteet, ohjelmistot, työvoima, tiedonhallintajärjestelmä, tallennustila, palvelun kilpailutus, palvelukokonaisuuden määrittely.

3

### Digitointi ulkoistetaan, valmistelu tehdään itse

Vaatimukset: oma työvoima, palvelun kilpailutus, palvelukokonaisuuden määrittely.

4

### Digitointi ja valmistelu ulkoistetaan

Vaatimukset: palvelun kilpailutus, palvelukokonaisuuden määrittely.

#### Lähteet:

Arkistolaki 23.9.1994/831.

CSC. 2022. Kansalliset pitkäaikaispalvelut. WWW-dokumentti. Saatavissa: <https://www.digitalpreservation.fi/> [viitattu 16.11.2022].

Digime-tietoarkkitehtuuriryhmä 2021. DIGIME-standardisalkku. PDF-dokumentti. Saatavissa: <https://urn.fi/URN:NBN:fi-fe202103258366> [viitattu 16.11.2022].

Kansallisarkisto. 2019. Kansallisarkiston vaatimukset hävittämiseen tähtäävään digitointiin. KA/3357/07.01.02.04.02/2019. PDF-dokumentti. Saatavissa: [https://kansallisarkisto.fi/uploads/normit/valtionhallinto/maarayksetjaohjeet/Kansallisarkiston\\_vaatimukset\\_havittamiseen\\_tahtaavaan\\_digitointiin.pdf](https://kansallisarkisto.fi/uploads/normit/valtionhallinto/maarayksetjaohjeet/Kansallisarkiston_vaatimukset_havittamiseen_tahtaavaan_digitointiin.pdf) [viitattu 16.11.2022].

Kansallisarkisto. 2021. Kansallisarkiston määräykset arkistoitavien erikoisaineistojen digitointiin. Ohje. KA/15780/07.01.02.04.02/2020. PDF-dokumentti. Saatavissa: <https://kansallisarkisto.fi/uploads/Viranomaisille/M%C3%A4%C3%A4r%C3%A4ykset%20ja%20ohjeet/Siirtom%C3%A4%C3%A4r%C3%A4ys/KA%20m%C3%A4%C3%A4ritykset%20erikoisaineistojen%20digitointiin.pdf> [viitattu 16.11.2022].

Laki julkisen hallinnon tiedonhallinnasta 906/2019.

Laki kulttuuriaineistojen tallettamisesta ja säilyttämisestä 28.12.2007/1433.

Museovirasto s.a. Museotalan kehittämisen ohjeita ja oppaita. WWW-dokumentti. Saatavissa: <https://www.museovirasto.fi/fi/palvelut-ja-ohjeet/museotalan-kehittaminen> [viitattu 16.11.2022].

Suomen Kotiseutuliitto s.a. Yhdistyksen arkistointiohje. WWW-dokumentti. Saatavissa: <https://kotiseutuliitto.fi/tietopankki/yhdistystoiminta/yhdistyksen-arkistointiohje/> [viitattu 16.11.2022].

Valtionvarainministeriö s.a. Tiedonhallintalautakunta. WWW-dokumentti. Saatavissa: <https://vm.fi/tiedonhallintalautakunta> [viitattu 16.11.2022].

Yksityiset keskusarkistot s.a. Järjestöjen/yritysten arkistointiohjeet. PDF-dokumentti. Saatavissa: <http://www.yksityisetkeskusarkistot.fi/wp-content/uploads/2018/05/arkistointiopas.pdf> [viitattu 16.11.2022].

# Digitointiprojektin hallinta

SARI JÄRN, EMMI LIIKANEN JA HENNA RISTOLAINEN

*Digitointiprojekteja voidaan tehdä pienessä tai suuressa mittakaavassa. Jokainen projekti on omanlaisensa. Jos digitoitavaa aineistoa on paljon tai oma osaaminen ei riitä halutun lopputuloksen aikaansaamiseksi, voidaan palvelu hankkia ulkopuolelta. Jokainen projekti kannattaa kuitenkin suunnitella huolellisesti.*

Digitointiprojektilla pitää aina olla päämäärä. Tämä selviää, kun kirjoitetaan auki taustat ja syyt digitoinnille. On tärkeää tunnistaa, miksi projekti kannattaa aloittaa. Jos tarpeeksi merkityksellistä syytä ei löydy, on syytä pohtia, kannattaako hankkeeseen lainkaan ryhtyä. Löytyykö hankkeelle merkityksellisyys ja sitä kautta oikeutus?

Digitointiprojekteja suunnitellaan ja toteutetaan erilaisissa toimintaympäristöissä. Pienen kotiseutumuseon digitointiprojekti on erilainen kuin suuren valtionviraston digitoinnin tiekartta. Kotiseutumuseon sidosryhmät ovat toiset kuin viraston.

## Digitointiprojekti alkaa valmistelulla

Digitointia suunniteltaessa on tärkeää tunnistaa sidosryhmät, jotta projektin tavoitteet ja perustelut voidaan kohdentaa oikein. Valmistelussa kirjoitetaan alustava projektin kuvaus, johon yksilöidään ja analysoidaan digitoinnin hyödyt. Sidosryhmien lisäksi on tunnistettava käyttäjien vaatimukset. Ketä ja miten aineiston digitointi hyödyttää?

Projektin aikana luodaan digitointiprosessi, jonka vaiheet tunnistetaan ja kirjoitetaan auki. Dokumentointi kannattaa aloittaa jo valmisteluvaiheessa. Muistioihin voi palata projektin aikana. Ideointiin kannattaa ottaa mukaan kaikki ne, joiden työhön digitointi vaikuttaa, joten on tärkeää varmistaa osallisuus kaikista käyttäjäryhmistä.

Projektin valmistelun aikana syntyneitä dokumentteja kannattaa käyttää hyväksi projektisuunnitelman laatimisessa. Projektioorganisaatiota suunniteltaessa kannattaa pohtia, löytyykö käyttäjistä tai sidosryhmistä ihmisiä, jotka voivat olla mukana projektissa – jos eivät työntekijöinä, niin ohjausryhmän jäseninä. Resurssit ja vastuut pitää suunnitella huolella: henkilöstö, tilat, laitteet, materiaalit ja työvälineet. Jo projektin valmistelun aikana pitää miettiä, onko tekemiseen varattu riittävästi sisäisiä tai ulkoisia resursseja. Tähän liittyy myös projektissa hyväksyttävä riskitaso. Suunnitelmaan pitää kirjata, miten toimitaan, jos resurssissa tapahtuu muutoksia. Resurssit pitää suunniteltaessa pitää myös



määritellä tarvittava osaaminen. Suunnitelmassa aineisto kuvataan tarkemmin kuin valmistelun aikana. Aineiston laajuus tulee selvittää mahdollisimman tarkasti, ja projektista on rajattava pois siihen kuulumaton aineisto. Tarkasti kuvattu ja rajattu aineisto helpottaa myös mahdollisen hankinnan tekemistä.

### Projektisuunnitelmaan kannattaa käyttää voimavaroja

Suunnitelmaan on kirjattava, mitkä määräykset, ohjeet, standardit ja lait vaikuttavat digitointiprojektiin. Nämä kaikki voivat vaikuttaa digitoinnin lopputulokseen ja siihen, onko digitoitu aineisto käyttäjien saatavilla.

Kustannukset on arvioitava mahdollisimman tarkasti. Budjetin pitää olla realistinen. Kustannukset vaikuttavat suoraan siihen, saako digitointiprojekti rahoituksen. Projektin oikeutuksen tulee toteutua myös kustannuksissa. Valmistelussa auki kirjoitettu hankkeen merkityksellisyys ei välttämättä riitä, jos kustannuksia tulee liikaa. Kustannuksia mietittäessä voi käyttää apuna markkinavuoropuhelua.

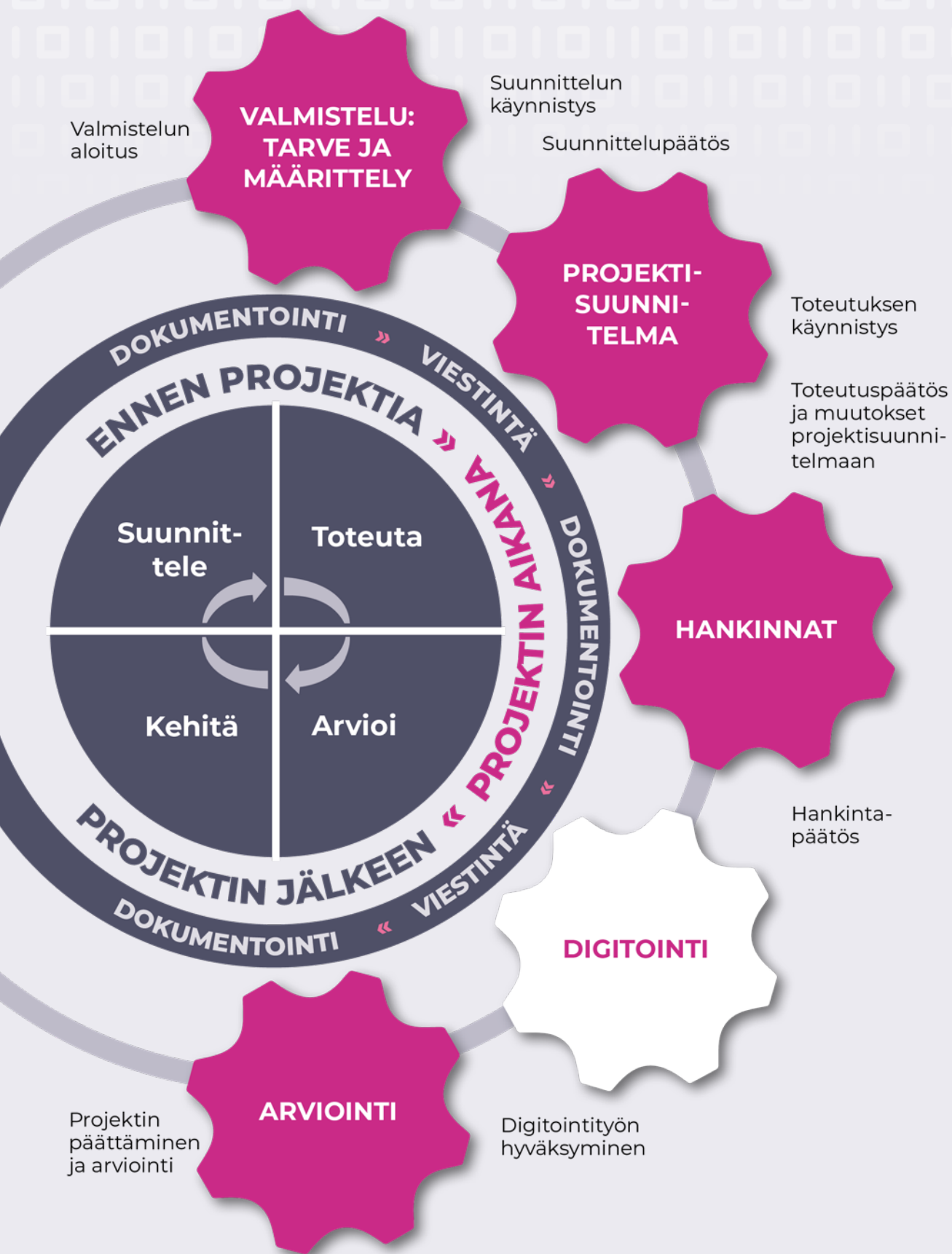
Projektin aikataulutusta helpottaa työpakettien tekeminen. Kun projekti jaetaan osiin, on myös helpompi huolehtia kokonaislaadun hallinnasta. Jo valmisteluvaiheessa toimintaympäristöä kartoitettaessa on pitänyt pohtia, miten laadun arviointi suunnitellaan ja varmistetaan sekä miten sitä valvotaan. Laadun lisäksi on syytä kiinnittää huomiota tieto-

**Katso liite sivulta 59:  
Esimerkki digitointi-  
projektin projekti-  
suunnitelman sisälly-  
luettelosta.**

ja työturvallisuuteen. Turvallisuus liittyy digitoinnin lopputulokseen myös laadullisesti. Salassa pidettävää aineistoa ei saa päätyä julkiseksi, ja työolosuhteiden tulee olla sellaiset, että on mahdollista tehdä laadukasta digitointia. Laadun vähimmäisvaatimusten määrittelyyn kirjataan, mitä määräyksiä, ohjeita, standardeja ja lakeja noudatetaan, sillä ne sanelevat myös laatuvaatimuksia. Kotiseutumuseon valokuvakokoelman digitoinnin laadun vähimmäisvaatimukset voivat olla erilaisia kuin viranomaisaineiston.

Viestintään ja tiedonhallinnan suunnitteluun pitää varata aikaa koko projektin ajaksi. Dokumentointi ja viestintä valmistelusta alkaen projektin päättämiseen saakka pitää tehdä järjestelmällisesti ja suunnitellusti.





Kuva 1. Esimerkki digitointiprojektin elinkaaresta ja kehittämisestä (soveltaen SFS-ISO 21502:2021 –standardi ja PDCA-malli)

## VALMISTELU: TARVE JA MÄÄRITTELY

- Taustan, syyn ja merkityksen tunnistaminen ja kuvaaminen
- Toimintaympäristön huomioiminen
- Sidosryhmien tunnistaminen
- Tavoitteet ja perustelut
- Hyötyjen yksilöinti ja analysointi
- Digitointiprosessin tunnistaminen ja dokumentointi
- Käyttäjän vaatimusten tunnistaminen, kuvaaminen ja täsmentäminen
- Ideointi, mahdollisuuksien kartoitus ja valmistelu

### DOKUMENTIT

- Alustava suunnitelma
- Projektikuvaus
- Projektiesitys
- Valmistelumuistiot
- Toteutettavuusanalyysi

## PROJEKTI- SUUNNITELMA

- Resurssien ja vastuiden suunnittelu: henkilöstö, tilat, laitteet, materiaalit, työvälineet
- Projektiorganisaation suunnittelu
- Aineiston kuvaus, laajuus ja rajaus
- Määräysten, ohjeiden, standardien ja lakien huomioiminen
- Kustannusten arviointi ja budjetin laadinta
- Tarvittavien ostopalveluiden suunnittelu
- Aikataulutus
- Kokonaislaadunhallinta: laadun ja laadun arvioinnin suunnittelu, varmistaminen ja valvonta
- Laadun vähimmäisvaatimusten määrittely
- Hyväksyttävän riskitason määrittely
- Turvallisuus ja työturvallisuus
- Tarvittavan ja uuden osaamisen määrittely
- Tarvittavien ohjeiden määrittely
- Viestinnän ja tiedonhallinnan suunnittelu
- Projektin päättämisen suunnittelu

### DOKUMENTIT

- Projektisuunnitelma
- Laatusuunnitelma
- Tehtäväjaon rakenne
- Riskienhallinta-suunnitelma
- Hankintasuunnitelma
- Budjetti
- Vaiheistettu ja aikataulutettu toteutussuunnitelma
- Tuloksen kuvaus
- Vaatimusmäärittelyluonnos
- Metatietomäärittelyluonnos
- Viestintäsuunnitelma
- Suunnittelumuistiot

## HANKINNAT

- Hankintojen suunnittelu ja ohjaus
- Vaatimusmäärittelyt
- Markkinavuoropuhelu
- Kilpailutusstandardien määrittely
- Tarjouskilpailutukset ja tarjouspyynnöt
- Tarjosten analyysi ja vertailu
- Toimittajien valinta
- Sopimuksista sopiminen ja neuvottelut
- Sopimusten valmistelu ja hallinnointi
- Hankintojen toteutus

### DOKUMENTIT

- Tarjousasiakirjat
- Palvelukuvaus
- Metatietomäärittely ja nimeämisohje
- Palvelutasovaatimukset
- Vaatimusmäärittely
- Hankinnan arviointiperusteet
- Käyttötapaesimerkit
- Täydennettävä projektisuunnitelmaluonnos
- Täydennettävä sopimusluonnos
- Sopimus

## DIGITOINTI

- Toteutuksen valmistelu ja suunnittelu
- Ohjeiden teko
- Aineiston valinta
- Säilytettävän sisällön tarkistaminen, aineistojen järjestäminen
- Testaus tai pilotointi
- Aineiston valmistelu
- Aineistologistiikka
- Laitteiden huolto ja ylläpito
- Digitointi: mm. skannaus/kuvaus
- Metatietojen tuottaminen ja tekstintunnistus
- Digitaalisen aineiston validointi/laadunvarmistus
- Digitaalisen aineiston arkistointi
- Käyttöön saattaminen
- Mittaus, analysointi ja toiminnan arviointi
- Sopimusten mukaisen toteutumisen valvonta ja muu etenemisen seuranta

### DOKUMENTIT

- Digitoinnin dokumentaatio
- Katselmusdokumentit
- Vaatimusmäärittelyjen toteutumisen seuranta-dokumentti
- Hankinnan laatu-kriteerien mukaisuuden seuranta-dokumentti
- Digitointiohjeet
- Perekäytösraportti
- Laitteiden huolto-raportti
- Seurantareportti
- Muistiot

## ARVIOINTI

- Tuotosten ja palvelujen arviointi
- Toteutuneiden hyötyjen ja tavoitteiden saavuttamisen arviointi
- Projektisuunnitelman toimenpiteiden ja tulosten toteutumisen arviointi
- Loppuraportoinnin toteutus
- Riskien ja muutosten hallinnan arviointi
- Asiakastytytyväisyyden arviointi
- Opittujen asioiden tunnistaminen ja tiedottaminen
- Sisäinen ja ulkoinen viestintä
- Kehittäminen

### DOKUMENTIT

- Vaatimusmäärittelyjen toteutumisen arviointidokumentti
- Hankinnan laatu-kriteerien mukaisuuden arviointidokumentti
- Loppuraportti

## VIESTINTÄ SEKÄ TIEDON JA DOKUMENTAATION HALLINTA

## Digitointiprojektien hallinnan piirteitä erilaisissa organisaatioissa

Digitoinnin taitajaksi -koulutukseen (2022) kuului projektinhallinnan suunnittelua, ja opiskelijat tekivät ryhmätöinä kuvitteellisia digitointiprojekteja. Digitointiprojektin hallintaan liittyvät esimerkit ovat mukailtuja koosteita ja yhdistelmiä näistä digitointiprojekteista. Ne edustavat erilaisia hankkeita ja toimijoita sekä tuovat kunkin organisaation osalta esille erilaisia piirteitä.

**Kuva 3.**  
Esimerkki valtion viranomaisen massadigitointiprojektin erityispiirteistä.



# VALTIO - massadigitointi

Massadigitointi on Kansallisarkiston vaatima toimenpide aineiston arkistointia varten. Tavoitteena on edistää tiedon saata- vuutta ja käytettävyyttä. Digitointi vähentää viraston paperiarkistojen vaatimaa tilaa.

### TAVOITTEET

Tavoitteena on täyttää valtion viraston velvoite pysyvästi säilytettävän asiakirja-aineiston luovuttamisesta Kansallisarkistoon. Massadigitoinnin jälkeen valtionhallinnon ja maakuntahallinnon viranomaisille ei jää paperiarkistoja.

### TOIMINTAYMPÄRISTÖ

Valtion viranomaisen valmistee ja toimittaa digitoitavan aineiston Kansallisarkiston massa- digitointikeskukseen.

### TUOTOKSET

Viranomaisen arkisto- aineisto on digitoitu ja siirtynyt Kansallisarkiston säilytettäväksi. Digitointi- projektin lisätavoitteena on pystyä luopumaan eräistä arkistotiloista ja keskittää jäljelle jäävä (määräajan säilytettävä) arkistoaineisto muutamalle paikka- kunnalle.

### TULOKSET

Viranomaisen arkistoaineistoa on digitoitu 500 hyllymetriä.

### SIDOS- RYHMÄT

Valtion viran- omaisen ja viraston asiakkaat.

### KÄYTETYT RESURSSIT

Valtion viranomaisen valmisteluun osallistunut henkilökunta, siirrot, massadigitointi ja mahdollinen aineis- ton tuhoaminen.

### MONI- MUTKAISUUS

Viranomaisen arkisto- aineiston valmistelussa on ilmennyt erikoisai- neistoja. Noudatettava Kansallisarkiston määräyksiä. ja ohjeita.

### RAJOITTEET

Massadigitoinnin ulkopuolelle rajautuu aineisto, joka ei fyysisten tai muiden ominaisuuksien perusteella sovellu massadigitoitavaksi. Kansallisarkiston mää- räykset ja ohjeet.

### KÄYTETYT PROSESSIT TAI MENETELMÄT

Tiekartta, jossa on mää- ritelty massadigitoinnin palveluun sisältyvät tieto- aineistot, niiden digitoin- tijärjestys ja digitoinnin aikataulu.

# KUNTA – rakennusvalvonta

Digitoinnille on ollut tarvetta säilytystilan vähyyden vuoksi, mutta digitointiin ei ole ollut osoittaa resursseja. Digitointia on rajoittanut sekin, ettei kunnassa ole ollut digitaaliselle aineistolle sopivaa arkistojärjestelmää.

## TAVOITTEET

Tavoitteena on saada rakennusluvut helpommin käytettäväksi kunnan rakennusvalvonnassa. Alkuperäistä aineistoa ei tuhota, vaan se säilytetään edelleen kunnan päätearkistossa.

## TOIMINTAYMPÄRISTÖ

Aineisto valmistellaan rakennusvalvonnan tiloissa digitointia varten. Palveluntuottaja digitoi aineiston.

## TUOTOKSET

Määritellyt rakennusluvut on digitoitu ja siirretty lupapisteeseen. Digitointiprosessi on testattu ja päivitetty.

## TULOKSET

Rakennuslupia on digitoitu 12 hyllymetriä.

## SIDOSRYHMÄT

Prosessien ja palvelualueiden johtoryhmät sekä aineistoja käyttävä henkilöstö.

## KÄYTETYT RESURSSIT

Projektin omistaja on rakennus- ja ympäristövalvonnasta vastaava viranhaltija. Projektipäällikkönä toimii tiedonhallinnan asiantuntija. Projektin asiantuntijoina ovat rakennusvalvonnan asiantuntijat ja lupasihteerit. Sihteerit digitoivat lupia omalla työajallaan kunnan laitteilla.

## MONIMUTKAISUUS

Digitoijat tekevät työtä oman työnsä ohella. Kiireisinä aikoina digitointiin ei olekaan käytettävissä resursseja.

## KÄYTETYT PROSESSIT TAI MENETELMÄT

Digitointiprosessin omistaja on hallintojohtaja, jonka alaisuudessa toimiva tiedonhallintatiimi koordinoi prosessia.

## RAJOITTEET

Ei hallita kaikkia digitoinnin vaiheita.

### Kuva 4.

Esimerkki kunnan rakennusvalvonnan digitointiprojektin erityispiirteistä.

# MUSEO – valokuvakokoelma

Valokuva-aineisto kootaan kokoelmätietokannasta. Käydään aineisto läpi ja valitaan, mitä digitoidaan. Kuva-valintoihin vaikuttavat tekijänoikeudet sekä kuvien laatu ja kunto. Arvioidaan käytettävyys ja yleinen kiinnostavuus.

## TAVOITTEET

Saada kuvia avoimesti käyttöön.

## TOIMINTAYMPÄRISTÖ

Paperikuvat taso-skannerilla. Mahdolliset negatiiviskannerilla. Tutkitaan myös vaihtoehtoa kameralla digitoimiseen.

## TUOTOKSET

Kuvia vapaaseen käyttöön. Tunnistetut kuvat kertovat tuleville sukupolville suomalaisista kulttuurintekijöistä.

## TULOKSET

Valokuvia on digitoitu 1 500.

## SIDOSRYHMÄT

Kulttuuriperintöorganisaatiot, suuri yleisö ja tutkijat.

## KÄYTETYT RESURSSIT

Arkistonhoitaja, harjoittelija ja järjestelmäasiantuntija.

## MONIMUTKAISUUS

Harjoittelija on perehdytettävä kunnolla digitointiin. Tarvitaan ohjeita.

## RAJOITTEET

Tekijänoikeudet (valokuvaajat, taiteilijat), julkisuuslaki (asiakirjojen julkisuus), tietosuojalaki ja -asetus sekä museolaki.

## KÄYTETYT PROSESSIT TAI MENETELMÄT

Arkistonhoitaja valikoi ja järjestää aineiston sekä skannaa ja luettelee. Harjoittelija järjestää aineiston, skannaa ja luettelee. Järjestelmäasiantuntijan vastuulla ovat PAS-siirto, skannerin kalibrointi ja määritykset. Arkistonhoitaja ja järjestelmäasiantuntija tekevät ohjeistukset muun muassa analogisen ja digitaalisten aineistojen käsittelystä sekä aineistojen siirrosta. Ohjauksen hoitaa arkistonhoitaja.

## Kuva 5.

Esimerkki museon valokuvakokoelman digitointiprojektin erityispiirteistä.



# YHDISTYS – Valokuvakokoelma

Kotiseutuyhdistyksen valokuva-arkistossa on valokuvia 1920-luvulta 1990-luvulle. Aineisto on analogista. Suurin osa on vedoksia, pieni osa lasinegatiiveja. Digitointiprojektissa digitoidaan skannaamalla kyseisen valokuva-aineiston keskeisin osa.

## TAVOITTEET

Digitointiprojektin tavoitteena on lisätä aineiston saavutettavuutta ja turvata aineiston pitkäikäisyys. Lasinegatiivaineiston käytettävyys lisääntyy merkittävästi digitoinnilla.

## TOIMINTAYMPÄRISTÖ

Vedokset skannataan tasoskannerilla, negatiivit negatiiviskannerilla. Tutkitaan myös vaihtoehtoa kameralla digitoimiseen.

## TUOTOKSET

Digitoinnin myötä valokuva-aineisto saadaan pakattua tiiviisti, jolloin aineisto saadaan siirrettyä aiempaa pienempään ja olosuhteiltaan parempaan tilaan. Digitoitu aineisto on yhdistyksen käytössä.

## TULOKSET

Kuvia on digitoitu 250.

## SIDOSRYHMÄT

Yhdistyksen jäsenet ja tutkijat.

## KÄYTETYT RESURSSIT

Yhdistyksen toimijat, harjoittelija, skanneri ja kamera.

## MONI-MUTKAISUUS

Harjoittelija on perehdytettävä kunnolla digitointiin. Tarvitaan ohjeita. Aineisto on haurasta.

## KÄYTETYT PROSESSIT TAI MENETELMÄT

Alueellinen museo ohjaa toimintaa.

## RAJOITTEET

Tekijänoikeudet (valokuvaajat, taiteilijat)

## Kuva 6.

Esimerkki yhdistyksen valokuvakokoelman digitointiprojektin erityispiirteistä.



# YKSITYINEN - taloyhtiön arkisto

Projektissa käydään läpi taloyhtiön paperiaineisto, seulotaan säilytettävät ja valitaan digitoitavat asiakirjat. Aineistoon kuuluu muun muassa rakennuspiirustuksia, sähköpiirustuksia, taloyhtiön pöytäkirjoja ja sopimuksia. Vanhimmat digitoitavat asiakirjat ovat 1970-luvulta.

## TAVOITTEET

Digitoinnilla turvataan aineiston säilyminen ja käytettävyys. Aineiston seulonnan yhteydessä saadaan myös tuhottua ne asiakirjat, joita ei tarvitse enää säilyttää.

## TOIMINTAYMPÄRISTÖ

Rakennuspiirustusten digitointityö ulkoistetaan. Hankitaan taloyhtiölle ulkoinen kovalevy sekä muutamia pahvikansioita. Yhtiöllä on oma skanneri, jolla voidaan skannata tavalliset asiakirjat.

## TUOTOKSET

Seulonnan ja digitoinnin myötä arkistoaineisto saadaan siirrettyä aiempaa pienempään ja olosuhteiltaan parempaan tilaan. Digitoitu aineisto on taloyhtiön käytössä.

## TULOKSET

Arkistomateriaalia on digitoitu kaksi hyllymetriä.

## SIDOSRYHMÄT

Taloyhtiön jäsenet.

## KÄYTETYT RESURSSIT

Yhtiön osakkaat osallistuvat toteutukseen. Isännöitsijä osallistuu aineiston läpikäyntiin. Oma skanneri.

## MONI-MUTKAISUUS

Yhdistysten aineistojen arkistointiin ja seulptaan on erilaisia ohjeita esimerkiksi Yksityisillä keskusarkistoilla, Kansallisarkistolla, Suomen valokuvataiteen museolla, Kotiseutuliitolla ja Digitalian digitointioppaassa.

## KÄYTETYT PROSESSIT TAI MENETLMÄT

Noudatetaan Kotiseutuliiton ja Yksityisten keskusarkistojen ohjeita.

## RAJOITTEET

Lainsäädäntö.

**Kuva 7.** Esimerkki yksityisen taloyhtiön digitointiprojektin erityispiirteistä.



#### Lähteet:

Artto K., Martinsuo M. & Kujala J. 2006. Projekttiliiketoiminta. WSOY, Helsinki. PDF-dokumentti. Saatavissa: <https://www.aalto.fi/sites/g/files/flghsv161/files/2020-08/Projekttiliiketoiminta.pdf> [viitattu 14.11.2022].

Lean Enterprise Institute. 2022. Plan, Do, Check, Act (PDCA). WWW-dokumentti. Saatavissa: <https://www.lean.org/lexicon-terms/pdca/> [viitattu 15.11.2022].

Mäntyneva, M. 2016. Hallittu projekti: Jäntevästä suunnittelusta menestykselliseen toteutukseen. Helsinki: Kauppakamari.

SFS-ISO 10006. 2018. Laadunhallinta. Ohjeistusta laadunhallintaan projekteissa.

SFS-ISO 21502. 2021. Projektin-, ohjelman- ja salkunhallinta. Ohjeita projektinhallintaan.

SFS-EN ISO 9004. 2018. Laadunhallinta. Organisaation laatu. Ohjeita jatkuvan menestyksen saavuttamiseen.

# Digitoinnin laatu

SARI JÄRN JA  
HENNA RISTOLAINEN

*Digitointi vaatii huolellista etukäteissuunnittelua jo hyvissä ajoin, ennen kuin varsinaiseen työhön ryhdytään. Suunnittelulla varmistetaan digitoinnin laadukas toteuttaminen ja aineiston entistä parempi käytettävyys asiakkaalle.*

*Laadunvarmistuksen tarkoituksena on huolehtia etenkin siitä, että digitoinnin lopputuloksena syntyy suunnitelman mukaista, tasalaatuaista digitoitua aineistoa. Ennen digitointiprosessia onkin syytä tehdä laaduntarkkailusuunnitelma sekä määrittellä ne kriteerit, joihin laaduntarkkailussa kiinnitetään erityistä huomiota.*

*Laatua tulee tarkkailla johdonmukaisesti suunnitelman mukaan. On myös tärkeää, että työntekijät tietävät vastualueensa ja heidät on perehdytetty asianmukaisesti.*

## Laatua varmistetaan koko prosessin aikana

Suunnittelussa laatu tarkoittaa sitä, että määritellään tarkasti lakien ja ohjeiden vaikutukset digitoinnin toteutukseen. Selvitetään, onko digitoinnin toteuttajalla riittävät ja ajantasaiset ohjeet. Suunnitelmaan kirjataan projektin ja aineiston laatukriteerit. Kriteereihin vaikuttaa digitoinnin päämäärä: digitoidaanko aineistoa pitkäaikaissäilytykseen tai vaikkapa yhteen painettavaan julkaisuun.

Jos kotiseutumuseon yksi valokuva halutaan saada yhdistyksen painettuun historiikkiin, käytetään erilaisia laatuvaatimuksia kuin pysyvästi säilytettävän viranomaisaineiston digitoinnissa.

Resurssien oikein kohdentaminen on laadun tae. Onko voitu hankkia laadukkaat laitteet digitointia varten? Onko työntekijät koulutettu ja ohjeistettu vaadittavalla tavalla?

Digitointiprosessia voidaan varmistaa selkeillä rooleilla ja vastuilla. Kun digitointiprosessi on jaettu riittävän pieniin osiin, voidaan tarkasti määrittellä kunkin vaiheen laatuun vaikuttavat tekijät.

Esimerkiksi digitoinnin valmisteluun pitää käyttää riittävästi aikaa, jotta aineistot tai laitteistot eivät vaurioidu tai peräti tuhoudu digitoinnin aikana. Laadukas valmistelu on tärkeässä osassa suunnitellun lopputuloksen kannalta.

Laadun vähimmäismäärittelyä suunniteltaessa on huomiotava, voidaanko määrittelyä noudattaa. Kotiseutumuseolla ei ole samoja mahdollisuuksia noudattaa kansallisia määräyksiä sähköisestä pitkäaikaissäilyttämisestä kuin viranomaisella. Laadun määrittelyn tulee aina tukea digitoinnin päämäärää.

Laadusta tulee huolehtia prosessin koko elinkaaren ajan. Laadukkaasti digitoitu aineisto ei ole kokonaisvaltaisesti laadukas, jos se ei esimerkiksi ole päämäärän ja suunnitelman mukaisesti käytettävissä.

Säännöllisyys on avainsana laadun varmistamisessa. Prosessin suorituskykyä on seurattava ja arvioitava säännöllisesti. Näin voidaan todentaa, onko laatuvaatimuksia noudatettu ja ovatko ne toteutuneet.



11

**Laadun määrittelyn tulee aina tukea digitoinnin päämäärää.**

## Muista seurata tuloksia

Digitointiprosessin tulosten ja aineiston laadun seuranta voidaan tehdä monella tavalla. Laadua voidaan tarkistaa silmämääräisesti tai ohjelmallisesti. Tekniikka kehittyy jatkuvasti, mutta silti tarvitaan vielä ihmissilmää varmistamaan lopputulos.

Laadua voidaan aina parantaa, joten mahdollisten virhelähteiden poistaminen kuuluu prosessiin. Laatu vaatii oman aikataulunsa. On varattava riittävästi resursseja ja aikaa, jotta jokainen digitointiprosessin vaihe voidaan tehdä huolella.

Digitointityötä tekeville yrityksillä ja organisaatioilla on erilaisia laatukriteereitä. Laadukkaan digitoinnin kehittäminen vaatii yhteistyötä eri toimijoiden välillä. Oma laatumäärittelyä laadittaessa kannattaa käydä läpi myös muiden organisaatioiden tapoja toimia.

Laatuun kuuluu olennaisena kaikenlainen dokumentointi. Laadukasta digitointia voidaan parhaiten toistaa, kun prosessi on kuvattu riittävällä tarkkuudella. Näin laatukriteereitä voi tarvittaessa esittää myös muille.

*Digitoinnin laatua ei voi koskaan korostaa liikaa. Laadukkaasti digitoitu aineisto palvelee sekä käyttäjää että digitoinnin tekijää.*

*Digitointi on prosessi, jolle on määritelty tarkat vaatimukset, jotta lopputulos olisi laadukas ja tarvittaessa pitkäaikaissäilytykseen kelpaava, arkistokelpoinen ja asiakkaiden helposti käytettävissä.*

*Käyttötarkoitus ja määräykset määrittävät digitoitun lopputuloksen laadun.*



**Kuva 8.** Laadunhallinta digitointiprojektissa (mukaiillen Artto ym. J. 2006, SFS-EN ISO 9004 & SFS-ISO 10006)

## Laadunhallinta hankinnoissa

Onnistunut ja korkealaatuinen hankinta on suunniteltu hyvin ja valmisteluun on käytetty riittävästi aikaa. Seuraavassa on esitetty korkealaatuisen hankinnan osa-alueita.

### Valmistelu

- Hankintaosaaminen, -resurssit ja -aikataulutus
- Määräyksiin ja ohjeisiin perehtyminen
- Ideointi, tarpeen tunnistaminen ja mahdollisuuksien tunnistaminen ja rajaaminen
- Markkinavuoropuhelujen käyminen toimittajien kanssa
- Vaatimusten määrittely ja dokumentointi
- Laatusuunnitelman vaatimien toimittajalta
- Ulkoisten toimittajien kartoitus ja arviointi
- Kilpailutuskriteerien määrittely
- Digitointiprojektin kuvaus
- Hankinnan kohteen kuvaus

## Hankinta ja sopiminen

- Tarjouskilpailutuksen avaaminen tai tarjouspyynnöt
- Tarjousten vastaanottaminen
- Tarjousten analyysi ja vertailu
- Toimittajien valinta
- Toimittajan laadunhallintajärjestelmän ja laatusuunnitelman vaatimustenmukaisuuden arviointi
- Sopimusprosessin mukaisen palvelusopimuksen teko

### Tilaus

- Sopimuksen toteutumisen valvonta ja laadunvarmistus
- Suorituskyvyn säännöllinen seuranta ja arviointi
- Vaatimusmäärittelyjen toteutumisen seuranta
- Tulosten seuranta
- Kriteerienmukaisuuden arviointi
- Virhelähteiden poistaminen
- Sopimusehtojen täyttämisen arviointi ennen sopimuksen päättämistä

### Hyväksyntä

- Lopputuloksen hyväksyminen
- Sopimusehtojen täyttämisen arviointi ennen sopimuksen päättämistä

### Lähteet:

Artto K., Martinsuo M. & Kujala J. 2006. Projekttiliiketoiminta. WSOY, Helsinki. PDF-dokumentti. Saatavissa: <https://www.aalto.fi/sites/g/files/flghsv161/files/2020-08/Projekttiliiketoiminta.pdf> [viitattu 14.11.2022].

FAGDI 2022. Guidelines: Technical Guidelines for Digitizing Cultural Heritage Materials. WWW-dokumentti. Saatavissa: <https://www digitizationguidelines.gov/guidelines/digitize-technical.html> [viitattu 16.11.2022].

Kansallisarkisto. 2022. Digitointi. WWW-dokumentti. Saatavissa: <https://kansallisarkisto.fi/fi/viranomaisille/digitointi> [viitattu 15.11.2022].

SFS-ISO 10006. 2018. Laadunhallinta. Ohjeistusta laadunhallintaan projekteissa.

SFS-ISO 21502. 2021. Projektin-, ohjelman- ja salkunhallinta. Ohjeita projektinhallintaan.

SFS-EN ISO 9004. 2018. Laadunhallinta. Organisaation laatu. Ohjeita jatkuvan menestyksen saavuttamiseen.

**Katso liitteet  
sivuilla 62-65:  
Esimerkit hankinta-  
suunnitelmasta.**



# Osa 2

–

# Näkökulmia





# Digitointi muuttuu ja kehittyy

HENNA RISTOLAINEN

*Digitointi ei koskaan ole pelkkä projekti, vaan sillä on tavoite, joka pyritään saavuttamaan ja jonka vaikutukset ovat pitkäaikaisia.*

Arkisto-, kirjasto-, ja museosektoreiden sekä yritysten kanssa käydyt keskustelut ja heille tehdyt selvitykset ovat antaneet ainutlaatuisen mahdollisuuden kurkistaa useiden erilaisten organisaatioiden tekemään digitointiin ja niihin haasteisiin, joihin käytännössä törmätään. Jo tehdyistä tai meneillään olevista digitointiprojekteista on mahdollisuus ottaa opiksi. Seuraavissa artikkeleissa esille tulee suunnitelmallisuuden tärkeys mutta toisaalta myös uudemmat kehityskulut. Teknologioiden ja tekoälyn kehittyminen ovat avanneet uudenlaisia mahdollisuuksia prosessin kehittämiseen sekä aineistojen hyödyntämiseen ja säilyttämiseen.

## Digitoinnin lopputuloksen hyödyntäminen

Yksi tärkeimmistä prosessin vaiheista on ylipäättään päätös digitointiprojektin aloittamisesta. Saatavat hyödyt olisi hyvä olla kirikkaana mielessä. Ehkä digitoinnilla säästetään tilakustannuksissa - ehkä sillä saadaan säilymään aineistoa, joka on vaarassa tuhoutua. Digitointiprojektiin lähteminen ei kuitenkaan yleensä ole läpihuutojuttu, vaan se vaatii työtä ja aikaa toteutuakseen. Onkin

hyvä miettiä myös ne muut hyödyt, jotka prosessin myötä voidaan saavuttaa, vaikka ne eivät alussa ensisijaisia syitä olisikaan. Saataisiinko digitoinnilla ratkaistua myös muita tarpeita? Ehkä ensisijainen tavoite on ollut saada huonokuntoiset aineistot säilymään, mutta aineistoja digitoitaessa niihin voidaan tehdä sisällöntunnistus, joka mahdollistaa asiakkaan itsenäisesti tekemät tiedonhaut, mikäli aineistot avataan. Lopputuloksen hyödyntämisessä asiaa olisi hyvä katsoa monen kohderyhmän osalta ja pitkällä aikavälillä. Seuraavissa artikkeleissa kerrotaan muun muassa pitkäaikaissäilyttämisen kokemuksista Kansallisliggeriassa.

Suunnitteluun ja laadunvarmistukseen kannattaa käyttää aikaa, jotta lopputulos on myös oikeasti hyödynnettävissä. Virheet digitoinnissa voivat tuottaa vuosiksi ennakoimatonta lisätyötä tai lopputulos ei vastaa ollenkaan toivottua. On esimerkiksi tapauksia, joissa on unohdettu digitoida kaksisivuisista dokumenteista toinen sivu (Mieto 2022) tai jo digitoituun dokumenttiin on tehty digitoinnin jälkeen merkintöjä (Pelli 2022). Virheet voivat vesittää kokonaan digitoinnin hyödyt ja dokumentin todistusvoimaisuuden.

## Digitointi muutoksessa ja tekoälyn hyödyntäminen digitointiprosessissa

Erityisesti viime vuosien aikana digitointi on ottanut suuria harppauksia eteenpäin. Yleinen digitalisaatiokehitys on edesauttanut myös digitointialan kehitystä sekä tekniseltä kannalta että asenteiden muutoksen myötä.

Kansallisarkiston massadigitoinnin toiminnot ovat kansainvälisestäkin merkittävä ja kunnianhimoinen ratkaisu, joka vaikuttaa koko viranomais-sektoriin ja välillisesti myös muihin toimijoihin.

Päätös massadigitointiin lähtemisestä ei varmastikaan olisi ollut mahdollista ilman teknologioiden kehittymistä tukemaan muun muassa automaattisia prosesseja. Esimerkiksi sisällön rakenteiden, tekstin ja käsin kirjoitetun tekstin tunnistuksessa käytettävät sovellukset ovat kehittyneet.

Tekoäly on tuonut mukanaan uusia vaihtoehtoja digitointiprosessien kehittämiseksi mutta myös digitoitujen aineistojen käsittelylle ja tulkinalle. Digitointi mahdollistaa tutkimuksen yhä uusiin aineistoihin, jotka aiemmin ovat jääneet tutkimatta työläytensä vuoksi. Tutkimustarpeet ovatkin nousseet yhä tärkeämmiksi syiksi digitoinnille.

## Strategiat ja politiikat digitointityön apuna

Digitoinnin tulisi olla linjassaan organisaation strategioiden kanssa. Vielä tärkeämpää on, että digitointi olisi suunnitelmallista. Mikäli organisaatiossa on tiedossa, että yhden digitointiprojektin lisäksi on aineistoja, joita tulevaisuudessa haluttaisiin digitoida, on järkevää toteuttaa suunnitelma pitkän aikavälin digitoinnille. Tällainen politiikka tai ohjelma luo suunnitelmallisuutta sekä auttaa digitointiprojektien ja digitoitavien aineistojen priorisoinneissa. Toisaalta ennakkoselvitykset ja suunnitelmallisuus varmistavat, että noudatetaan lakeja ja ohjeita.

## Digitointi osana digitalisaatiokehitystä

Osana digitalisaatiokehitystä on muun muassa Kansallisarkiston Määräys arkistoitavien asiakirjojen muodosta (2021), mutta toisaalta myös visiot digitaalisesta tulevaisuudesta (Auvinen 2020). Korona-ajan myötä moni organisaatio otti harppauksen digitaalisuudessa. Yleistynyt etätyö ei ole yleensä mahdollista ilman digitaalisia aineistoja.

Digitoimattomat aineistot aiheuttavat, että etätyö ei ole mahdollista osalle ihmisistä. Etätyöstä on kuitenkin tullut jo monelle organisaatiolle tärkeä työnteon tapa ja valttikortti palkata osaavia työntekijöitä muilta paikkakunnilta.

Yleinen digitalisaatiokehitys on tuonut digitointia uudelleen pinnalle ja tärkeäksi osaksi digitaalisen toiminnan mahdollistamista. Opetus- ja kulttuuriministeriön (2022) muutosohjelma vuosille 2023–2026 nostaa digitoinnin ja aineistojen pitkäaikaisäilytyksen merkittäviksi digitalisaation kehittämiskohteiksi. Seuraavissa, asiantuntijoiden kirjoittamissa, artikkeleissa pääset tutustumaan käytännössä toteutettuihin ratkaisuihin, joilla digitointia kehitetään. Artikkelit tuovat hyvin esille digitoinnin osa-alueiden monipuolisuutta ja mahdollisuuksia.

### Lähteet:

Auvinen, T. 2020. Digitaalinen ja ihmisläheinen kunta 2035. Toni Auvinen Avarasti Oy. E-kirja. Saatavissa: <https://luode.cld.bz/kunta-2035/4/> [viitattu 11.11.2022].

Pelli, S. 2022. Sales Specialist. Digitointiprojektin sisältö. Luento 8.9.2022. Canon Oy.

Kansallisarkisto. 2021. Määräys arkistoitavien asiakirjojen muodosta. KA/15906/07.01.01.00/2021. PDF-dokumentti. Saatavissa: [https://kansallisarkisto.fi/uploads/Viranomaisille/M%C3%A4%C3%A4r%C3%A4ykset%20ja%20ohjeet/M%C3%A4%C3%A4r%C3%A4ys%20arkistoitavien%20asiakirjojen%20muodosta\\_22.12.2021.pdf](https://kansallisarkisto.fi/uploads/Viranomaisille/M%C3%A4%C3%A4r%C3%A4ykset%20ja%20ohjeet/M%C3%A4%C3%A4r%C3%A4ys%20arkistoitavien%20asiakirjojen%20muodosta_22.12.2021.pdf). [viitattu 11.11.2022].

Mieto, J. 2022. Kokoelmavastaava. Haastattelu 29.9.2022. Päivälehdien arkisto.

Opetus- ja kulttuuriministeriö. 2022. Sivistyshallinto 2030. Opetus- ja kulttuuriministeriön konsernin kehittämishankkeen loppuraportti. Opetus- ja kulttuuriministeriön julkaisuja 2022:35. PDF-dokumentti. Saatavissa: <http://urn.fi/URN:ISBN:978-952-263-816-8> [viitattu 15.11.2022].

# Kansalliskirjaston digitointiohjelma digitoinnin lähtökohtana

MINNA KAUKONEN

*Artikkeli on lyhennelmä Johanna Liljan, Jussi-Pekka Hakkaraisen, Pirjo Karppisen ja Pekka Lampisen julkaisusta Kansalliskirjaston digitointiohjelma 2021-2024. Koko julkaisu on luettavissa <https://urn.fi/URN:ISBN:978-951-51-6993-8>.*

*Kansalliskirjaston rikkaisiin ja monimuotoisiin aineistoihin kohdistuu runsaasti kirjaston omia digitointitoiveita sekä yhteistyökumppaneiden ja asiakkaiden toiveita. On välttämätöntä priorisoida digitoitavan aineiston valintaa, koska resurssit ovat rajalliset. Digitointiohjelman 2021-2024 valintoja on suunniteltu asiantuntijoiden, Tutkimuskirjasto-palvelualueen esihenkilöiden ja Kansalliskirjaston johdon kanssa sekä Kansalliskirjaston neuvottelukunnissa.*

## **Digitoinnin mittarit ovat määrällisiä, toiminnallisia ja eettisiä**

Kansalliskirjastossa on käytössä kriteereitä ja mittareita, joilla voidaan arvioida digitoinnin onnistumista jo digitoitavaa aineistoa valittaessa. Määrällisten mittarien, kuten lataus- ja tuotantolukujen, lisäksi on haluttu nostaa esiin myös toiminnallisia ja eettisiä mittareita. Tavoitteena on, että digitoitu aineisto tehostaisi kirjastotyötä ja palvelisi laajasti erikielisiä ja erilaisista taustoista tulevia ihmisiä. Tutkimuksen tarpeet nousevat vahvasti esiin Kansalliskirjaston strategiassa. Palvelemalla tutkimusta digitointi tuottaa välillistä hyötyä kaikille.

Digitoitavan aineiston valinnassa on otettava huomioon mahdollisuus tuottaa laadukasta aineistoa, jonka hyödyntäminen on käyttäjälle sujuvaa ja vaivatonta. Erityisesti tutkimuskäytön kannalta on olennaista, että digitoidaan ehjiä ja kattavia kokonaisuuksia. Riittävät tuotannon resurssit turvaavat parhaiten digitoinnin korkean teknisen laadun. Luotettavat kuvailutiedot varmistavat aineiston löytyvyyden ja pitkäaikais säilytyksen. Selkeät käyttöehdot sekä helppokäyttöinen ja luotettava aineistopalvelu parantavat käytettävyyttä.

## **Kansalliskirjaston digitoinnin prosessit huomioivat erilaiset aineistotyypit**

Digitointia tehdään Kansalliskirjaston omana työnä. Mikkelin toimipisteessä on laaja ja monipuolinen valikoima erilaisia skannereita eri aineistojen digitointiin. Hyväkuntoiset kirjat on mahdollista skannata lähes automaattisesti tätä varten varatulla automaattiskannerilla. Hauraalle paperi- ja irtolehtiaineistolle on käytössä manuaaliset, kirjahisseillä varustetut skannerit, joilla sivunvaihto tapahtuu käsin. Näillä skannereilla voidaan skannata myös muita vaativia aineistoja, mm. isoja karttoja.

Mikrofilmeille kuvattuja lehtiaineistoja skannataan rullafilmiskannereilla. Vapaakappaleina luovutettavien sanomalehtien vastaanotossa digitoidaan kaikki vain painettuina saatavat sanomalehdet sekä vanhempia lehtiä sanomalehtiskannerilla.



Skannattu aineisto vaatii aina myös jälkikäsittelyä ennen sen julkaisua ja pitkäaikaistallennusta. Mikkelissä jälkikäsittely tehdään erityisesti masadigitointiin suunnitellulla laitteistolla. Jälkikäsittelyn yhteydessä tehdään sivujen rajausta, tekstiosuukien suoristusta, kuvien ja tekstiosuukien automaattinen tunnistus ja tekstintunnistus (OCR). Lopputuotteena syntyy tiedostokokonaisuus, joka sisältää kaikki teoksen tai lehden julkaisuun ja pitkäaikaissäilytykseen tarvittavat tiedostot, metatiedot mukaan lukien. Digitointi vaatii useiden Kansalliskirjaston yksiköiden ja asiantuntijoiden yhteistyötä

### **Aineistoja digitoidaan perusrahoituksella ja täydentävällä rahoituksella**

Kansalliskirjaston kansalliskokoelmaan kuuluvat sanomalehdet tarjoavat ihmisistä, historiasta, kielestä, kulttuurista ja yhteiskunnasta kiinnostuneille rikkaan lähteen. Valtakunnalliset, maakunnalliset, paikalliset lehdet ja kaupunkilehdet, päivä-, yleissanoma-, erikoissanoma- sekä ilmaisjakelulehdet kertovat ajastaan eri näkökulmista. Kokoelman koko on 50 miljoonaa sivua ja se kasvaa vuosittain noin 300 sanomalehtinimekkeen sivumäärällä eli 850 000 sivulla. Kaikki Suomessa vuosina 1771–1939 ilmestyneet sanomalehdet on digitoitu. Tekijänoikeusjärjestön kanssa tehtyjen sopimusten mukaan Kansalliskirjasto on voinut avata yleisön käyttöön kaikki ennen vuotta 1940 ilmestyneet lehdet sekä tutkijakäyttöön kaikki digitoitut lehdet vuoteen 2018 asti digi.kansalliskirjasto.fi -palvelussaan. Kaikki sanomalehdet on digitoitu vuoden 2017 alusta ilmestyneistä eteenpäin. Uusin aineisto on käytössä vapaakappalekirjastoissa. Kansalliskirjasto on kuitenkin digitoinut vasta 30 % sanomalehdistä.

Vuosina 1940–2016 ilmestyneitä, vielä digitoimattomia sanomalehtiä, digitoidaan perusrahoituksella, apurahoilla ja yhteistyössä kustantajien, julkaisijoiden ja säätiöiden kanssa. Meneillään on muun muassa ruotsinkielisten lehtien suuri digitointiprojekti, jossa kaikki Suomessa ilmestyneet ruotsinkieliset sanomalehdet digitoidaan kahdeksan ruotsinkielisen säätiön ja rahaston rahoituksella. Projektin päättyessä ruotsinkielistä lehtiaineistoa on käytettävissä digitaalisena liki 6 miljoonaa sivua.

Täydentävällä rahoituksella toteutettuja kirja-aineistojen digitointikohteita ovat olleet Keisarillisen Aleksanterin yliopiston väitöskirjat ja klandestiinikäsikirjoitukset, jotka ovat 1600–1700-lukujen kiellettyä filosofista kirjallisuutta. Ruotsin ajan kansalliskokoelman digitointi, kartografian klassikoiden Ptolemaios-atlasten digitointi ja nuottikokoelman digitointi ovat myös lähiaikojen projekteja.

**Lähde:** Lilja, J., Hakkarainen, J.-P., Karppinen, P. & Lampinen, P. 2021. Kansalliskirjaston Digitointiohjelma 2021–2024. Kansalliskirjasto. E-kirja. Saatavissa: <https://urn.fi/URN:ISBN:978-951-51-6993-8> [viitattu 1.11.2022].

# Tekoälyvälineiden käyttöön- oton tavoitteet ja suden- kuopat massadigitoinnissa

VILLE KAJANNE JA MAIJU POHJOLA

*Tekoälyratkaisut voivat merkittävästi tehostaa digitointia ja tuoda sille lisäarvoa. Mitä laajamittaisemmasta digitointituotannosta on kyse, sitä suurempia ovat hyödyt. Tekoälykomponenttien avulla voidaan vähentää tai kohdentaa paremmin henkilötyötä, luoda erilaisia työkaluja aineiston tai tietojen löytämiseen sekä parantaa digitaalisen aineiston laatua. Myös pitkäaikaissäilytykseen siirrettävän datan määrää voidaan supistaa tai ainakin hallita paremmin, myös jälkikäteen, jos aineistoa on tekoälyn avulla eri kriteerein luokiteltu tai määritetty.*

## Tekoälyn hyödyt massadigitoinnissa

Tekoälykomponenttien käyttöönotolla Kansallisarkiston massadigitoinnissa tavoitellaan kolmenlaista hyötyä. Digitoinnin laadunvalvontaan ja myöhemmin pitkäaikaissäilytykseen siirtyvän kuvamassan määrää pyritään kontrolloimaan tyhjien sivujen tunnistuksella. Tyhjät sivut voidaan ohittaa virheiden tunnistuksessa (validointi), sisällöntunnistuksessa (OCR) ja muissa digitoinnin jälkiprosesseissa sekä lopulta asiakaskäyttöliittymässä (Astia). Tyhjät kuvat voidaan niin halutessa myöhemmin kokonaan poistaa.

Toiseksi halutaan laajentaa digitoinnin erilaisten virheiden automaattista tunnistusta. Tämä vähentää merkittävästi henkilötyön määrää, koska tällä hetkellä 100 % tuotetuista kuvista validoidaan ihmissilmin. Myös validoinnin tasalaatuisuus vahvistuu, koska virhetyyppejä on lukuisia ja henkilötyönä tapahtuvassa validoinnissa tehdään

aina hiukan erilaisia ratkaisuja. Kun puhutaan tuhansien kuvien läpikäynnistä päivittäin, voidaan koneoppimisratkaisujen avulla myös havaita virheet paremmin kuin ihmissilmällä.

Kolmanneksi halutaan lisätä aineiston käytettävyyttä. Automatisoidulla kuvankäännöllä lisätään sisällöntunnistusta ja luettavuutta. Aineiston käytettävyyttä ja haettavuutta pyritään lisäämään rakentamalla luokittimia asiakirjan rakenteiden tunnistamiseen, nimientiteettien tunnistamiseen (NER) eli automaattiseen metatietojen generointiin. Käytettävyyttä ja haettavuutta pyritään kehittämään myös parantamalla OCR- (tai HTR-) eli sisällöntunnistuksen luotettavuutta esimerkiksi NLP-menetelmien avulla (OCR-tunnistuksen virheiden korjaaminen) tai kehittämällä ns. sivusegmentoijaa, jonka avulla voidaan ohjata kuvia oikeisiin prosesseihin sisällön mukaan.

Tekoälyratkaisuja kehitetään ja otetaan käyttöön Kansallisarkistossa sisäisinä projekteina esimerkiksi massadigitoinnin tuotannossa tai tiettyjen aineistojen verkkopalveluina (esim. Tuomiokirjahaku), tutkimusyhteistyöhankkeina (esim. READ, FIN-CLARIAH) sekä EU-rahoitteisina kumppanuushankkeina (esim. DALAI).

## Tekstintunnistustyökaluja

- **Transcribus**  
käsinkirjoitetun tekstin tunnistus
- **Pero OCR**  
painettujen ja käsinkirjoitettujen tekstien tunnistus
- **Tesseract OCR**
- **EasyOCR**

## Tekoälyn haasteet

Tekoälysovellusten usein mainittu haaste on luokittimien opettamisessa tarvittavan aineiston laaja määrä. Tähän voi liittyä myös käytännön ongelmia, kuten datan hallinta tai tietosuojan varmistaminen ja siihen liittyvät menettelyt. Laaja aineistomäärätarve koskee kuitenkin vain komponentin kehittämistä. Sen ohella on useita niiden varsinaiseen käyttöönottoon ja ylläpitoon liittyviä pulmia. Niihin tarvittavia resursseja ja osaamista ei kenties ole riittävästi huomioitu, vaan huomio on kiinnittynyt enemmän teknologioiden tarjoamiin välineellisiin mahdollisuuksiin. Toteutusvaiheet ja näkymätön taustatyö on ikään kuin jäänyt jännittävien ja lupaavien teknologioiden varjoon.

Ennen kuin komponentteja kytketään prosesseihin, tulee varmistua niiden käsittelynopeudesta ja edellyttämästä prosessointikapasiteetista. Laadukas komponentti ei ole kovin hyödyllinen, jos se muodostaa hitautensa vuoksi pullonkaulan kokonaisprosessiin tai edellyttää kokonaistoiminnan näkökulmasta suhteetonta käsittelykapasiteettia. Tätä arvioidessa on huomioitava komponenttien modulaarisuus: useimmiten käytössä on useampia komponentteja, ja niiden määrä voi alati kasvaa.

Käyttöönotto ei siis ole vain tekniikkaa tai mustavalkoista päätöksentekoa sen perusteella, toimii ko komponentti vai ei. Tavoitetason määrittäminen voi vaihdella eri komponenteilla. Laadun tai käytettävyyden kannalta kriittiset virheet tulee tunnistaa ja asettaa tavoitetasot niissä korkeammiksi kuin muissa. Joskus taas tarjolla on hyvin käyttökelpoisia ja hyödyllisiä komponentteja, mutta niiden toimintalogiikka saattaa edellyttää muutoksia esimerkiksi kuvaformaatin vaatimusmäärittelyyn. Käyttöönotto voi pysähtyä alkuunsa, jos selkeä roolitus tai päätöksentekokyky vastuiden tai vaatimustason osalta uupuu.

Lukuisien komponenttien käyttöönotto vaatii huolellista suunnittelua. Niiden tulee seurata toisiaan prosessissa modulaarisesti ja loogisesti. Esimerkiksi tyhjien sivujen tunnistus tulee sijoitua ennen virheiden määrittystä ja sen ennen OCR-käsittelyä. Joskus yksittäisen komponentin edellyttämät prosessointivaatimukset ovat ristiriidassa toisen kanssa. Esimerkiksi kuvankäännön logiikka voi haitata yksittäistä virhetunnistusta.

Komponenttien kytkeminen prosessiin aiheuttaa muutostarpeita sovelluksiin ja sen myötä uusia toimenpiteitä. Jos API tunnistaa virheellisessä lukusuunnassa olevan kuvan, tulee selvittää, miten ja missä vaiheessa se käännetään oikeaan suuntaan. Käyttöönotosta voi seurata suuri määrä kehitystarpeita varsinaisiin digitointiprosessin sovellusten toimintoihin ja toiminnallisuuksiin.

## Seuranta ja jatkokehittäminen on yhtä tärkeää kuin käyttöönotto

Komponenttien toimivuuteen pitää uskoa, mutta samalla sitä pitää jatkuvasti seurata. Jos komponentin käyttöönotolla halutaan aidosti vähentää työmääriä, tulee sen toimivuuteen luottaa. Tämä voi tarkoittaa virheiden tai puutteiden päättymistä loppukäyttäjille tai -säilytykseen. Komponentit voivat myös alkuvaiheessa toimia huonommin, mutta toimivuuden arvioimiseksi niiden vieminen tuotantoon kannattaa silti, koska se mahdollistaa jatkokehityksen. Tämä voi hetkellisesti tarkoittaa jopa kokonaisprosessin työmäärän kasvua. Toisaalta erittäin hyvin toimiva komponentti voi alkaa toimia huonommin esimerkiksi uuden, erilaisen aineiston vuoksi tai algoritmin taustalla vaikuttavan syväoppimisen kirjaston versio vanhentumisen vuoksi. Siksi komponenttien toimivuutta pitää jatkuvasti seurata vakioituilla mittareilla ja kerätä opetusaineistoa niiden päivittämistä varten.

### Lähteet:

European Association for Digital Humanities s.a. READ. WWW-dokumentti. Saatavissa: <http://read.transkribus.eu/> [viitattu 9.11.2022].

Kansallisarkisto. 2021. DALAI - Digitaalisten aineistojen laadun ja käytettävyyden parantaminen tekoälyavusteisesti. WWW-dokumentti. Saatavissa: <https://kansallisarkisto.fi/dalai> [viitattu 9.11.2022].

Kansallisarkisto. 2022. Fin-Clariah. Saatavissa: <https://www.kansalliskirjasto.fi/fi/projektit/fin-clariah> [viitattu 9.11.2022].

Kansallisarkisto s.a. Tuomiokirjahaku. WWW-dokumentti. Saatavissa: <https://tuomiokirjat.kansallisarkisto.fi/> [viitattu 9.11.2022].

# Digitoinnissa keinoäly tunnistaa tyhjät sivut

ANSSI JÄÄSKELÄINEN

*Keinoäly eli tekoäly on tehnyt tuloaan teoriassa jo pitkään. Ensimmäiset akateemiset maininnat löytyvät vuodelta 1955, jolloin John McCarthy esitteli termin artificial intelligence (McCarthy ym. 1955, 10). Kuitenkin vasta hiljattain teknologian kypsyys ja etenkin sen helpohko hyödynnettävyys ovat mahdollistaneet käytännönläheiset keinoälyratkaisut ja niiden laajamittaisen hyödyntämisen.*

## Keinoälyn mahdollisuudet

Keinoälyyn perehtymättömille ihmisille keinoäly on edelleen jotakin mystistä, mahdollisesti Terminator, I Robot tai 2001: Avaruusseikkailu -elokuvi-ta tuttua. Keinoäly esitetään jonakin yli-inhimillisenä sekä ympäristöstään tietoisena ja oppivana toimijana. Vaikka todellisuus on kaukana tästä, tosiasia on, että esimerkiksi AlphaZero-keinoäly voitti maailman parhaan sääntöihin perustuvan shakkiohjelman jo vuonna 2017 opeteltuaan peliä vain 24 tuntia (Silver ym. 2017). Kansalaiset hyödyntävät kuitenkin keinoälyratkaisuja täysin tietämättään. Keinoälyratkaisuja ovat muun muassa Applen Siri -puhehaku, pienempien luottohakemuksien käsittelyt, semiautonominiset autot sekä Facebookissa käytetty henkilöiden tunnistus.

Teoriassa mikä tahansa matemaattinen tai matemaattiseksi muutettavissa oleva ongelma on ratkaistavissa keinoälyn avulla. GPU-korttien eli näytönohjainten kehittyminen riittävän tehokkaiksi on mahdollistanut tämän. Siinä missä CPU:ssa eli prosessorissa voidaan ajaa maksimissaan kymmeniä rinnakkaisia operaatioita, näytönohjaimen ytimissä näitä voidaan ajaa tuhansia rinnakkain. Saavutettava teoreettinen nopeus-hyöty on valtava.

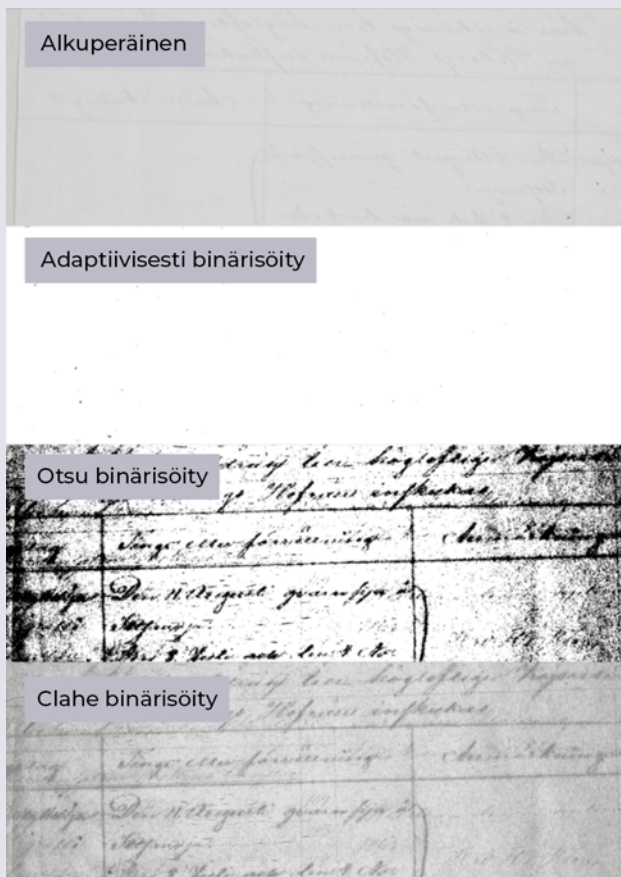
## Keinoälyn opettamisen prosessi

Digitalia, digitaalisen tiedonhallinnan tutkimus- ja kehittämiskeskus, on hyödyntänyt keinoälyä tyhjien sivujen tunnistamisessa yhteistyössä Kansallisarkiston kanssa sekä tekstisisällön automaattisessa asiasanoittamisessa. Keinoälyn näkökulmasta prosessi etenee yleisellä tasolla seuraavasti.

1. Hanki riittävä määrä opetusaineistoa tai käytä valmiiksi opetettua mallia.
2. Valitse keinoälymalli ja opeta se aineistolla.
3. Testaa testiaineistolla.
4. Jos tulos hyväksyttävä, siirry eteenpäin, muuten opeta uudelleen tai vaihda keinoälymallia.
5. Käytä oikealla aineistolla.

”

*Jos keinoälyratkaisut ja niiden mahdollinen hyödyntäminen kiinnostaa, ole yhteydessä Digitaliaan ([www.digitalia.fi](http://www.digitalia.fi)).*



**Kuva 9.** Esimerkki kuvan esikäsittelyn merkityksestä ennen keinoälyn hyödyntämistä. Kuva: Anssi Jääskeläinen.

Kuten kuvasta 9 voi havaita, kuvan esikäsittelyn merkitystä ennen keinoälyn tekemää päätöstä ei voi aliarvioida. Sivun tunnistuisi tyhjäksi sekä alkuperäisenä että adaptiivisen binärisöinnin jälkeen, mutta Otsu- ja Clahe -binärisöintien jälkeen keinoäly tunnistaisi sivut tekstiä sisältäviksi.

## Tyhjien sivujen tunnistaminen keinoälyn avulla

Tyhjien sivujen tunnistamisen testissä aineistona käytettiin alkuperäisiä käsittelemättömiä kuvatedostoja. Keinoälymallit oli opetettu sekä noin 2 000 sivun että noin 30 000 sivun opetusaineistoilla. Opetusaineiston koolla on merkitystä tiettyyn pisteeseen asti. Kun opetusaineiston kasvaa, kasvaa myös sen koostamiseen kulunut aika.

Isommalla opetusaineistolla päästiin erinomaiseen tulokseen, kun painotettiin sitä, että dataa ei tunnistuisi tyhjäksi. Testiaineiston kokonaismäärä oli 17 359 sivua, joista 99,879 prosenttia tunnistettiin tekstiä sisältäviksi ja vain 21 tekstiä sisältävää sivua tunnistui tyhjiksi. Toiseen suuntaan tark-

kuus kärsi edellä mainitun painotuksen vuoksi, eli 11 928 tyhjää sivua, joista 90,979 prosenttia tunnistui tyhjiksi. Tietoa sisältäviksi tunnistui siis reilut 1000 sivua. Pienemmällä opetusaineistoilla opetetuilla malleilla noin 80 prosenttia sivuista tunnistui oikein molempiin suuntiin.

## Digitalian ja Memory Labin uudet laitteet tukevat keinoälyn hyödyntämistä

Tutkittaessa keinoälyratkaisuja pelkästään teknisestä näkökulmasta, maailman ehdotonta huipua edustavat Nvidian DGX-A100- ja H100 -laskentayksiköt. Digitaliassa ja Memory Labissa ollaan siirtymässä kuusi vuotta vanhasta palvelinkannasta ja näihin istutetusta vanhasta GPU-kannasta uuteen laitekantaan. Tämän hankinnan yhteydessä hankitaan DGX-A100- tai sitä vastaava GPU-laskentayksikkö. Näin ollen Digitalia ja Memory Lab ottavat suuren harppauksen kohti keinoälykästä ihmeellistä maailmaa .

### Testaustulokset

Opetusaineisto 29287 sivua, data ei tunnistuisi tyhjäksi -painotus

17359 sivua sisälsi tekstiä	99,879% sisälsi tekstiä, 21 tyhjää sivua tunnistui
11928 tyhjää sivua	90,979% tyhjää sivua, 1076 tekstiä sisältävää sivua tunnistui

### Lähteet:

McCarthy, J. Minsky, M. L. Rochester, N. & Shannon, C.E. 1955. A proposal for the Dartmouth summer research project on artificial intelligence. PDF-dokumentti. Saatavissa: <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf> [viitattu 1.11.2022].

Silver, D. Hubert, T. ym. 2017. Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm. ArXiv:1712.01815v1 [cs.AI] 5 Dec 2017. WWW-dokumentti. Saatavissa: <https://doi.org/10.48550/arXiv.1712.01815> [viitattu 1.11.2022].



# Kokemuksia tiedostojen PAS-paketoinnista

KARI PEIPONEN

*Kansallisgalleriassa on paketoitu mediataiteen teoksia Kulttuuriperintö-PAS-palvelun pitkäaikaissäilytykseen vuodesta 2020 lähtien. Käytännön tasolla on kyse erilaisten tiedostojen keräämisestä ja paketoinnista yhteen pakettiin, joka voidaan lähettää säilytykseen. Tämä artikkeli kuvaa mediataideteosten pakettiin liittyviä tehtäviä ja haasteita.*

## Paketoitava aineisto

Kansallisgallerian kokoelmissa on mediataiteen teoksia jopa vuodelta 1955. Vuosien varrella teoksia on muunnettu vanhoista magneettinauhaan perustuvista medioista tietokoneen tiedostoiksi. Suurin osa teoksista sisältää video-, ääni-, kuva- sekä tekstitiedostoja. Tärkeä osa teosten tallentamisessa ovat sopimukset ja ohjeet, joiden avulla teos voidaan säilyttää ja palauttaa uudelleen esityskuntoon (Sääskilahti 2022).

Nykyisin mediataiteen teokset voivat muodostua monimutkaisista ohjelmistoista ja myös tekoälyä hyödyntävistä sovelluksista. Jos tavoitteena on teoksen tallentaminen ja mahdollisesti uudelleen esittäminen vuosien kuluttua, voi tehtävä osoittautua hyvin haasteelliseksi. Tämä johtuu siitä, että organisaatiolla ei ole tarpeeksi resursseja kaikkien teknisten yksityiskohtien selvittämiseen ja tarvittavien tiedostojen sekä teknisten alus- teosten tallentamiseen. Ehkäpä kaikkein monimutkaisimpien teosten osalta täytyy resurssien puuttuessa tyytyä vain teoksen mahdollisimman hyvään dokumentoimiseen.

## Tehtävän rajaus ja toiminnan organisointi

Pitkäaikaissäilytykseen hyväksytään vain siirto- ja säilytyskelpoisia tiedostomuotoja. Tiedostomuodot on kuvattu kattavasti Säilytys- ja siirtokelpoiset tiedostomuodot -dokumentissa (Opetus- ja kulttuuriministeriö 2022a). Lisäksi tiedostojen täytyy läpäistä PAS-palveluun rakennettu tiedostojen validointiprosessi.

Yksi mediataideteos tiedostoitteineen muodostaa yhden objektin. Keskeinen osa pakettia on METS-tiedosto, joka sisältää objektiin liittyvät tekniset ja kuvailevat metatiedot XML-formaatissa (Opetus- ja kulttuuriministeriö 2022b).

**Kuva 10.**  
Mediataiteen PAS-paketoinnin tehtävät ja projektityöskentelyn roolit

1

### PROJEKTI-PÄÄLLIKKÖ

- Toiminnan organisointi
- Säännöllisten tilannepalaverien järjestäminen

2

### PAKETOIJA

- Videoiden normalisointi säilytyskelpoiseen muotoon
- Paketoinnin tekeminen ja siirto PAS-säilytykseen

3


### VIDEOEDITOIJJA

- Alkuperäisten tiedostojen etsiminen
- Tiedostojen palauttaminen nauhoilta

4

### TEKNINEN ASIAINTUNTIJA

- Tekninen tuki ja paketoinnin vianselvitykset
- Mediahillo ohjelman ylläpito ja muutokset ohjelmaan



Muut paketoituvaiheessa käsiteltävät ja normaali-soitavat tai suoraan liitettävät tiedostot ovat:

- Teokseen liittyvät metatiedot kokoelmahallintajärjestelmästä API-rajapinnan kautta luettuna.
- Teosta kuvailevat metatiedot museolle suunnitellussa Lido-XML-skeemassa (Cidoc 2010-2021).
- Teokseen sisältyvät ohjeet ja sopimukset pääosin PDF-tiedostoina.
- Alkuperäiset video- tai äänitiedostot yleensä MPEG, MOV, MP4 tai MP3 säiliömuodoissa.
- Teokseen liittyvät kuvatiedostot pääosin TIFF, JPEG tai PNG tiedostoina.

Kansallisgalleriassa mediataiteen teosten paketoitua pitkäaikaissäilytykseen on toteutettu projekteissa. Toiminnan organisointi projektissa on tärkeää, ja mediataiteen paketoituprojektin erilaiset roolit sekä tehtävät on esitetty kuvassa 10.

## Tarvittavat työkalut ja paketoinnin vaiheet

Tärkeä työkalu on varsinaisen PAS-paketoinnin suorittava ohjelma. CSC Tieteen tietotekniikan keskus on kehittänyt tehtävää varten ”Pre-Ingest Tool” -nimisen työkalun, joka otettiin käyttöön (The Finnish National Digital Preservation Services 2022). Sen päälle koodattiin itse käyttöliittymä, jonka avulla voidaan suorittaa paketointiin liittyviä komentoja verkkoselaimella. Kansallisgallerian käyttöliittymän nimeksi annettiin Mediahillo, ja se sisältää myös tietokannan, johon kerätään onnistuneiden paketoitien tiedot. Työkalulla voidaan myös käynnistää PAS-palveluun tallennettujen paketoitien palautus.

**Kuva 11.**  
PAS-paketoinnin vaiheet

### VALMISTELU ▶

- Tiedostojen kerääminen (MPEG, MOV, MP4, TIFF, JPEG, PDF,...)
- Normalisointi = videoiden muunnos FFV1/Matroska muotoon FFMPEG ohjelmalla
- Apple ProRes videot ovat erikoistapaus

Suuria tiedostoja

### PAKETOINTI ▶

- Mediahillo sovelluksella
- Tiedostojen on oltava siirto- tai säilytyskelpoisia
- Metatietojen (XML) liittäminen MuseumPlus API-rajapinnasta
- Apple ProRes videot liitetään natiivina (vain bittitason säilytys)

### SIIRTO ◀◀ SÄILYTYS

- Pakettien siirto SFTP-ohjelmalla
- Validointi OK > paketti säilytykseen
- Validointi EI OK > virheiden selvitys ja uudelleen paketointi
- **PAS tuki**  
**pas-support@csc.fi**
- **Yhteistyö!**

Videotiedostot sisältävät erillisen videon ja äänen osuuden omissa formaateissaan yhden säiliömuodon sisällä. Tiedostopääte riippuu säiliömuodosta, kuten esim. MOV, MKV tai MP4. Taiteilijat käyttävät usein Apple ProRes -formaatteja videoissa, ja ne ovat olleet merkittävä tekijä Kansallisgallerian kontekstissa. Apple ProRes on kuitenkin ns. suljettu formaatti, josta ei ole saatavilla kattavaa avointa dokumentaatiota. Tämän vuoksi Apple ProRes -koodatut videot eivät ole säilytys- tai siirtokelpoisten tiedostoformaattien joukossa (Opetus- ja kulttuuriministeriö 2022b).

Videoiden osalta teimme ratkaisun kaikkien videoiden muuntamisesta säilytyskelpoiseen muotoon, ja tätä prosessia kutsutaan PAS-palvelussa normalisoinniksi. Säilytyskelpoiseksi muodoksi valitsimme FFV1-formaatissa olevan videon Matroska-säiliömuodossa. FFV1 on avoimen lähdekoodin FFMpeg-projektin vuodesta 2003 alkaen kehittämä häviötön pakkausmenetelmä elävälle kuvalle (Opetus- ja kulttuuriministeriö 2022b). Ääniraidat tallensimme saman säiliömuodon sisälle FLAC- tai LPCM-formaatissa. Videoiden normalisointi FFV1-formaattiin Matroska-säiliössä toteutettiin avoimen lähdekoodin FFMpeg-ohjelmalla (Bellard 2022).

Kulttuuriperintö-PAS-palvelun kanssa yhteistyössä saimme paketoitutyökaluun mukaan myös natiivitiedoston tallentamisen mahdollisuuden. Jos toimitettavassa paketissa on mukana säilytyskelpoinen videotiedosto Apple ProRes -formaatissa olevasta videosta, niin mukaan saa siinä tapauksessa toimittaa myös natiivimuotoisen tiedoston, vaikka se ei olisi siirto- tai säilytyskelpoinen tiedostomuoto. Liitimme Apple ProRes -formaatissa olevat videot mukaan pakettiin natiivimuodossaan.

Paketoinnin vaiheet ja työn aikana tehtävät toimenpiteet on esitetty kuvassa 11.



## Paketoinnin haasteet

Ensimmäinen haaste työssä oli paketointityökalun toimintakuntoon saattaminen. CSC:n kehittämä paketointityökalu toimii Unix-käyttöjärjestelmässä, ja tarvittavien riippuvuuksien selvittäminen sekä asentaminen vaatii tietoteknistä osaamista. Käyttöjärjestelmän sekä riippuvuuksien asennukseen liittyvää dokumentaatiota on vähän, ja ongelmien selvittäminen vaatii sitkeyttä. Sen jälkeen, kun perustyökalut oli saatu toimimaan, piti kehittää vielä oma käyttöliittymä komentojen käyttämiseksi.

Paketoitavien tiedostojen nimet aiheuttivat ongelmia työkalujen käytössä, ja välilyönnit, skandinaaviset merkit sekä erikoismerkit täytyi poistaa tiedostonimistä. Tämä johtui eri kerroksissa olevien ohjelmien yhteen toimivuuden haasteista Unix-käyttöjärjestelmässä. Ongelman voisi ehkä kiertää koodaamalla työkaluja enemmän, mutta se ei ollut mahdollista tehtävän työläyden vuoksi. Tiedostojen nimeämisessä kannattaakin noudattaa hyviä käytäntöjä, ja PAS-palveluun on tehty myös erillinen ohje tiedostojen nimeämisestä (Lehtonen ym. 2022).

Erilaisten tiedostojen paketoinnissa kohtasimme tiedostojen validointiin liittyviä ongelmia. PAS-palvelussa validointiin käytetään muun muassa JHOVE-ohjelmistoa (Open preservation foundation 2020), joka tutkii palveluun lähetettyjä tiedostoja. Käyttäjän näkökulmasta se on "musta laatikko", joka hyväksyy tai hylkää tiedostot. Ainoa vaihtoehto on korjata tai muuntaa hylätyt tiedostot hyväksyttävään muotoon.

Kuvatiedostojen käsittelyssä tuli esiin yllättävä ongelma. Totesimme, että vanhemmilla Photoshop ohjelmilla tallennetut TIFF-tiedostot eivät mene läpi PAS-palvelun validoinnista. Tarkemman tut-

kimisen jälkeen ongelman syyksi paljastui ennen vuotta 2018 käytössä olleen Photoshop-ohjelman virheellinen toiminta (Peterson 2018). Vanhemmilla versioilla tallennetuissa TIFF-tiedostoissa on virheellisellä tavalla muotoiltua metatietoa. Vika on korjattu helmikuun 2018 Photoshopin päivityksessä. Ongelman ratkaisu on se, että viallinen tiedosto pitää avata uudemmassa Photoshop-ohjelmassa ja tallentaa se takaisin samalla nimellä. Ongelman korjaus on sinänsä helposti tehtävissä, mutta työlääksi korjauksen tekeminen muodostuu, kun palvelimella on tuhansia viallisia kuvatiedostoja lukuisissa eri hakemistoissa.

PDF-tiedostojen validointiin liittyvät ongelmat aiheuttivat myös paljon työtä. Useilla ohjelmilla on mahdollista luoda PDF-tiedostoja, mutta ne eivät välttämättä ole kelvollisia JHOVE-validointiohjelman tarkistuksessa. Verkosta löytyy myös palveluja, joissa voi tarkistaa, onko tiedosto oikein muodostettu. Löysimme PDF-XChange Editor-ohjelman maksullisesta versiosta toiminnon, jolla pystyimme siivoamaan ja puhdistamaan tiedostot hyväksyttävään muotoon (Tracker Software Products 2022). Dokumentin voi tallentaa ohjelmalla "Sanitize Document" -toiminnolla, joka korjasi meidän tapauksissamme virheet tiedostoissa.

FFV1-koodattujen Matroska-säiliössä olevien videoiden paketoinnissa tuli vastaan useita ongelmia, mutta se kuului asiaan, koska vastaavia paketoitteja ei ollut aikaisemmin tehty. Teimme siis jonkinlaista pioneerityötä. Ongelmia ratkottiin yhteistyössä PAS-palvelun kanssa, ja ne onnistuttiin ratkaisemaan projektin edetessä.



”

*Digitaalisen aineiston säilyttäminen pitkäaikaisesti vaatii jatkuvasti päivitettävää ja ylläpidettävää teknologiaa sekä resursseja, jos se tehdään omin voimin.*

## PAS -palvelun hyödyt

Siirrettäessä aineistojen pitkäaikaissäilytykseen organisaatio joutuu selvittämään perinpohjaisesti olemassa olevan aineiston ja siihen kuuluvat tiedostomuodot sekä formaatit. Aineiston läpikäyminen ja korjaaminen voi olla työlästä, mutta tiedostojen säilyvyyden kannalta myös välttämätöntä työtä. Aineiston laatu paranee työn edetessä, kun pitkäaikaissäilytyksen prosessi on aloitettu.

Aineiston laadun parantamisen tehtävä lähtee perusasioista, kuten tiedostojen nimeämisen ja tiedostomuotojen selvittämisestä. Mahdollisuuksien mukaan tulisi selvittää tiedostojen oikeamuotoisuus validoimalla ja luoda prosesseja tarvittavien korjaavien toimenpiteiden tekemiseen.

Digitaalisen aineiston säilyttäminen pitkäaikaisesti vaatii jatkuvasti päivitettävää ja ylläpidettävää teknologiaa sekä resursseja, jos se tehdään omin voimin. Pitkäaikaissäilytyksessä kymmenien vuosien aikajänne on lyhyt ajanjakso, jolloin organisaation tulisi pystyä hoitamaan tekniset toimenpiteet jatkuvalla periaatteella sekä sitoutumaan prosessien ja tekniseen infrastruktuuriin ylläpitoon.

Keskitetty ja kansallisesti hoidettu palvelu tuntuu paljon paremmalta vaihtoehdolta verrattuna organisaatiokohtaiseen toteutukseen. Silloin organisaation ei itse tarvitse huolehtia säilyttämisen teknisestä osuudesta aineiston siirron jälkeen, ja tarvittaessa aineisto on palautettavissa takaisin säilytyksestä.

### Lähteet:

Bellard, F. 2022. FFmpeg. <https://bellard.org/>. WWW-dokumentti. Saatavissa: <https://ffmpeg.org/about.html> [viitattu: 23.10.2022].

Cidoc. 2010-2021. LIDO Schema. ICOM International Committee for Documentation. WWW-dokumentti. Saatavissa: <https://cidoc.mini.icom.museum/workinggroups/lido/lido-overview/lido-schema/> [viitattu: 24.10.2022].

Lehtonen, J., Parland-von Essen, J., Pousi, S., Hindersson-Söderholm, T. & Rauste, P. 2022. Manage well and get preserved – 6. Managing files and file naming. CSC - IT Center for Science. 27.1.2022. Video. Saatavissa: [https://video.csc.fi/media/t/0\\_tr04kqy2](https://video.csc.fi/media/t/0_tr04kqy2) [viitattu: 24.10.2022].

Open preservation foundation. 2020. JHOVE open source file format identification, validation & characterisation. 9. 6 2020. WWW-dokumentti. Saatavissa: <https://jhove.openpreservation.org/> [viitattu: 21.10.2022].

Opetus- ja kulttuuriministeriö. 2022a. Säilytys- ja siirtokelpoiset tiedostomuodot. WWW-dokumentti. CSC – tieteen tietotekniikan keskus oy, 27.2.2022. Saatavissa: <https://www.digitalpreservation.fi/specifications/fileformats> [viitattu: 20.10.2022].

Opetus- ja kulttuuriministeriö. 2022b. Aineistojen ja niiden metatietojen paketointi pitkäaikaissäilytykseen. CSC - tieteen tietotekniikan keskus, 27.2.2022. WWW-dokumentti. Saatavissa: <https://www.digitalpreservation.fi/specifications/metadata> [viitattu: 20.10.2022].

Peterson, D. 2018. Adobe fixes bug uncovered by JHOVE. 1.3.2018. WWW-dokumentti. Saatavissa: <https://dtculturalheritage.com/adobe-fixes-bug-uncovered-by-jhove/> [viitattu: 21.10.2022].

Sääskilahti, S. 2022. Mediataiteen teosten säilyttämisestä, osa ii. 5.9.2022. Saatavissa: WWW-dokumentti. <https://www.digitalpreservation.fi/2022-mediataiteen-teosten-sailyttamisesta-osa-ii> [viitattu: 20.10.2022].

The Finnish National Digital Preservation Services. 2022. Pre-ingest tool. WWW-dokumentti. Saatavissa: <https://github.com/digital-preservation-finland/dpres-siptools> [viitattu: 23.10.2022].

Tracker Software Products. 2022. PDF-XChange by Tracker Software. WWW-dokumentti. Saatavissa: <https://www.tracker-software.com/product/pdf-xchange-editor> [viitattu: 23.10.2022].

# Digitointikoulutus kehittyä palvelemaan digitointitarpeita

HENNA RISTOLAINEN

*Digitalisaation kehityksen myötä digitoinnille on kysyntää, mutta samalla digitoinnin osaamistarve kasvaa.*

Yhteiskunnan digitalisaatiokehitys, tutkimustarpeet, aineistojen säilyvyyden takaaminen ja kansalaisten palvelujen kehittäminen ovat vaikuttaneet tarpeeseen saada analogisia aineistoja digitaalisiksi. Digitaalisilla aineistoilla palvelemaan enenevässä määrin asiakkaita mutta edelleen myös organisaation sisäisiä tarpeita. Tekoäly mahdollistaa tiedon yhdistelemisen uudella tavalla ja suurien massojen käsittelyn, mikä mahdollistaa myös uudenlaisen tutkimuksen teon. Tekoälyn hyödyntäminen vaatii analogisten aineistojen digitointia, ja toisaalta tekoälyä voidaan hyödyntää digitointiprosesseissa. Molemmat vaativat sekä ymmärrystä että osaamista.

## Koulutustarjonta Suomessa ja joissakin muissa maissa

Suomessa ei ole vakiintuneita opintopolkua digitointiosaamisen saamiseen. Tutkintoon johtava koulutus on yksi tapa kerryttää digitoinnin osaamista. Ammattiopistojen tarjottimella on Media-alan perustutkintoon kuuluva Tallennetuotannossa työskentelyn (30 osp) tutkinnon osa (Opetushallitus 2021), mutta sitä suorittaneita on hyvin vähän hankkeissa tekemiemme kartoitusten perusteella. Suurin ongelma lienee koulutuksen tunnettuudessa ja siinä, että se mielletäisiin digitointityöhön valmistavaksi koulutukseksi. Koulutuksen tunnettuus vaatisi yhteistyötä digitointiosaajia tarvitsevien organisaatioiden ja ammattiopiston välillä ja toisaalta kouluttajilta digitointiosaamista. Suomessa Jyväskylän yliopisto on tarjonnut Digitaaliset arkistoaineistot (5 op) -opintojaksoa (Jyväskylä yliopisto 2022), jossa sivutaan digitointia. Opintojakson lähestymistapa

on kuitenkin enemmän historiallisen ja etnologisen tutkimuksen tekoa tukeva kuin asiantuntijoiden digitointityöhön valmistava.

Ruotsissa Boråsissa kirjastoalan maisteritutkinnossa on mahdollista opiskella 15 op kulttuuriperintöaineiston digitointia (Högskolan i Borås 2022, 2). Toisaalta Tanskassa ja Virossa ei ole tarjolla ollenkaan digitointikoulutusta. Esimerkiksi Tanskassa, jossa ollaan pitkällä digitaalisuudessa, ollaan vasta heräämässä digitoinnin tarpeellisuuteen.

## Aiemmat koulutukset

Muu Suomessa tarjottu digitointikoulutus on pääsääntöisesti ollut hankkeiden tuottamaa tai muutoin väliaikaista koulutusta. Digitointikoulutusta on tarjottu Mikkelin ammattikorkeakoulussa. Uusimpana koulutuksena on ollut muun muassa Digitoinnin ammattilaiseksi (digitoija) -koulutus (15 op), joka on ollut Kaakkois-Suomen ammattikorkeakoulun ja MikseiMikkelin hankkeessa järjestämää (Kaakkois-Suomen ammattikorkeakoulu 2021). Ensimmäinen asiantuntijoille suunnattu digitointikoulutus Suomessa on ollut Digitoinnin ammattilaiseksi (digitoinnin asiantuntija) -koulutus (15 op), joka on järjestetty vuonna 2021 ja joka on ollut osittain integroitu Tiedonhallinnan ja sähköisen arkistoinnin YAMK-opintoihin. Vuonna 2022 toteutettu Digitoinnin taitajaksi -koulutus (10 op) on suunnattu työelämässä toimiville asiakirjahallinnon, tiedonhallinnan ja arkistoalan työntekijöille, ja se on pyrkinyt tarjoamaan ymmärrystä erityisesti digitointiprosessista sekä digitointiprojektien läpiviennistä eli pääasiassa asiantuntijataso tarpeisiin (Kaakkois-Suomen ammattikorkeakoulu 2022).

## Muut tavat kerryttää digitointiosaamista

Digitointiorganisaatioille tehtyjen kartoitusten ja keskustelujen perusteella on selvää, että Suomessa tarvitaan sekä asiantuntija- että digitoijatason digitointikoulutusta. Suomessa digitointia tekevät työntekijät ovat päätyneet digitointitöihin moninaisista erilaisista lähtökohdista, mutta harvoilla heistä on töihin tullessaan taustalla digitointikoulutusta. Digitointityöhön tai digitoinnin suunnittelutyöhön on päädytty yleisimmin mm. siviilipalveluksen tai valokuvausosaamisen kautta tai mm. arkisto-, museo-, historia-, ja kirjastokoulutustaustalla. Osaaminen on hankittu organisaation sisäisessä perehdytyksessä, oman alan tarjoamissa yksittäisten aihepiirien esitelmissä ja sitä on mahdollisesti ylläpidetty konferenssien kautta. Erityisesti kunnissa ja pienemmissä organisaatioissa digitoinnin suunnittelu saatetaan kuitenkin antaa tehtäväksi kenelle tahansa asiakirjojen kanssa työskentelevälle, ja osaamistarve tulee ilmeiseksi.

## Uusia osaamistarpeita

Nykyään yhä useammin massoja digitoitaessa ammatillinen koulutus voi vastata digitoijien osaamistarpeeseen parhaiten. Tämä voi olla sekä yksilön että organisaation mielestä hidas tapa tuottaa osaamista, mutta toisaalta ammatillinen koulutus antaa mahdollisuuden ymmärtää syvällisesti digitoinnin perustietoja ja -taitoja. Toisaalta olemassa olevaa osaamista voi tunnistaa ja tunnustaa sekä osoittaa näytteinä.

Kuitenkin esimerkiksi organisaation sisäisen digitointiosaamisen puutteen vuoksi digitointityö voidaan päätyä ulkoistamaan. Digitointityön ulkoistaminen vaatii projektin suunnittelijalta ymmärrystä kokonaisprosessista sekä tietoa digitoinnin työkulusta ja digitoinnilla saavutettavista hyödyistä, jotka tulee huomioida kilpailutuksissa. Digitointikoulutuksen tulisikin olla helposti saavutettavaa, ja opintoja olisi hyvä päästä opiskelemaan nopeasti, kun työelämän tilanteet sitä vaativat. Erityisesti Suomen tilanteessa, jossa koulutusta ei ole työelämässä jo toimiville, toimiva ratkaisu voisivat olla lyhyet perustason, syventävämmän osaamisen sekä asiantuntijaosaamisen opintojaksot, joita voisi suorittaa joustavasti.

”

*Digitointityön ulkoistaminen vaatii projektin suunnittelijalta ymmärrystä kokonaisprosessista ja tietoa digitoinnin työkulusta ja digitoinnilla saavutettavista hyödyistä, jotka tulee huomioida kilpailutuksissa.*

## Laatua digitointityöhön

Muun muassa kansallisen tason digitoinnin laatuksiteristöjen määrittely edesauttaa ymmärtämään digitoinnin vaatimuksia. Koulutuksen puute on osaltaan vaikuttanut siihen, että digitointityötä ei nähdä omana ammattialanaan tai edes, että se vaatisi erityisosaamista. Vakiintunut koulutus auttaisi digitointialan profiiliin nostamisessa ja ammatti-identiteetin ja -ylpeyden kehittämisessä, jotka vaikuttaisivat kansallisesti digitointityön ymmärtämiseen, kiinnostavuuteen ja arvostukseen.

### Lähteet:

Högskolan i Borås. 2022. Master program: Biblioteks- och informationsvetenskap, Digitala bibliotek och informationstjänster Utbildningsplan. PDF-dokumentti. <https://kursinfodoc.hb.se/PdfMaker.aspx?type=program&code=NGBIB&revision=13,000&language=SV> [viitattu 1.11.2022].

Jyväskylän yliopisto. 2022. AHAS350 Digitaaliset arkistoaineistot (5 op). WWW-dokumentti. Saatavissa: <https://opinto-opas.jyu.fi/2022/fi/opintojakso/ahas350/> [viitattu 1.11.2022].

Kaakkois-Suomen ammattikorkeakoulu. 2021. Digitoinnin ammattilaiseksi. WWW-dokumentti. Saatavissa: <https://www.xamk.fi/tutkimus-ja-kehitys/digitoinnin-ammattilaiseksi/> [viitattu 1.11.2022].

Kaakkois-Suomen ammattikorkeakoulu. 2022. Digitoinnin taitajaksi. WWW-dokumentti. Saatavissa: <https://www.xamk.fi/tutkimus-ja-kehitys/digitoinnin-taitajaksi/> [viitattu 1.11.2022].

Opetushallitus. 2021. Media-alan ja kuvallisen ilmaisun perustutkinto OPH-5059-2021. Tallennetuotannossa työskentely, 30 osp. WWW-dokumentti. Saatavissa: <https://eperusteet.opintopolku.fi/#/fi/amatillinen/4610160/tutkinnonosat/4610082> [viitattu 1.11.2022].

# Digitointiprojektien haasteet – case-esimerkit kuvaamassa digitointityön monimuotoisuutta

EMMI LIIKANEN

*Digitointiprojektien toteutus vaihtelee riippuen toteuttavasta organisaatiosta, digitointitarpeesta ja digitoitavista aineistoista.*

Arkistotoimijoille, kunnille, digitoinnin palveluntarjoajille sekä yhdistyksille ympäri Suomen lähetettiin kysely tehdyistä digitointiprojekteista. Osalle digitoinnin toteuttajista toteutettiin myös tarkentavat haastattelut, jotka valikoituivat tähän oppaaseen. Tuomme esille erilaisia, oikeita projektiesimerkkejä. Esimerkkien tavoitteena on herättää ajatuksia siitä, miten asioita voi tehdä, ja auttaa peilaamaan oman organisaation tulevan digitointiprojektin tavoitteita sekä toimenpiteitä ja tuoda esille opittuja asioita. Kaiken kaikkiaan vastauksia saatiin yhteensä kymmeneltä toimijalta ja näistä haastateltiin erikseen vielä viisi. Case-esimerkkeinä esitellään Lahden, Mikkelin ja Turun kaupungit, Miehikkälän museot sekä Päivälehdän arkisto.

Esimerkkiprojekteissa nousi esiin samanlaisia haasteita, vaikka osa projekteista oli hyvin erityyppisiä keskenään. Kolme vaihetta nousi haastateluissa useimmiten esille:

- aineiston valmistelu
- kilpailutustekniset seikat
- laadunvalvonta ja varmistus

## Aineiston valmistelu

Haastateltujen esimerkkien kohdalla esiin nousi erityisesti aineiston valmistelu usein aikaa vievimpänä osuutena koko projektissa. Aineistosta, sen

laadusta ja fyysisestä kunnosta riippuen tämä voi sisältää niittien, klemmarien, liimojen, seläkkeiden, tarralappujen ym. poistamista tuhansista asiakirjoista tai vedosten ja negatiivien vertailua. Lisäksi valmisteluun kuuluu tärkeänä osana aineiston järjestely, joka aineiston arkistointitavasta riippuen vie enemmän tai vähemmän aikaa. Järjestelyn jälkeen kerätään talteen meta- ja tunnistetiedot. Vaikka vanhakin aineisto olisi hyväkuntoista, voi se ikänsä vuoksi vaatia aikaa taustatyön tekemiseen, jotta se saadaan järjestelyä ja merkittyä oikein. Tästä yhtenä esimerkkinä on tapaus, jossa digitoitiin rakennuslupia. Osa vanhimmasta aineistosta oli arkistoitu siten, että aineiston ja arkistoinnin tulkintatapa oli kadonnut hiljaisena tietona työntekijöiden eläköityessä.

## Kilpailutuksen valmistelu

Projekteissa, joissa osa toteutusta ostettiin ulkopuolisena työnä, nousi esiin kilpailutuksen valmistelun merkitys. Kuntien apuna kilpailutusteknisissä asioissa ovat omat hankintayksiköt, mutta tarjouspyynnön sisältö jää projektin toteuttajan vastuulle. Tämä edellyttää projektin toteuttajalta hyvää ymmärrystä siitä, mitkä vaiheet digitointiprojekti sisältää. Tarjouspyynnössä tulee osata määrittellä vastuut ja työvaiheet niin, ettei tarjouksen jättäjille jää liikaa tulkinnan varaa. Tällä vältetään, että ei tule hyvin erisisältöisiä tarjouksia, joita voi olla vaikea arvioida keskenään. Kun kilpailutus tehdään aineiston alustavan valmistelun jälkeen tai valmistelun loppupuolella ja tiedossa on aineiston laatu, laajuus sekä määrä, tarjouspyynnön tekeminen ja vastuiden määrittely helpottuu.

Katso liitteet  
sivuilla 50-55:  
Case-esimerkit.



## Laadunvalvonta ja -varmistus

Laadunvarmistuksen määritelmät ja mahdollisen ohjeistuksen tekeminen oli yleisesti erityisen tärkeää jo aineiston valinnan aikana. Tähän vaikutti suuresti se, miksi kyseinen aineisto digitointiin ja mihin käyttötarkoitukseen se mahdollisesti lopulta haluttiin. Kun digitointiprojekti tehtiin kokonaan omana työnä, laadunvalvontaan ja -varmistukseen oli tehty oma sisäinen ohjeistus. Ulkoistetuissa digitointiprojekteissa oli tärkeää, että nämä oli määritelty tarjouspyyntöön sekä lopulliseen hankintasopimukseen molempien osapuolten osalta. Ostopalveluina toteutettujen digitointien osalta osassa projekteissa oli etukäteen määritelty ja sovittu koedigitoinnista, jossa laatu varmistettiin sovitun mukaiseksi. Joissain projekteissa tehtiin koeotantoja tietyissä erissä, joissa aineisto tarkistettiin tarkemmin.

## Työelämäverkostot digitointityön tukena

Suurin osa esimerkeistä on kunnilta, joissa projekteja oli joko jo tehty tai käynnistetty. Kuntien projektivastaavien kanssa käydyissä haastatteluissa kävi ilmi, että moni oli omaa projektia suunnitellessaan tutkinut, mitä muualla on tehty, sekä saanut neuvoja ja vinkkejä muilta kunnilta. Työelämäverkostot ovatkin tärkeä tiedonsaantikanava. Toisaalta se tuo esiin saatavilla olevan tiedon rajallisuutta ja osaamistarpeita.

### Lähteet:

- Hyvönen, T. 2022. Digitoinnin projektityöntekijä. Haastattelu 18.10.2022. Turun kaupunki.
- Hyvönen, T. 2022. Digitoinnin projektityöntekijä. Sähköpostiviesti 31.5.2022. Turun kaupunki.
- Hyvönen, T. 2022. Digitoinnin projektityöntekijä. Sähköpostikeskustelu 4.-8.11.2022. Turun Kaupunki.
- Häyrinen, T. 2022. Asianhallintapäällikkö. Haastattelu 7.10.2022. Lahden kaupunki.
- Häyrinen, T. 2022. Asianhallintapäällikkö. Sähköpostiviesti 30.5.2022. Lahden kaupunki.
- Häyrinen, T. 2022. Asianhallintapäällikkö. Sähköpostikeskustelu 10.-11.11.2022. Lahden kaupunki.
- Mieto, J. 2022. Kokoelmavastaava. Haastattelu 29.9.2022. Päivälehden arkisto.
- Mieto, J. 2022. Kokoelmavastaava. Sähköpostiviesti 31.5.2022. Päivälehden arkisto.
- Mieto, J. 2022. Kokoelmavastaava. Sähköpostikeskustelu 4.-8.11.2022. Päivälehden arkisto.
- Mustapää, M. 2022. Asiakirjahallintapäällikkö. Haastattelu 21.10.2022. Mikkelin kaupunki.
- Mustapää, M. 2022. Asiakirjahallintapäällikkö. Sähköpostiviesti 1.6.2022. Mikkelin kaupunki.
- Mustapää, M. 2022. Asiakirjahallintapäällikkö. Sähköpostikeskustelu 9.-10.11.2022. Mikkelin kaupunki.
- Oinonen, A. 2022. Museotutkija. Haastattelu 15.11.2022. Miehikkälän museot.
- Oinonen, A. 2022. Museotutkija. Sähköpostikeskustelu 15.-16.11.2022. Miehikkälän museot.





# Liitteet



# Case: Lahden kaupunki

## TAVOITTEET

Tietopalvelun suorittamisen mahdollistaminen sähköisesti ja aineistoon pääsyn varmistaminen

## AINEISTO

Peruskoulun päättötodistukset, erotodistukset ja oppilaskortit

## RAHOITUS

Oma rahoitus

## PROJEKTIN TOTEUTUSTAPA

Toteutettu kokonaan omana työnä

## LAITTEISTO

Monitoimilaite ja tekstintunnistuslaite

## OHJAAVAT DOKUMENTIT

Arkistolaitoksen säilytysmääräaika-määräykset

## Digitointi-projekti

### Päättötodistukset ja oppilaskortit Digiin

#### Projektin lähtökohdat ja kulku

- Koulutodistuksien tietopyyntöjä on paljon, josta lähti tarve digitoinnille ja työn sujuvoittamiselle.
- Aineiston digitointi aloitettiin todellisen käyttötarpeen arvioinnilla.
- > Näissä aineistoissa käyttötarve on uusimmasta vanhimpaan, eli digitointi aloitettiin nykypäivän aineistosta taaksepäin.
- Aineiston valmistelussa siivottiin ja varmistettiin, että aineisto on yhteismitallista.
- Aineistosta kerättiin metatiedot, jotka dokumenttien luonteen vuoksi ovat selkeät ja yhteismitalliset.
- Aineiston skannaus ja tallennus verkokolevyille.
- Digitoinnin jälkeen paperiset todistukset tuhottiin, analogista arkistoa ei digitoinnin jälkeen näistä enää säilytetty.
- Digitoinnin jälkeen paperiset oppilaskortit arkistoitiin ja säilytettiin.

#### Kokemukset projektista

*“Aineisto oli yhdenmukaista ja helposti käsiteltävää.”*

#### Omat tärkeimmät muistisäännöt

- Aineiston tarvemääritys on alussa tärkeää: mitä aineistoa digitoida, miksi ja missä järjestyksessä.
- Metatiedoissa kannattaa miettiä, mikä on oleellista ja mikä ei.

# Case: Miehikkälän museot

## TAVOITTEET

- Luoda museon kuvakokoelmille digitointiprosessin käytännöt ja menettelytavat
- Avata kulttuurihistoriallisesti merkittävä ja kysytty kokoelma Finnassa, sekä rikastaa kokoelmakuvien metietoja
- Tehostaa museon kuvapalvelua

## AINEISTO

Valokuvavedoksia 1910-30-luvuilta, osa kopioitu yksityisistä kokoelmista. Osa aineistosta mahdollisesti originaalina tai reprotuna muilla museoilla.

## RAHOITUS

Oma rahoitus

## PROJEKTIN TOTEUTUSTAPA

Projektisuunnitelma, aineiston valmistelu, valinta, metatietojen täydennys museotietojärjestelmään, skannaus ja avaus Finnaan tehdään itse. Tarvittaessa tehdään yhteistyötä muiden näitä kuvia kokoelmissaan omaavien museoiden kanssa.

## LAITTEISTO

Oma laitteisto: tasoskanneri

## OHJAAVAT DOKUMENTIT

- EU:n tietosuojasetus ja kansallinen tietosuojalaki
- Kansallisarkiston määräykset
- Laki digitaalisten palveluiden tarjoamisesta
- Museoiden luettelointiohje
- Tekijänoikeuslaki

## Pilotti-digitointiprojekti

### Kalliokosken lasitehtaan valokuvien digitointi

## Projektin esittely ja vaiheistus

- Pilottiprojekti, jossa käydään läpi museon valokuvakokoelmien digitointiprosessi, kuvien avaaminen Finnassa. Huomioidaan myös kulttuuriperintöaineistojen pitkäaikaisäilyttämiseen (PAS) vaadittavat digitoinnin laatuvaatimukset ja standardit.
- Projektisuunnitelma tehty keväällä 2022.
- Aineiston kokoaminen eri lähteistä kesä-syyskuu 2022.
  - > Oma aineisto on pääosin kopiovedoksia alkuperäisistä valokuvista.
  - > Aineistoa on itsellä, sekä osittain Suomen Lasimuseolla ja Kymenlaakson museolla.
  - > Osa myös samoja kuvia, joko originaaleja tai jo digitoituja -> tärkeää varmistaa laadukkaita versioita.
- Metatietojen täydentäminen aineiston valinnan jälkeen (täydennysluettelointi).
  - > Koska kuvia voi olla useammasta museosta, niiden metatietojen laatu ja määrä voi vaihdella. On tärkeää täydentää tiedot yhteneväisesti.
- Skannaus ja mahdollinen yhteistyönä tehty repro digitointi muiden museoiden kanssa.
- Kuvien avaaminen Finnassa.

Projektista saadut kokemukset kirjataan loppuraportiksi, jonka pohjalta voidaan laatia yksityiskohtaisempia ohjeistuksia (esim. laaduntarkkailusuunnitelma).

## Kokemukset projektista

- Projektisuunnitelman tekoon on hyvä varata riittävästi aikaa.
- Tällä hetkellä aineistoa digitointiprosessin toteutuksen eri osa-alueista on paljon, mutta se on hyvin hajallaan ja hyvin eri tasoista (osa yleisemmällä tasolla, osa yksityiskohtaisempaa).
- Välillä oli vaikea hahmottaa oman projektin kannalta tärkeimmät tietolähteet.

# Case: Mikkelin kaupunki

## TAVOITTEET

Viranomaistoiminnan tehostaminen ja tietopalvelun nopeuttaminen

## AINEISTO

Lupahakemuksia, päätöksiä, pääpiirustuksia, rakennussuunnitelmia ja lvi-suunnitelmia

## RAHOITUS

Oma rahoitus

## PROJEKTIN TOTEUTUSTAPA

Projektin suunnittelu ja hallinnointi, aineiston valmistelu ja laadun valvonta omana työnä. Digitointi ja aineiston kuljetus ulkoistettu.

## LAITTEISTO

Palveluntuottajan laitteet

## OHJAAVAT DOKUMENTIT

- Oma alustava projektisuunnitelma
- Muilta kunnilta saadut projektisuunnitelmat, -raportit ja hankinta-asiakirjat
- Kansallisarkiston määrätykset

## Digitointi- projekti

Kaupungin rakennus-  
valvontaviranomaisen  
lupa-arkiston  
digitointi

### Projektin resurssit ja suunnitelma

- Projektisuunnitelman tekoon käytettiin apuna muilta kaupungeilta saatuja suunnitelmia, raportteja ja hankinta-asiakirjoja.
  - Projektia tekemässä 3 vuoden aikana 2-3 arkistonhoitajaa.
  - Itse tehdään aineiston valmistelu (järjestely lupa-akteittain ja aineiston fyysinen siivous).
- > Palveluntarjoaja tekee teknisen digitoinnin sekä hoitaa logistiikan.

### Laadunvalvonta

- Omille sisäisille prosesseille ohjeistukset.
- Tarjouskilpailussa valittavalle palveluntuottajalle määritellyt laatuvaatimukset.
- Tarjouskilpailussa edellytettiin referenssejä sekä tietoturvallisia toimintiloja ja palvelimia.
- Tarjouskilpailun voittajalla eli palveluntuottajalla teetätettiin koedigitointi, jotta voitiin testata ja varmistua että palvelun laatu vastaa tarjouskilpailussa edellytettyä, ennen sopimuksen allekirjoittamista.

### Kokemukset projektista

Projekti on vielä kesken, mutta tähänastisen työn kokemuksia:

- Suunnittelu on kaiken A ja O, kokemuksia kannattaa kysyä muilta.
- Kilpailutuksen valmisteluun kannattaa käyttää aikaa, mahdollisesti markkinavuoropuhelu potentiaalisten yritysten kanssa etukäteen.
- Tarjouspyyntö kannattaa laatia huolella siten, että keskeiset vastuut on määritelty selkeästi, eikä tulkinnanvaraa palvelun toteutuksessa ja esim. aikatauluissa jää liikaa.
- Tärkeät määriteltävät vastuut ja roolit tarjouspyynnössä: kuka tekee, mitä tekee, missä vaiheessa tekee.

# Case: Päivälehdteen arkisto

## TAVOITTEET

Aineiston käytettävyyden parantaminen ja säilyvyyden takaaminen

## AINEISTO

A4 tai pienempiä piirustusoriginaaleja

## RAHOITUS

Omasta budjetista varattu vuosittainen summa digitointiin

## PROJEKTIN TOTEUTUSTAPA

Projektin suunnittelu ja aineiston esivalmistelu omana työnä. Varsinainen skannaustyö ulkoistettu kevyellä kilpailuksella.

## LAITTEISTO

Palveluntuottajan laitteet

## OHJAAVAT DOKUMENTIT

- Oma digitointisuunnitelma
- Oma laadittu kaava tiedostonimeämiselle
- Kansallisarkiston vuoden 2016 digitointisuositus

## Digitointi-projekti

Rafael Rindellin piirustukset

### Projektin yksinkertaistettu eteneminen

- Aineiston valinta.
  - Aikataulutus ja suunnittelu.
  - Aineiston valmistelu (luettelointi, numerointi/tunnistus).
  - Logistiikan hahmotus ja aineiston pakkaaminen.
  - Palveluntarjoajan kilpailutus.
- > Palveluntarjoaja tekee teknisen digitoinnin.
- > Palveluntarjoaja toimittaa testiversion laadunvalvontaa varten.
- Laadunvalvonta.
  - Korjauspyynnöt.
- > Palveluntarjoaja palauttaa aineiston.
- Aineiston julkaisukappaleiden ja luettelointitietojen vieminen arkistotietojärjestelmään.
  - Digitaalisten tallekappaleiden arkistointi.
  - Sähköisen arkistohallintajärjestelmän ajantasaistaminen.

Projektin kokonaiskesto oli noin vuoden ja siinä työskenteli yksi henkilö muun työn ohella.

### Kokemukset projektista

*“Tämän projektin osalta kaikki onnistui lopulta melko hyvin. Sen olen oppinut kantapään kautta, että valmistelu voi viedä yllättävän kauan aikaa, vaikka aineisto olisi luetteloitukin. Digitointia valmistellessa on joka tapauksessa identifioitava ja kuvailtava jokainen asiakirja tai piirustus ja se vie aikaa. Lisäksi palveluntarjoajan kilpailutus kannattaa ajoittaa valmistelun loppupuolelle, jolloin aikataulua on helpompi arvioida.”*

### Omat tärkeimmät muistisäännöt

- Kilpailutus kannattaa tehdä kun aineisto on valmisteltu (helpottaa aikataulun hahmottamista).
- Kertaa aikaisemmin opitut, ennen kun aloitat uuden projektin.

# Case: Turun kaupunki

## TAVOITTEET

Aineiston käytettävyyden parantaminen ja tietopalvelun helpottaminen sekä viranomaistoiminnassa että muita asiakkaita palveltaessa

## AINEISTO

Pääpiirustuksia ja katselmuspöytäkirjoitteita. Yli 150 000 asiakirjaa.

## RAHOITUS

Oma rahoitus

## PROJEKTIN TOTEUTUSTAPA

Projektin suunnittelu ja hallinnointi, aineiston valmistelu sekä rakennuslupa-aineiston paikannus kartalle omana työnä. Digitointi ulkoistettu.

## LAITTEISTO

Palveluntuottajan laitteet

## OHJAAVAT DOKUMENTIT

- Oma ohjeistus sisäiselle prosessille ja eri työvaiheille
- Arkistolaitoksen suositukset digitoinnin laatukriteereiksi
- Kansallisarkiston määrätykset

## Digitointi-projekti

### Rakennuslupa-aineiston digitointi

#### Projektin resurssit ja suunnitelma

- Projektille tehtiin suuntaa antava suunnitelma, jossa mietittiin henkilöstön sijoittaminen ja kustannusten jako.
  - Projektissa on työskennelty kerrallaan 4-6 hlöä.
  - Projektissa on määritetty palveluntarjoajaa sitovat laatuvaatimukset ja ohjeistukset, minkä jälkeen kilpailutus on hoidettu yhdessä kaupungin hankintapalveluiden kanssa.
- > Palveluntarjoaja tekee teknisen digitoinnin.

#### Laadunvalvonta

- Omille sisäisille prosesseille ohjeistukset.
- Kilpailutuksessa palveluntarjoajaa koskevat ohjeet ja määrätykset.
- Digitoinnin laatua on seurattu tarkemmin projektin alussa ja aina palveluntarjoajan vaihtuessa.

#### Kokemukset projektista

*“Suurten massa-aineistojenkin digitointi sujuu omalla painollaan, kun prosessien luomiseen ja metatietojen määrittämiseen on jo suunnitteluvaiheessa panostettu.”*

#### Omat tärkeimmät muistisäännöt

- Projekti kannattaa suunnitella heti alussa hyvin, koska korjausten tekeminen on jälkikäteen hankalaa.
- Muutosten tekemistä kesken projektin ei silti kannata pelätä, sillä varsinkin vuosia kestävässä projektissa huomataan ajan mittaan, mikä toimii ja mitä voisi parantaa.

### Case-esimerkkien lähteet:

#### Lahden kaupunki

Häyrinen, T. 2022. Asianhallintapäällikkö. Haastattelu 7.10.2022. Lahden kaupunki.

Häyrinen, T. 2022. Asianhallintapäällikkö. Sähköpostikeskustelu 10.-11.11.2022. Lahden kaupunki.

Häyrinen, T. 2022. Asianhallintapäällikkö. Sähköpostiviesti 30.5.2022. Lahden kaupunki.

#### Miehikkälän museot

Oinonen, A. 2022. Museotutkija. Haastattelu 15.11.2022. Miehikkälän museot.

Oinonen, A. 2022. Museotutkija. Sähköpostikeskustelu 15.-16.11.2022. Miehikkälän museot.

#### Mikkelin kaupunki

Mustapää, M. 2022. Asiakirjahallintapäällikkö. Haastattelu 21.10.2022. Mikkelin kaupunki.

Mustapää, M. 2022. Asiakirjahallintapäällikkö. Sähköpostikeskustelu 9.-10.11.2022. Mikkelin kaupunki.

Mustapää, M. 2022. Asiakirjahallintapäällikkö. Sähköpostiviesti 1.6.2022. Mikkelin kaupunki.

#### Päivälehdien arkisto

Mieto, J. 2022. Kokoelmavastaava. Haastattelu 29.9.2022. Päivälehdien arkisto.

Mieto, J. 2022. Kokoelmavastaava. Sähköpostikeskustelu 4.-8.11.2022. Päivälehdien arkisto.

Mieto, J. 2022. Kokoelmavastaava. Sähköpostiviesti 31.5.2022. Päivälehdien arkisto.

#### Turun kaupunki

Hyvönen, T. 2022. Digitoinnin projektityöntekijä. Haastattelu 18.10.2022. Turun kaupunki.

Hyvönen, T. 2022. Digitoinnin projektityöntekijä. Sähköpostikeskustelu 4.-8.11.2022. Turun Kaupunki.

Hyvönen, T. 2022. Digitoinnin projektityöntekijä. Sähköpostiviesti 31.5.2022. Turun kaupunki.

# Digitointiprosessin vaiheita

Varsinaiseen digitointiprosessiin voi kuulua erilaisia toimenpiteitä, joista esimerkkejä on seuraavassa listauksessa. Vaiheet voivat tapahtua tai olla tapahtumatta prosessista riippuen prosessin eri vaiheissa tai eri järjestyksessä. Osa vaiheista voidaan hankkia ulkoiselta palveluntarjoajalta.

## SUUNNITTELU

- Digitoitavan aineiston valintaperiaatteet
- Digitointijärjestyksen suunnittelu ja priorisointi
- Digitointimenetelmän valinta
- Digitoinnin jälkeisen säilytystarpeen ja -vaatimusten selvittäminen
- Digitoinnin jälkeisen järjestelmän vaatimusten huomiointi ja digitaalisten aineistojen siirron ja käyttöoikeuksien suunnittelu haluttuun tietojenhallintajärjestelmään
- Metatietojen ja tiedostomuotojen määrittely
- Digitoinnin laadun suunnittelu ja haluttavan lopputuloksen suunnittelu
- Digitoinnin aikaisen aineistoista annettavan tietopalvelun suunnittelu

## AINEISTON VALINTA

- Säilytettävän sisällön seulonta ja säilytysaikojen tarkistaminen
- Aineistojen järjestäminen
- Digitoitavan aineiston valinta
- Ei-digitoitavien aineistojen erottelu
- Erittäin huonokuntoisen, huonokuntoisen tai erikoisaineiston erottelu (erilliset valmistelu- ja/tai digitointimenetelmät)

## SIIRROT JA KULJETUKSET

- Siirrosta sopiminen ja kuljetuksen suunnittelu
- Siirrettävän aineiston kartoittaminen
- Tarvittaessa tunnistetietojen merkintä ja/tai siirtoon suojaaminen
- Aineiston kuljetukseen valmistelu
- Aineiston nouto ja kuljetus
- Aineiston vastaanotto, tarkistus ja siirron hyväksyminen
- Aineiston nouto valmisteluun



## **AINEISTON VALMISTELU**

- Digitointia haittaavien esineiden poisto aineistosta (mm. klemmarit, niitit, teipit)
- Huonokuntoisen tai erikoisaineiston käsittely digitointiin
- Määräajan säilytettävien aineistojen seulonta ja esineiden poistaminen
- Aineiston kuvailu ja hyväksyntä
- Aineiston esikäsittely/valmistelu skannaukseen/valokuvaukseen
- Digitointijärjestykseen järjestäminen
- Skannaustavan valinta ja toimitus digitointiin

## **DIGITOINTI/SKANNAUS/VALOKUVAUS**

- Työtilojen valmistelu digitointiin sopiviksi
- Pölyttömien tilojen varmistaminen säännöllisellä siistinä pidolla ja siivouksella
- Laitteiden testaus ja kalibrointi
- Laitteiden ja ohjelmistojen valmistelu ja optimointi skannaukseen/valokuvaukseen
- Skannaus/valokuvaus
- Kuvien värisävyjen tarkastus
- Järjestelmien ja laitteiden suorituskyvyn tarkkailu
- Laadunvarmistuksen menetelmien hyödyntäminen (mm. mittataulut, referenssiarvot ja analysointiohjelmistot)

## **AINEISTON JA MENETELMIEN KUVAILU SEKÄ DOKUMENTOINTI**

- Aineiston kuvailu ja metatietojen tallennus
- Menetelmien ja rajoitusten dokumentointi eheyden varmistamiseksi ja metatietojen tallennus
- Digitoinnissa käytettävien laitteistojen ja ohjelmistojen dokumentointi ja metatietojen tallennus
- Metatietojen täydentäminen
- Muu dokumentointi

## **AUTOMAATTISET MENETELMÄT JA SISÄLLÖNTUNNISTUS**

- Automaattinen indeksointi
- Automaattinen tekstintunnistus ja sen tarkistus (mm. OCR, HTR)
- Automaattinen asiansanoitus ja sen tarkistus (mm. Annif)
- Muiden automaattisten tai tekoälyn menetelmien käyttö

## **DIGITAALISEN AINEISTON HALLINTA JA LAADUNVARMISTUS**

- Kansiorakenteen luominen ja päivittäminen
- Vastaanottavan järjestelmän vaatimat aineiston hallinnan toimenpiteet
- Tiedostojen nimeäminen
- Tiedostojen tai digitaalisten ilmentymien (tallekappale, käyttökappale) siirto kansiorakenteeseen tai järjestelmään
- Digitaalisen aineiston eheyden tarkistaminen, digitoidun aineiston validointi (esim. otannat, visuaalinen tarkastus, tunnistustulos) ja laadunvarmistus
- Laadunvarmistuksen menetelmien hyödyntäminen (mm. mittataulut, referenssiarvot ja analysointiohjelmit)

## **DIGITAALINEN ARKISTOINTI JA KÄYTTÖÖN SAATTAMINEN**

- Digitoidun aineiston digitaaliseen säilytykseen tai arkistointiin siirto
- Säilytysmenetelmien ja -järjestelmien migraatiot
- Tiedon eheyden varmistaminen järjestelmissä pitkällä aikavälillä
- Digitoidun aineiston käyttöön ja jakeluun saattaminen

## **AINEISTON JÄLKIKÄSITTELY JA TUHOAMINEN**

- Analogisen aineiston jälkikäsitteily ja suojaus sekä uudelleen arkistointi
- Tietoturvallinen tuhoaminen

# Esimerkki digitointiprojektin projektisuunnitelman sisällysluettelosta

## PROJEKTISUUNNITELMA

Projektin nimi

### 1. TAUSTA

- 1.1. Projektin tausta ja muutostarve
- 1.2. Kohderyhmien ja sidosryhmien tunnistaminen eli keitä muutos palvelee
- 1.3. Huomioitava lainsäädäntö, standardit ja muut ohjeet
- 1.4. Projektiorganisaation suunnittelu, projektivastuut ja ohjaus

### 2. TAVOITELTAVAT HYÖDYT

- 2.1. Digitoinnin tavoitteet, saavutettavuuden ja käytettävyyden huomioiminen
- 2.2. Vaikutukset ja hyödyt

### 3. AINEISTOT

- 3.1. Digitoitavan aineiston kuvaus ja laajuus. Aineistokokonaisuuden rajaus. Perustelut valitulle digitointimenetelmälle.

### 4. RESURSSIT

- 4.1. Henkilöstö ja osaaminen
- 4.2. Tilat
- 4.3. Laitteet, järjestelmät, ohjelmistot, materiaalit ja työvälineet
- 4.4. Palvelutarjonta ja mahdolliset toimittajat ja alihankkijat

### 5. BUDJETTI JA HANKINTOJEN HALLINTA

- 5.1. Budjetti. Kustannukset ja tarvittavat hankinnat. Kustannus-hyötysuhde.
- 5.2. Hankintojen hallinta. Hankinta- ja kilpailutusperiaatteet. Hankintavaltuudet ja sopimusten hallintatapa.
- 5.3. Kokonaiskustannukset ja tehtäväkokonaisuuksien kustannukset

## 6. TYÖSUUNNITELMA JA AIKATAULUTUS

### 6.1. Projektin ositus, vastuut ja työsuunnitelma

6.1.1. Digitointiprosessin vaiheistus: itse tehtävät ja hankittavat

6.1.2. Aikataulun ja resurssien raportointi- ja ohjausperiaatteet

### 6.2. Tehtäväkokonaisuuksien tunnistaminen, aikataulut, järjestys ja vastuut

6.2.1. Aineiston valinta

6.2.2. Metatietojen määrittely

6.2.3. Aineiston valmistelu

6.2.4. Siirrot ja logistiikka

6.2.5. Digitoinnin valmistelu

6.2.6. Digitointi

6.2.7. Tekstintunnistus

6.2.8. Digitoidun aineiston validointi

6.2.9. Digitaalisten ilmentymien siirto

6.2.10. Digitoidun aineiston arkistointi ja/tai käyttöön saattaminen. Saavutettavuus.

6.2.11. Analogisen aineiston jälkikäsittely

6.2.12. Projektin päättäminen

## 7. LAADUNHALLINTA JA RISKIEN KOKONAISARVIOINTI

### 7.1. Laadunhallinta ja laatusuunnitelma

7.1.1. Laadun suunnittelu, varmistus ja ohjaus. Kokonaislaadun hallinta.

7.1.2. Digitoidun aineiston laadun mittauksen, analysoinnin, arvioinnin ja valvonnan suunnittelu

### 7.2. Strategisten, taloudellisten, toiminnallisten ja vahinkoriskien tunnistaminen ja arviointi. Hyväksyttävä riskitaso. Riskienhallintasuunnitelma.

7.2.1. Työsuhteen ja osaamisen riskit. Tarvittavan osaamisen määrittely ja uuden osaamisen tarve

7.2.2. Tuotettavat ohjeet ja perehdytystoimenpiteet

7.2.3. Työympäristön ja toimitilojen turvallisuus ja kuormituksen riskit

7.2.4. Tietojärjestelmien suojaus, tietoturvallisuus ja salassapito

7.2.5. Alihankinnan ja sopimusten riskit

### 7.3. Muutostenhallinta, kehittäminen ja toiminnan jatkuvuus

## **8. VIESTINTÄ, TIEDONHALLINTA JA DOKUMENTOINTI**

8.1. Välttämättömien dokumenttien määrittely

8.2. Raportointisuunnitelma

8.3. Viestintäsuunnitelma

### **LIITTEET**

Vaatimusmäärittely

Metatietomäärittely

Käyttötapaesimerkit

Kilpailutuskaavat

Sopimusluonnos

# Hankintasuunnitelma

**Ostajaorganisaatio**

**Hankinnan otsikko**

**Hankinnan lyhyt kuvaus**

**Hankinnan tai hankinnan osien kuvaukset**

**Kuvaus lisähankintamahdollisuuksista**

**Hankinnan liitteet (valitse tarpeelliset)**

- Palvelukuvaus digitointipalvelusta
- Vaatimusmäärittely
- Käyttötapausesimerkit
- Palvelutasovaatimukset
- Hankinnan pitkä kuvaus
- Hankinnan arviointiperusteet
- Täydennettävä projektisuunnitelmaluonnos
- Täydennettävä sopimusluonnos

# Esimerkki hankintasuunnitelmasta: Mikkelin kaupunki

## Ostajaorganisaatio

Mikkelin kaupunki

## Hankinnan otsikko

Mikkelin kaupungin rakennusvalvontaviranomaisen lupa-arkiston digitointi

## Hankinnan tyyppi

Palvelut

## Hankinnan lyhyt kuvaus

Hankinnan kohteena on Mikkelin kaupungin rakennusvalvontaviranomaisen lupa-arkiston digitointi osana kaupungin digitointiprojektia tarjouspyyntöasiakirjojen mukaisesti. Hankintaan kuuluu digitoitavien asiakirjojen noutaminen kaupungilta, asiakirjojen skannaaminen sekä metatietojen poimiminen ja tuottaminen niihin, analogisten asiakirjojen palauttaminen kaupungille sekä digitoitujen asiakirjojen siirtäminen kaupungin osoittamalle palvelimelle tai tietojärjestelmään tarjouspyyntöasiakirjoissa esitetyllä tavalla. Hankintaan kuuluu lisäksi edellä mainittuihin toimenpiteisiin liittyvät projekti- ja asiantuntijapalvelut. Digitoinnilla tavoitellaan viranomaistoiminnan tehokkuuden ja tuloksellisuuden lisäämistä, sillä digitaalisessa muodossa olevat asiakirjat ovat paremmin käytettävissä kuin analogisessa muodossa olevat asiakirjat, jolloin työaikaa ei tarvitse käyttää tiedon etsimiseen arkistosta. Lisäksi viranomaisten asiakirjoihin kohdistuvien tietopyyntöjen toteuttamiseen käytetyn ajan arvioidaan vähentyvän huomattavasti edellä esitetyistä syistä.

## Hankinnan kuvaus

Hankinnan kohteena on Mikkelin kaupungin rakennusvalvontaviranomaisen lupa-arkiston digitointi osana kaupungin digitointiprojektia tarjouspyyntöasiakirjojen mukaisesti.

Kaupungin digitointiprojektissa tarkoituksena on muuttaa digitaaliseen muotoon Mikkelin kaupungin rakennusvalvontaviranomaisen lupa-arkiston lupa-aktiasiakirjat. Rakennusvalvontaviranomaisen lupa-aktit ovat tärkeitä ja käytettyjä viranomaisasiakirjoja, ja niille on käyttötarvetta niin rakennusvalvonnan viranomaistoiminnassa kuin ulkopuolisille toimijoille, kuten kuntalaisille ja yksityisille yhteisöille.

Rakennusvalvontaviranomaisen lupa-arkistossa on noin 720 hyllymetriä lupa-aktiasiakirjoja. Lupa-aktien lisäksi, jotka sisältävät muun muassa rakennuslupahakemuksia, pääpiirustuksia, lausuntoja ja katselmuspöytäkirjoja, on rakennusvalvontaviranomaisen arkistossa myös rakennuslautakuntien pöytäkirjoja sekä muita rakennuslautakuntien asiakirjoja. Lupa-arkisto muodostuu Mikkelin kaupungin eli niin sanotun kantakaupungin sekä liitoskuntien rakennusvalvontojen asiakirjoista. Digitointiprojektissa tarkoituksena on digitoida ensisijaisesti lupa-arkiston lupa-aktiasiakirjat. Digitoinnin ulkopuolelle rajataan lupa-akteista kuitenkin ne viranomaisten toiminnan julkisuudesta annetun lain (621/1999) 24 §:n perusteella salassa pidettävät asiakirjat, jotka ovat julkisen hallinnon tiedonhallinnasta annetun lain (906/2019) 18 §:n mukaisesti turvallisuusluokiteltuja.

Valtaosa lupa-arkiston hyllymetreistä koostuu kantakaupungin lupa-akteista, joita on noin 560 hyllymetriä. Kantakaupungin suuri asiakirjamäärä selittyy sillä, että kaupunkialueelle on rakennettu eniten, jonka lisäksi vuoden 2001 Mikkelin kaupungin, Mikkelin maalaiskunnan ja Anttolan kunnankuntaliitoksen jälkeen, on liitettyjen kuntien uusia lupa-akteja alettu säilyttää osana kantakaupungin lupa-akteja. Mikkelin maalaiskunnan lupa-akteja on noin 75 hyllymetriä, Ristiinan kunnan lupa-akteja on noin 55 hyllymetriä ja Anttolan kunnan, Haukivuoren kunnan sekä Suomenniemen kunnan lupa-akteja on noin 10 hyllymetriä kuntaa kohden. Kantakaupungin lupa-aktien kohdalla lupa-arkistossa on käytetty paljon ohuita arkistokoteloita. Tämä aiheuttaa sen, että lupa-arkiston lupa-aktiasiakirjat täyttävät 720 hyllymetriä tilaa, mutta todellisuudessa itse aineisto ei tiiviimmin koteloituna täytä koko tätä tilaa, vaan todennäköisesti lupa-aktiasiakirjoja on noin 600-700 hyllymetriä tai mahdollisesti vähemmän. Mikkelin kaupunki ei sitoudu tämän tarjouspyynnön perusteella mihinkään minimiostoon, vaan palvelua hankitaan digitoitavan aineiston todelliseen määrään nähden tarvittava määrä.

Digitointiprojektissa tilaajan henkilöstö suorittaa digitoitavan aineiston digitointikuntoon saattamisen, mikä pitää sisällään aineiston järjestämisen, tiivistämisen, jälkiseulonnan ja kuvailevien metatietojen keräämisen.

Hankinnan kohteena on digitointiprojektin varsinainen digitointi- eli skannaustyö ja siihen liittyvät kuljetus-, siirto-, projekti- ja asiantuntijatehtävät. Tarjouskilpailun voittaneen toimijan tehtävinä on noutaa digitoitavat asiakirjat tilaajalta, skannata ne vaadituin kriteerein, poimia ja tuottaa niihin vaaditut metatiedot, palauttaa analogiset asiakirjat takaisin tilaajalle ja siirtää digitoidut asiakirjat tilaajan osoittamalle palvelimelle tai tietojärjestelmään. Edellä mainitut työvaiheet toistuvat useamman kerran projektin aikana, koska lupa-aktiasiakirjoja on tarkoitus digitoida erissä. Projektin alkaessa sovitaan tarkemmin, millaisissa erissä aineistoa on mahdollista digitoida.

Tilaajan henkilöstö tarkastaa palautettujen asiakirjojen kunnon, digitoitujen asiakirjojen laadun, lisää niihin tarvittavat kuvailevat metatiedot ja siirtää ne sähköiseen arkistoon.

### **Kuvaus lisähankintamahdollisuuksista**

Mahdollinen optiovuosi:

Rakennuslupa-aktiasiakirjojen lisäerän tai muiden lupa-arkiston asiakirjojen digitointi alkuperäisen vaatimusmäärittelyn mukaisesti ja alkuperäisessä palvelussa tarjotuilla hinnoilla.

Lisäksi tilaaja voi hankkia palvelun seuraavilla täydentävillä lisäpalveluilla:

Rakennuslupa-aktiasiakirjojen digitointi TIFF- tai JPEG-tiedostomuotoon metatietoineen vaatimusten mukaisesti. Digitointi TIFF- tai JPEG-tiedostomuotoon sisältää myös yhden kuvailevan metatiedon lisää alkuperäiseen vaatimusmäärittelyyn: nimike. Lisäpalvelu toteutettaisiin muilta osin alkuperäisen vaatimusmäärittelyn mukaisesti (tarjouspyyntö liitteineen, erityisesti liite 1 Palvelukuvaus ja vaatimukset).

Rakennuslupa-aktiasiakirjoista muiden kuin tarjouspyynnön metatietojen poimiminen. Lisäpalvelu toteutettaisiin muilta osin alkuperäisenvaatimusmäärittelyn mukaisesti.

Lisähankinta tehdään, mikäli tilaaja tahtoo toteuttaa lisähankinnan.

Optio-oikeus ja lisäpalveluiden hankinta on ainoastaan tilaajan harkinnassa.



## **Hankinnan liitteet**

Liite 1 Palvelukuvaus ja vaatimukset

Liite 1.1 Kansallisarkiston vaatimukset hävittämiseen tähtäävään digitointiin 2019

Liite 1.2 PAS säilytys- ja siirtokelpoiset tiedostomuodot v1.9.0

Liite 2 Palvelun toteutuskuvaus

Liite 3 Projekti- ja asiantuntijaresurssit

Liite 4 Referenssit

Liite 5 Sopimusluonnos rakennusvalvontaviranomaisen lupa-arkiston digitointi

Liite 5.1 Henkilötietojen käsittelyn ehdot

Liite 5.1.1 Henkilötietojen käsittelytoimien kuvaus

Liite 6 JYSE 2014 palvelut

# Käsitteet ja termit

Käsitteet ja termit koonneet Kati Saltiola, Sari Järn ja Henna Ristolainen.

Käsite/termi	Selitys
<b>Aineisto</b>	Analoginen tai digitaalinen objekti.
<b>Aineistotyyppi</b>	Aineiston tyyppi sisällön mukaan luokiteltuna, esimerkiksi teksti, ääni, video, kuva, verkkoarkisto, tietokanta.
<b>Analoginen aineisto</b>	Paperisessa tai muussa käsin kosketeltavassa muodossa laadittu aineisto, esimerkiksi paperiset dokumentit, kirjeet, kirjat, valokuvat, vinyylilevyt, VHS-videonauhut ja C-kasetit.
<b>Arkisto</b>	Asiakirjoista koostuva kokonaisuus, joka on järjestelty loogisesti tiettyjen ennalta määriteltyjen ehtojen mukaan.
<b>Arkistointi</b>	Arkistointi on asiakirjahallinnan peruskäsite. Arkistointi tapahtuu säilytysajan päättymisen jälkeen (tiedonhallintalaki 906/2019, 21 §) siirtämällä asiakirjat arkistoon. Arkistoinnin tehtävänä on varmistaa asiakirjojen käytettävyys ja säilyminen, huolehtia asiakirjoihin liittyväs-tä hakupalvelusta, määritellä asiakirjojen säilytysarvo ja tuhota tarpeeton aineisto sovittavalla, oikea-aikaisella tavalla. Arkistoinnilla tarkoitetaan myös arkistolain mukaista pysyvää säilyttämistä (8 § 3 momentti), jos se perustuu Kansallisarkiston päätökseen.
<b>Arvonmääritys</b>	Arvonmääritys on asiakirjan elinkaarihallinnan prosessi, jossa määritellään, mitkä asiakirjat arkistoidaan ja säilytetään pysyvästi, ja mitä asiakirjoja säilytetään määräajan sekä kuinka kauan määräajan säilytettäviä asiakirjoja säilytetään. Arvonmääritys voi olla säänneltyä tai omaan harkintaan perustuvaa. Arvonmäärityksen periaatteita voidaan soveltaa aineistoihin. Arvonmääritys ohjaa seulontaa.
<b>Asiakirja</b>	Asiakirja on yksittäisen asian käsittelyä varten laadittu, kirjallisessa muodossa oleva kokonaisuus, jossa voi olla teksti- tai taulukkomuotoista sisältöä tai esimerkiksi kuvamateriaalia. Asiakirja voi olla alun perin analogisena (paperisena), mutta myös sähköisenä, kuten esimerkiksi Word-, Power Point-, Excel- tai jpeg-muotoisena aineistona.
<b>Digitaalinen aineisto</b>	Digitaalisessa muodossa laadittu tai syntynyt aineisto, tai digitaaliseen muotoon teknologian avulla muunnettu aineisto.
<b>Digitalisaatio</b>	Digitoidun tuotteen avulla uusien tuotteiden tai palveluiden inno-voiminen.
<b>Digitaalinen transformaatio</b>	Digitalisoidun tuotteen tai palvelun uudelleenajattelu ja järjestely kokonaan uudeksi konseptiksi.
<b>Digitointi</b>	Digitointi on suppeasti määriteltynä analogisessa muodossa olevan tietoaineiston muuntaminen digitaaliseen muotoon. Se on yksittäisen aineiston käsittelyä siten, että lopputulokseksi tulee indeksoitu, määrämuotoinen digitaalinen aineisto. Laajasti määriteltynä digitointi alkaa suunnittelusta ja digitoitu aineisto tallennetaan joko operatiiviseen järjestelmään tai sähköiseen arkistoon.

<b>Käsite/termi</b>	<b>Selitys</b>
<b>Eheys</b>	Eheys tarkoittaa tiedon yhtäpitävyyttä alkuperäisen tiedon kanssa eli tietoa ei ole muutettu luvatta, se ei ole muuttunut vahingossa ja että mahdolliset muutokset voidaan todentaa. Eheys tarkoittaa myös tietojen aitoutta, väärentämättömyyttä, sisäistä ristiriidattomuutta, kattavuutta, ajantasaisuutta, oikeellisuutta ja käyttökelpoisuutta. Aineiston eheys voi olla joko sisällöllistä ja/tai teknistä.
<b>Elinkaari</b>	Aineistolle tehtävät toimenpiteet alkaen muodostamisesta tai vastaanottamisesta ja päättyen pysyvään säilyttämiseen tai tuhoamiseen.
<b>HTR</b>	Handwritten Text Recognition. Käsien kirjoitetun tekstin koneellinen tunnistus.
<b>Ilmentymä</b>	Aineiston analoginen tai digitaalinen (alkujaan digitaalinen tai digitoitu) olomuoto.
<b>Indeksointi</b>	Indeksoinnilla voidaan tarkoittaa aineiston kuvailua tai tietokannan suunnitteluun liittyviä toimenpiteitä suorituskyvyn parantamiseksi indeksien avulla.
<b>Kuvailu</b>	Kuvailu on prosessi, jossa tuotetaan metatietokokonaisuus, joka mahdollistaa aineiston löytämisen, tunnistamisen, valinnan sekä käyttöön saamisen ja ymmärrettävyyden.
<b>Käytettävyys</b>	Tiedon on oltava siihen oikeutettujen saatavilla sovitusti ja silloin, kun tarve on. Käytettävyys on tiedonsaannin helppoutta, johon vaikuttavat tiedon paikallistettavuus, esiin saatavuus, esitettävyys ja tulkittavuus. Metatiedoilla varmistetaan käytettävyys.
<b>Laadunhallinta</b>	Laadunhallinnan toimenpiteillä pyritään määrittämään ja saavuttamaan organisaation laatutavoitteet. Laadunhallinta sisältää johtamista, suunnittelua, arviointia ja toiminnan kehittämistä. Laadunhallinta kattaa laadun suunnittelun ja ohjauksen, laadunvarmistuksen, laadun arvioinnin ja kehittämisen. Määriteltyä laatua voidaan tarkistaa, tarkkailla ja varmistaa eri toimenpiteillä. Poikkeamat laatumäärittelyihin verrattuna pyritään estämään ja korjaamaan.
<b>Markkinakartoitus</b>	Markkinoilla olevan tiedon selvittämistä, jota hankintayksikön on mahdollista saada ja hyödyntää hankintaa suunnitellessaan ja valmistellessaan.
<b>Markkinavuoropuhelu</b>	Markkinoiden kartoittamiseksi järjestettäviä tapaamisia potentiaalisten tarjoajien kanssa sekä hankinta-asiakirjojen tai muun materiaalin kommentointikierroksia tarjoajien kanssa.
<b>Metatieto</b>	Metatieto on tietoa tiedosta, eli se kuvailee tiedon rakennetta ja kontekstia. Metatiedon avulla aineistoon voidaan kohdistaa hakuja. Aineiston kontekstia, sisältöä ja rakennetta sekä niiden hallintaa ja käsittelyä koko elinkaaren ajan kuvaavaa tietoa. Metatietoa voidaan käyttää muun muassa aineiston hakuun, paikallistamiseen, pitkäaikaissäilyttämiseen ja tunnistamiseen.
<b>Migraatio</b>	Migraatiossa tiedot siirretään tiedostoformaattista ja/tai tietojärjestelmästä toiseen tai siirrytään tietojenkäsittely-ympäristöstä toiseen.
<b>OAIS-viitemalli</b>	Open Archival Information System eli pitkäaikaissäilytysarkiston viitemalli. Viitemallia määrittelevää standardia ISO 14721:2012 (suomenos SFS 5972) käytetään useimpien pitkäaikaissäilytys- ja digitaalisten pitkäaikaissäilytysjärjestelmien suunnittelun lähtökohtana.

<b>Käsite/termi</b>	<b>Selitys</b>
<b>OCR</b>	Optical Character Recognition, eli tekstintunnistus. OCR-tekniikalla digitaalisista kuvista pyritään havaitsemaan tekstiä, jonka tietokone tallentaa tekstiksi.
<b>PAS</b>	Pitkäaikaissäilyttäminen. Toiminta, jolla tietoa säilytetään pitkäaikaisesti kohdeyhteisölle itsenäisesti ymmärrettävässä muodossa ja siten, että tiedon autenttisuus voidaan taata.
<b>Pysyvä säilyttäminen</b>	Pysyvällä säilyttämisellä ei ole määräaikaa. Arkistolaissa käsitteellä pysyvä säilyttäminen tarkoitetaan tiedonhallintalain mukaista arkistointia.
<b>Seulonta</b>	Aineistolle tehdyn arvonmäärityksen toimeenpanoa joko ennakoivasti tai taannehtivasti. Päätetään aineiston säilytysmuodosta ja arvosta, erotellaan määräajan säilytettävät aineistot pysyvästi säilytettävistä sekä tuhotaan aineistot säilytysajan päättymisen jälkeen. Arvonmääritykseen vaikuttavat esim. eettiset, lainsäädännölliset ja taloudelliset kysymykset.
<b>Siirtokelpoinen tiedostomuoto</b>	Pysyvästi säilytettävien aineistojen siirtokelpoiset tiedostomuodot ovat sellaisia, joissa pitkäaikaissäilytettävää aineistoa on runsaasti tallennettu. Kaikki säilytyskelpoiset tiedostomuodot ovat myös siirtokelpoisia, mutta kaikki siirtokelpoiset eivät ole säilytyskelpoisia.
<b>SIP</b>	Submission Information Package. Yleensä viitataan OAI-viitemallin mukaiseen tiedonsiirtopakettiin.
<b>Skannata</b>	Skannerilla tapahtuvaa lukemista ja tunnistamista.
<b>Skanneri</b>	Laite, joka lukee kohdetta osio kerrallaan ja luo siitä digitaalisen kuvan tai kolmiulotteisen mallin tai tunnistaa koodin. Skannereissa käytetyt tekniikat vaikuttavat lopputulokseen. Yleensä skanneri lukee kohteen siitä heijastuvan tai sen läpäisevän valon avulla. Jos kohde on tekstiä, kuvanlukija muodostaa tekstistä kuvan. Jotta kuvan tekstiä voisi käsitellä, se täytyy erikseen muuntaa kirjoitusmerkeiksi tekstintunnistusohjelmalla. Skannereita voivat olla esim. tasoskanneri, filmiskanneri, dia- ja negatiiviskanneri, valokuvaskanneri, suurkuvaskanneri, mastoskanneri, rumpuskanneri, dokumentiskanneri.
<b>Sähköinen arkisto</b>	Tietojärjestelmä, jossa säilytetään aineistoja pitkään tai pysyvästi ja joka on suunniteltu takaamaan aineiston pysyvä säilyvyys. Järjestelmän avulla pidetään huolta aineistojen käyttöoikeuksista, suojataan ne asiattomilta muutoksilta, metatietojen avulla pidetään ne hallittavissa sekä huolehditaan aineistojen ja järjestelmän migraatioista.
<b>Säilyttäminen</b>	Säilyttäminen ja säilytysaika tarkoittavat sitä, miten kauan tietoa säilytetään siihen käyttötarkoitukseen, johon asiakirjat ja muut vastaavat tiedot on kerätty. Säilytysajat määrittelee viranomainen. Kun tietoaineiston tai siinä olevien asiakirjojen säilytysaika on päättynyt, ne tulee arkistoida tai tuhota viipymättä tietoturvalisella tavalla.
<b>Säilytyskelpoinen tiedostomuoto</b>	Säilytyskelpoisten tiedostomuotojen arvioidaan olevan käyttökelpoisia vielä pitkään. Tietosisällön säilyminen ja ymmärrettävyys voidaan taata, koska tiedostomuodon migraatioon on todennäköisimmin tekniset valmiudet.
<b>Säilytysmuoto</b>	Aineistolla tai asiakirjalla on digitaalinen ja/tai analoginen muoto, ilmentymä, joka voi muuttua elinkaaren eri vaiheissa.

<b>Käsite/termi</b>	<b>Selitys</b>
<b>Tekoäly</b>	Tekoäly (AI = Artificial Intelligence) ei ole yksi selkeä käsite, vaan erilaisten teknologioiden, sovellusten ja menetelmien ominaisuus. Tekoälyn määrä voi vaihdella ja sen ominaisuuksiin kuuluu kyky tehdä tehtäviä monimutkaisessa ympäristössä ilman jatkuvaa ohjausta ja kyky parantaa suorituskykyä oppimalla kokemuksesta.
<b>Tiedostomuoto</b>	Tiedostomuoto ilmaisee tiedoston rakenteen. Monista tiedostomuodoista on olemassa julkinen määritelmä (standardi), jossa kuvataan, miten kyseisessä formaatissa oleva informaatio on tallennettava digitaaliseen muotoon, esimerkiksi Portable Document Format (PDF), Tagged Image File Format (TIFF).
<b>Tiekartta</b>	Tiekartta (road map) voi edustaa suunnitelmaa tai hahmotelmaa. Tiekartalla kuvataan laajoja ja pitkällä aikavälillä toteutettavia suunnitelmia. Tiekartan tarkoitus on ohjata reitti haluttuun tulokseen tai päämäärään.
<b>Tietoturva</b>	Järjestelyt, joilla pyritään varmistamaan aineistojen ja tiedon saatavuus, eheys ja luottamuksellisuus. Tietoturvan järjestelyjä ovat esimerkiksi kulunvalvonta, tilojen lukitus, aineistojen turvallinen säilytys ja tuhoaminen. Tietoturvaan kuuluu muun muassa tietoaineistojen, laitteistojen, ohjelmistojen, tietoliikenteen ja toiminnan turvaaminen.
<b>Tuhoaminen</b>	Tuhoamisella tarkoitetaan sitä, että tietoaineisto on poistettava käytöstä siten, ettei sitä enää voida palauttaa uudelleen käyttöön. Tuhoamisella tarkoitetaan toimia, joilla tietoaineisto tosiasiaa tuhoutuu. Määräajan säilytettävät asiakirjalliset tiedot tuhoetaan vahvistettujen säilytysaikojen umpeuduttua.
<b>Validointi</b>	Tiedon luotettavuuden toteaminen tai kelpuutus, jota voidaan tehdä eri menetelmillä. Menetelmästä riippuen validoidaan koko aineistoa tai osaa aineistoista tai otos aineistosta. Menetelmiä voidaan luokitella eri tavoin esim. automatisoitu vs. manuaalinen; objektiivinen vs. subjektiivinen; rakenteellinen vs. sisällöllinen. Validoinnissa voidaan tarkistaa esim. kentän pituutta ja numeroarvorajoja, arvojen olemassaoloa, duplikaatteja. Säännöt ovat usein ehdollisia, eli tiedon validiteetti tarkistetaan, jos tietyt ehdot täyttyvät.

## Lähteet:

- Ahoranta, T., Haavisto, J., Kecskeméti, I. & Mäenpää, M. 2017. Kohti Kansallisarkiston asiakirjojen vapaaehtoisdigitointia. Blogi. Päivitetty helmikuussa 2017. Saatavissa: <http://nakokulmiaarkistosta.blogspot.com/2018/02/kohti-kansallisarkiston-asiakirjojen.html> [viitattu 16.11.2022].
- CSC – Tieteen tietotekniikan keskus Oy. 2019. Sanasto. WWW-dokumentti. Saatavissa: <https://digitalpreservation.fi/specifications/sanasto> [viitattu 10.11.2022].
- ESS. 2018. ESS Handbook. Methodology for data validation. CROS - European Commission. PDF-dokumentti. Saatavissa: [https://ec.europa.eu/eurostat/cros/content/ess-handbook-methodology-data-validation-version-20-revision-2018\\_en](https://ec.europa.eu/eurostat/cros/content/ess-handbook-methodology-data-validation-version-20-revision-2018_en) [viitattu 2.12.2022].
- Helsingin yliopisto s.a. Elements of AI. Mitä tekoäly on? WWW-dokumentti. Saatavissa: <https://course.elementsofai.com/fi/1/> [viitattu 2.12.2022].
- Hovi, A., Huotari, J. & Lahdenmäki, T. 2003. Tietokantojen suunnittelu & indeksointi. Docendo: Jyväskylä.
- Hänninen, T. 2019. Digitointiopas. Kaakkois-Suomen ammattikorkeakoulu. WWW-dokumentti. Saatavissa: <https://digitalia.xamk.fi/digiopas/sanasto> [viitattu 1.11.2022].
- Kansallisarkisto. 2013. Kuvailu. Arkistowiki. WWW-dokumentti. Saatavissa: <http://wiki.narc.fi/arkistowiki/index.php/Kuvailu> [viitattu 25.11.2022].
- Kansallisarkisto s.a. Arvonmääritys ja seulonta, keskeiset käsitteet. WWW-dokumentti. Saatavissa: <https://kansallisarkisto.fi/fi/viranomaisille/arvonmaaritys-ja-seulonta/kasitteet> [viitattu 1.11.2022].
- Kansallisarkisto. 2022. Digitointi. WWW-dokumentti. Saatavissa: <https://kansallisarkisto.fi/fi/viranomaisille/digitointi> [viitattu 25.11.2022].
- Kaukovuo, H. 2022. Yhdenmukainen tiedon validointi käyttöliittymässä ja mikropalvelussa. Jyväskylän yliopisto. Kandidaatintutkielma. PDF-dokumentti. Saatavissa: <http://urn.fi/URN:NBN:fi:juu-202201051007> [viitattu 2.12.2022].
- Klemettinen, R. 2016. Pysytkö kartalla? Kielikello 2/2016. Verkkolehti. Saatavissa: <https://www.kielikello.fi/-/pysytko-kartalla-> [viitattu 16.11.2022].
- Koivusalo, S. 2021. Miksi markkinavuoropuhelu on niin tärkeää ”työllistämisen ehdon käyttämisessä” – oppeja käytännön työstä. PDF-dokumentti. Saatavissa: <https://www.kuntaliitto.fi/sites/default/files/media/file/Hankinnoilla%20ty%C3%B6llist%C3%A4misen%20vauhditusohjelma%20seminaari%2C%20miksi%20markkinavuoropuhelu%20on%20niin%20t%C3%A4rke%C3%A4%2010.11.2021%2C%20Koivusalo.pdf> [viitattu 16.11.2022].
- Kuntaliitto, Kuntasektorin arkkitehtuuriryhmä. 2016. Sanasto. PDF-dokumentti. Saatavissa: <https://docplayer.fi/23520265-Liite-1-asianhallinnan-sanasto.html> [viitattu 1.11.2022].
- Lybeck, J. ym. 2006. Arkistot yhteiskunnan toimiva muisti. Asiakirjahallinnon ja arkistotoimen oppikirja. Arkistolaitos. Saatavissa: [https://www.arkisto.fi/uploads/Julkaisut/oppaat/asiakirjahallinnon\\_oppikirja.pdf](https://www.arkisto.fi/uploads/Julkaisut/oppaat/asiakirjahallinnon_oppikirja.pdf) [viitattu 30.11.2022].
- OAIS s.a. OAIS Reference model (ISO 14721), fundamental standard for digital preservation. WWW-dokumentti. Saatavissa: <http://www.oais.info/> [viitattu 25.11.2022].
- Pölkki, T. 2022. Digitaalisuus ilman muutosta on turhaa. Blogi. Päivitetty 11.3.2022. Saatavissa: <https://blogit.jamk.fi/turbinaattori/2022/03/11/digitaalisuus-ilman-muutosta-on-turhaa/> [viitattu 16.11.2022].
- Suominen, V., Saarti, J. & Tuomi, P. 2009. Bibliografinen valvonta. Johdatus luetteloinnin ja sisällönkuvailun menetelmiin. BTJ Kustannus: Helsinki.
- Sanastokeskus. 2005. TEPA-termipankki. Erikoisalojen sanastojen ja sanakirjojen kokoelma. Saatavissa: <https://termipankki.fi/tepa/fi/> [viitattu 30.11.2022].
- Valoro Oy. 2022. Asiakirjahallinta – millä sanoilla siitä puhumme? WWW-dokumentti. Saatavissa: <https://valoro.fi/palvelut/asiakirjahallinta-sanasto/> [viitattu 1.11.2022].
- Valtiovaraministeriö. 2022. Suositus tietoaineistojen säilytysajasta ja toimenpiteistä säilytysajan päätyttyä. Valtiovaraministeriön julkaisuja 2022:54. Saatavissa: <https://urn.fi/URN:ISBN:978-952-367-220-8> [viitattu 2.12.2022].

# Kiitokset

Kiitämme erityisesti oppaassa esiteltyjen case-esimerkkien tarjoajia Päivälehdien arkistoa, Lahden kaupunkia, Turun kaupunkia, Mikkelin kaupunkia ja Miehikkälän museoita sekä muita kartoitukseen vastanneita tahoja, joilta saimme arvokkaita esimerkkejä ja kokemuksia. Kiitämme myös asiantuntija-artikkeleiden kirjoittajia Kansallisgalleriasta, Kansallisarkistosta, Kansalliskirjastosta ja Kaakkois-Suomen ammattikorkeakoulusta. Kiitos Digitoinnin taitajaksi -koulutuksen opiskelijoille arvokkaista keskusteluista ja kiinnostuksestanne digitoinnin osaamisen kehittämistä kohtaan.

