



Euroopan unioni
Euroopan aluekehitysrahasto

Vipuvoimaa
EU:lta
2014–2020



Digitalia
Digitaalisen tiedonhallinnan
tutkimus- ja kehittämiskeskus

Digitointiopas

Tämä on pdf-versio julkaisusta <https://digitalia.xamk.fi/digiopas>

Tekijä: Teemu Hänninen
Palaute julkaisusta: etunimi.sukunimi(at)xamk.fi
Julkaisusarja: Xamk Kehittää 82
Julkaisija: Kaakkois-Suomen ammattikorkeakoulu, Mikkeli
Julkaisuvuosi: 2019
ISBN: 978-952-344-184-2 (PDF)
ISSN: 2489-3102 (verkkopublication)

Abstrakti: Digitoinnilla tarkoitetaan analogisen aineiston muuntamista digitaaliseen muotoon. Aineiston digitointi parantaa sen säilyvyyttä, saatavuutta ja käytettävyyttä, koska digitoituihin aineistoihin pääsee nopeammin käsiksi kuin paperiseen aineistoon ja siihen on helpompi kohdistaa hakuja. Digitaalinen arkisto eli digiarkisto ei eroa tavoitteiltaan tavanomaisesta fyysisestä arkistosta. Ainoastaan tiedon esitysmuoto on erilainen: digiarkistossa tieto säilytetään tiedostoina jollakin digitaalisella tallennusvälineellä, kuten tietokoneen kovalevyllä, kun perinteisessä fyysisessä arkistossa tieto säilytetään paperisissa asiakirjoissa ja muilla analogisilla tallennusmedioilla. Tässä oppaassa kerrotaan lyhyesti, kuinka digitoidaan ja millä menetelmillä turvataan digitaalisen tiedon pitkäaikainen säilyminen.

Asiasanat: digitointi, pitkäaikaissäilytys, sähköinen arkistointi, digitalisaatio

Digitointiopas

Digitalisaation myötä yksityishenkilöt, yritykset ja yhdistykset ovat alkaneet nähdä sähköisen pitkäaikaissäilyttämisen tärkeyden. Kansalaisten, pienyhteisöjen, järjestöjen ja sukujen historiallisten aineistojen säilyttäminen onkin arvokas tavoite, jonka toteuttamiseksi tarvitaan digiaikana uusia tietoja ja taitoja (Hänninen 2019). Tämä opas tarjoaa helposti lähestyttävän tietopaketin digitoinnin ja pitkäaikaissäilyttämisen perusteisiin.

Johdanto digitointiin ja digiarkistointiin

Nykyään yhä suurempi osa tiedonvälityksestä ja tiedon tallentamisesta tapahtuu sähköisesti. Otamme digikuvia älypuhelimilla, lähetämme sähköposteja ja kirjoitamme blogitekstejä. Kaikki tämä tieto on digitaalisessa muodossa jossakin tallennusvälineellä ykkösinä ja nollina. Emme kuitenkaan aktiivisesti pyri turvaamaan kaiken tämän tiedon pitkäaikaista säilymistä, mikä saattaa johtaa siihen, että suuri osa tiedosta katoaa. Samalla katoaa merkittävä osa kulttuuriperintöä. Tästä syystä tiedon pitkäaikaissäilytys eli arkistointi on tärkeä ja tarpeellinen pyrkimys, jonka aloittamiseen vaadittaviin toimenpiteisiin pystyy kuka tahansa.

Arkisto ja arkistointi ovat siinä mielessä hassuja käsitteitä, että jokaisella on varmaankin jonkinlainen mielikuva siitä, mitä ne tarkoittavat. Arkistoon liitetään ihmiselle tyypillisen intuitiivisesti mielikuva pölyisestä kammioista, missä lukemattomat asiakirjoja sisältävät hyllyt jatkuvat silmäkantamattomiin. Tällainen homeinen mielikuva ei välttämättä ensimmäisenä herätä suurta ihastusta arkistointia kohtaan. Arkistolla on kuitenkin tärkeä tarkoitus: arkistoimalla jokainen meistä voi jättää itsestään jäljen, joka säilyy pitkiä aikoja ja on jälkipolvien saatavilla. Arkistossa säilötään historiaa, jonka avulla yksilö tai organisaatio voi kertoa itsestään tarinan. Arkiston olennaisia ominaisuuksia ovat seuraavat:

- Arkistossa säilytettävät asiakirjat muodostavat loogisen kokonaisuuden, mikä edesauttaa tiedon saatavuutta ja käytettävyyttä.
- Arkiston tarkoitus on säilyttää tietoa pitkiä aikoja, mistä on tulevaisuudessa hyötyä sekä arkistonmuodostajalle että tutkijoille.
- Arkistossa säilytettävä tieto on arkistonmuodostajan toiminnan yhteydessä syntynyttä toiminnallista tietoa.

Jos aivan tarkasti lain mukaan määritellään, niin arkistolaissa sanotaan seuraavaa: "arkistoon kuuluvat asiakirjat, jotka ovat saapuneet arkistonmuodostajalle sen tehtävien johdosta tai syntyneet arkistonmuodostajan toiminnan yhteydessä". Arkistossa säilytettävä tieto on siis arkistonmuodostajalle ainutkertaista ja yksilöllistä.

Digitoinnilla tarkoitetaan analogisen aineiston muuntamista digitaaliseen muotoon. Esimerkiksi vanhoja valokuvia voidaan digitoida skannaamalla ja tallentamalla ne tietokoneelle digitaalisina tiedostoina. Digitaalinen arkisto eli digiarkisto ei eroa tavoitteiltaan tavanomaisesta fyysisestä arkistosta. Ainoastaan tiedon esitysmuoto on erilainen: digiarkistossa tieto säilytetään tiedostoina jollakin digitaalisella tallennusvälineellä, kuten tietokoneen kovalevyllä, kun perinteisessä fyysisessä arkistossa

tieto säilytetään paperisissa asiakirjoissa ja muilla analogisilla tallennusmedioilla, kuten vaikkapa vanhoilla videonauhoilla.

Digitointi tuo monia hyötyjä. Aineiston digitointi parantaa sen säilyvyyttä, saatavuutta ja käytettävyyttä, koska digitoituihin aineistoihin pääsee nopeammin käsiksi kuin paperiseen aineistoon ja siihen on helpompi kohdistaa hakuja. Digitointi myös tehostaa viranomaisten toimintaa.

Digitaalinen aineisto ei kulu käytössä samalla tavalla kuin analoginen aineisto. Esimerkiksi kirjat ja valokuvat kuluvat ja haalistuvat ajan saatossa, kun ajan hammas puree hitaasti mutta varmasti jälkensä. Lisäksi digitaalinen arkisto voi samalla toimia varmuuskopiona alkuperäiselle aineistolle.

Digitaaliset tiedostot ovat kuitenkin riippuvaisia laitteistosta ja ohjelmistoista. Paperisia aineistoja saatetaan arkistoida sadoiksi vuosiksi, mutta sähköisessä arkistossa puhutaan kymmenien vuosien elinkaaresta, koska laitteistot ja ohjelmistot muuttuvat nopealla tahdilla. Siksi oikean tiedostomuodon valitseminen on tärkeää, koska tiettyjen tiedostomuotojen avaamiseen tarvittavia välineitä ei välttämättä ole saatavilla kymmenenkään vuoden kuluttua. Tästä syystä sähköistä arkistoa pitää myös uudistaa tasaisin väliajoin niin, että digitaaliset tiedostot muutetaan sillä hetkellä käytössä olevien standardien mukaisiksi.

Jokaisen digitaalisiin aineistoihin panostavan yksilön ja organisaation on järkevää panostaa myös aineiston pitkäaikaissäilytykseen. Sähköistä arkistointia ei voida pitää ainoastaan museoiden, kirjastojen ja muiden arkistointiorganisaatioiden vastuualueena, vaan se on haaste kaikille tahoille, niin yksityishenkilöille kuin organisaatioillekin, jotka ovat kiinnostuneita digitaalisten aineistojen kehittämisen, ylläpitämisen ja saatavuuden varmistamisesta.

Arkistossa tieto on jäsennelty ja säilytty loogisena kokonaisuutena.

Arkiston tarkoitus on säilyttää tietoa pitkiä aikoja, mistä on tulevaisuudessa hyötyä sekä arkistonmuodostajalle että tutkijoille.

Digitointi tarkoittaa paperisen aineiston muuntamista digitaaliseen muotoon

Esimerkiksi vanhoja valokuvia voidaan digitoida skannerilla.

Digitointi tuo monia hyötyjä

Tiedon saatavuus helpottuu ja siihen voidaan kohdistaa hakuja.

Kuinka digitoidaan ja arkistoidaan

Digitointia voi tehdä itse kotona, ja lisäksi useat kirjastot tarjoavat kansalaisille maksuttoman mahdollisuuden digitoida henkilökohtaisia aineistoja. Mikäli digitoitavaa aineistoa on paljon tai tekninen toteutus vaikuttaa liian haastavalta, voi digitointipalvelun myös ostaa siihen erikoistuneelta yritykseltä. Yksityishenkilö pystyy usein toteuttamaan oman digitointiprojektin ilman ulkopuolista palveluntarjoajaa, koska aineiston määrä on suhteellisen pieni. Yhdistysten ja pienten yritysten digitointiprojekti saattaa kuitenkin vaatia useiden eri osapuolten yhteistyötä, ja varsinainen digitointi on vain yksi osa kokonaisuutta. Digitointiin kuuluu huolellista suunnittelua, valmistelua, testausta, varsinaista digitointia, tiedostojen käsittelyä ja jatkuvaa laadunvalvontaa. (Laustela 2013)

Digitoinnissa kannattaa myös ottaa huomioon se, onko digitoinnin tarkoituksena ainoastaan luoda kopiot alkuperäisestä aineistosta vai halutaanko samalla varmistaa tiedostojen pitkäaikainen säilyminen. Tähän tarkoitukseen voidaan muodostaa oma digiarkisto, tai tallentaa tiedostot johonkin valmiiseen arkistopalveluun. Digitointi onkin aina järkevää tehdä arkistonäkökulmasta ja tähdätä siihen, että digitaalisten tiedostojen pitkäaikainen säilyminen varmistetaan.

Suunnitteluvaiheessa tehdään **digitointisuunnitelma**. Suunnitelmassa otetaan huomioon kaikki projektin osat. Suunnittelun jälkeen ensimmäinen työvaihe on aineiston valmistelu digitointia varten. Valmistelussa aineisto seulotaan ja luetteloidaan. Valmisteluun kuuluu myös aineiston fyysisen kunnon varmistaminen digitointiin sopivaksi. Aineistosta siis poistetaan haitalliset esineet, kuten klemmarit, teipit ja niitit yms. Kansioissa ja muovitaskuissa säilytetyt asiakirjat puretaan. Huonokuntoiset asiakirjat merkataan ja niille valmistellaan sopivat digitointimenetelmät, jotka eivät vahingoita asiakirjoja.

Aineiston käsittelyn aikana tulee huomioida myös työergonomia ja **työturvallisuus**. Vanhoissa paperisissa aineistoissa saattaa esiintyä homeetta, joka on terveydelle haitallista. Homeisten asiakirjojen käsittelyssä tulee käyttää asianmukaisia välineitä, kuten hengityssuojaimet ja suojakäsineet. Homeisten aineistojen käsittelystä tarjoo lisää tietoa Kansalliskirjaston ohjeet home-epäilyyn tai homeisen aineiston käsittelyyn.

Ennen kuin dokumentteja aletaan työntää skanneriin, niin digitointivälineet tulee testata ja kalibroida. Näiden työvaiheiden jälkeen aloitetaan varsinainen digitointi. Digitoinnin tuloksena syntyvien tiedostojen laatua tarkkaillaan koko työvaiheen ajan, ja mahdolliset laatuvirheet korjataan. Yksi työläimmistä vaiheista on tiedostojen metatietojen luominen. Osa metatiedoista pystytään luomaan automaattisesti, mutta myös manuaalista metatietojen syöttämistä tarvitaan. Tiedostot nimetään kuvailevasti ja tallennetaan oikeissa formaateissa. Seuraavassa taulukossa on yksinkertainen muistilista suositelluille tiedostoformaateille aineiston tyyпин mukaan. Tiedot perustuvat KDK-PAS -palvelun määrityksiin suositelluista tiedostomuodoista.

Aineiston tyyppi	Tiedostomuoto
Tekstidokumentit	PDF/A-1,2,3
Kuvat	TIFF, JP2 (JPEG 2000)
Video	JPEG2000, MPEG4
Ääni	WAV, BWF

Digitoinnissa on tärkeää, että alkuperäisen aineiston tieto säilyy muuttumattomana, kun se muutetaan digitaaliseen muotoon. Aineiston digitointia aloitettaessa ensimmäisiä kysymyksiä on varmaankin se, missä järjestyksessä alkuperäistä aineistoa pitäisi alkaa digitoida. Tässä on hyvä priorisoida aineisto sen fyysisen kunnan perusteella, eli kaikkein huonokuntoisin aineisto, joka lähestyy elinkaarensa loppua, kannattaa digitoida ensimmäisenä (Mäkelä 2017). Digitoitava aineisto saattaa nimittäin olla hyvin huonossa kunnossa, kuten vaikkapa vanhat kirkonkirjat. Sivut saattavat olla hauraita ja lahonneita, joten niiden käsittely vaatii huolellisuutta, mikä puolestaan hidastaa digitointiprosessia. Usein voi olla aiheellista antaa konservointiasiantuntijan arvioida aineisto ennen digitointia, jotta saadaan selville, tarvitseeko aineistoon kohdistaa joitakin toimenpiteitä sen säilymisen turvaamiseksi. Luonnollisesti myös aineiston samankaltaisuus nopeuttaa digitointia, koska tarvitaan vähemmän erilaisia laitteita ja ohjelmia.

Paperisen aineiston, kuten dokumenttien ja kuvien, digitointiin on käytännössä kaksi vaihtoehtoa: valokuvaaminen tai skannaaminen. Molemmilla menetelmillä on mahdollista tuottaa hyvälaatuisia ja tarkkoja digitaalisia jäljenteitä alkuperäisestä aineistosta. Eroja löytyy kuitenkin käytettävyydestä, kustannuksista ja siinä, kuinka paljon aikaa operaatio vaatii. Dokumenttien valokuvaamiseen tarvitaan hyvä valaistus, jonka asettelu vaatii aika ja vaivaa, kun taas skannerilla tätä seikkaa ei tarvitse ottaa huomioon. Lisäksi kamera ja dokumentin alusta täytyy asettaa tukevasti niin että ne eivät pääse liikkumaan kuvauksen aikana. Valokuvaamisen etuna on se, että se on skannaamista hienovaraisempi menetelmä. Useimmissa älypuhelimissa on riittävän hyvä kamera dokumenttien digitointia varten, tai tarvittaessa voi ostaa järjestelmäkameran, joiden hinnat alkavat noin 400 eurosta.

Skannerit maksavat halvimmillaan alle sata euroa, ja tavanomainen skanneri on riittävän hyvä henkilökohtaiseen digitointiin. Skannerin asentaminen varsinkin Windows-ympäristössä onnistuu helposti USB-liittimen kautta. Useimmiten ei tarvitse kuin kytkeä skannerin USB-kaapeli tietokoneen USB-porttiin, minkä jälkeen Windowsin asetuksista etsitään kohta "laitteet" ja valikon kohdasta "lisää tulostin tai skanneri". Windows etsii saatavilla olevia skannereita, ja kun listaan on löytynyt haluttu skanneri, valitaan se ja klikataan "lisää laite".

Digitointimenetelmä kannattaa valita sen perusteella, millaista aineistoa aiotaan digitoida. Valokuvaaminen sopii kuluneille ja hauraille asiakirjoille, jotka eivät välttämättä säilyisi ehjinä skannerin läpi. Skanneri on kätevä vaihtoehto hyväkuntoisille dokumenteille. Valaistukseen sopivat vähän lämpöä tuottavat valonlähteet, koska perinteiset hehkulamput tuottavat suuria määriä lämpöä, joka on haitallista paperille ja musteelle.

Skannauksen yhteydessä on joskus mahdollista valita tiedostomuoto, jossa kuva tallennetaan. Mikäli valittavissa on TIFF-muoto, on se paras vaihtoehto kuvien pitkäaikaissäilytystä varten. Yleisimmin käytetyt tiedostomuodot ovat JPEG ja TIFF, joista

ensin mainittu on laadultaan heikompi, mutta toisaalta vie vähemmän tilaa tietokoneen kovalevyiltä. Yleisesti voidaan sanoa, että TIFF on pitkäaikaissäilyttämistä silmällä pitäen parempi valinta. Tiedostomuodoista voi lukea lisää täältä.

Yleensä kuvat digitoidaan 24-bittisinä (väridigitointi) ja tekstidokumentit 8-bittisinä (harmaasävydigitointi). Varsinkin väridigitoinnissa on tärkeää, että digitaalisen jäljenteen värimaailma vastaa mahdollisimman todenmukaisesti alkuperäisen kuvan värimaailmaa.

Kuvien skannaamisessa olennaista on kuvan resoluutio, joka ilmoitetaan arvoilla PPI (Pixels Per Inch) tai DPI (Dots Per Inch), eli pikseleiden tai pisteiden määrällä tuumaa kohti. Käytännössä suurempi resoluutio tarkoittaa tarkempaa kuvaa. 300 PPI on riittävän suuri resoluutio useimmille dokumenteille, jotka ovat vähintään kokoa A4. Arkistolaitos suosittelee, että tavallista A4-kokoista paperiarkkia pienemmille dokumenteille käytetään suurempaa resoluutiota kuin 300 PPI. Tällaisia pieniä dokumentteja voivat olla esimerkiksi kuitit ja diakuvat. Suurempi resoluutio on tarpeellinen siksi, että pienet dokumentit saadaan tarvittaessa suurennettua alkuperäistä koko suuremmiksi vaikkapa tietokoneen näytöllä, ja suuremman resoluution ansiosta dokumentit pysyvät tarkkuudeltaan luettavina.

Oheisessa taulukossa on listattu Arkistolaitoksen (nyk. Kansallisarkisto) käyttämiä resoluutioita eri kokoisille dokumenteille. Vaatimuksiin voi perehtyä tarkemmin osoitteessa <https://www.arkisto.fi/uploads/normit/valtionhallinto/suosituksset/Arkistolaitoksen%20vaatimukset%20digitoidulle%20aineistolle.pdf>.

Originaalin koko	Resoluutio
6x9 cm	1600 ppi
9x12 cm	1200 ppi
13x18 cm	900 ppi
18x24 cm	600 ppi
A3 (tai suurempi)	300 ppi

Kun dokumentit on muunnettu digitaaliseen muotoon, täytyy niistä vielä erotella teksti. Skannaaminen ja valokuvaaminen nimittäin tuottavat digitaalisia kuvatiedostoja, eli tietokone ei ymmärrä, että kuvat sisältävät tekstiä, vaikka ihmiselle teksti olisikin luettavissa ja ymmärrettävissä. Alkuperäisen aineiston tekstin tallentaminen digitaaliseen muotoon on elintärkeää arkiston hyödyntämisen kannalta, koska tällöin tekstiä voidaan kopioida ja sen perusteella aineistoon voidaan kohdistaa hakuja.

Tekstintunnistus eli OCR (Optical character recognition) on menetelmä, jossa digitaalisista kuvista etsitään kirjoitettua sisältöä, joka tunnistuksen jälkeen tallennetaan tekstiksi. Tekstintunnistamista varten on tärkeää, että skannauksen aikana skannattavien dokumenttien kaikki sisältö on skannerin havaittavissa. Eli dokumentin sivuissa ei saa olla esimerkiksi taitettuja kulmia, joiden alle saattaa jäädä piiloon sisältöä. OCR-palveluja on

saatavilla ilmaisina verkkosovelluksina, kuten Online OCR osoitteessa <https://www.onlineocr.net/>.

Äänen ja kuvan digitointiin tarvitaan sellainen laite, joka osaa lukea tietoa alkuperäiseltä tallenteelta. Esimerkiksi vanhojen VHS-nauhojen toistamiseen käytettävä videonauhuri pitää kytkeä tietokoneeseen, joka pystyy tallentamaan sisällön digitaalisessa muodossa tietokoneen kovalevyille. Vanhoja videonauhureita löytyy läjäpäin esimerkiksi kirpputoreilta, mikäli sellaista ei löydy omasta varastosta. Video- ja äänisignaalin siirtämistä varten löytyy markkinoilta videosieppareita, joka liitetään videonauhurin analogisesta ulostuloliitännästä tietokoneen USB-liitäntään.

Videosieppareiden hinnat liikkuvat halvimmillaan noin 40 euron paikkeilla, ja mukana toimitetaan yleensä myös aineiston toistamiseen, käsittelyyn ja tallentamiseen tarvittava ohjelmisto. Ohjelmistot ovat useimmiten helppokäyttöisiä ja tarjoavat myös lisäominaisuuksia videon editointiin.

Äänitteiden, kuten vanhojen C-kasettien ja vinyylilevyjen, digitointi onnistuu samalla tavalla kuin VHS-nauhojen digitointi. Helppimmillaan ei tarvita erillistä siepparia tai adapteria, vaan alkuperäisen äänitteen toistamiseen käytettävä laite liitetään kaapelilla suoraan tietokoneen äänikortin sisääntuloliitännään (Line in), joka mahdollistaa ulkoisen äänilähteen signaalin kaappaamisen. Äänikorttien liitännät on värikoodattu, ja Line in on yleensä sininen liitäntä. Äänen käsittelyyn ja tallentamiseen on saatavilla ilmaisia ohjelmistoja, kuten Audacity.

Tiedonvälityksestä suuri osa tapahtuu sähköisesti esimerkiksi sähköpostin ja sosiaalisen median välityksellä. Varsinkin sähköpostissa saattaa olla runsaasti sellaisia dokumentteja ja viestejä, jotka olisi hyvä saada arkistoitua. Sähköpostien arkistointiin ei vielä ole olemassa mitään yhteistä käytäntöä. Yksi tapa on muuttaa jokainen yksittäinen sähköpostiviesti arkistokelpoiseksi PDF-tiedostoksi, mikä kuitenkin vie aikaa ja resursseja. Microsoftin Outlook-sähköpostiohjelmasta pystyy tallentamaan sähköpostit erillisenä pst-tiedostona, jonka avaamiseen kuitenkin aina vaaditaan Outlook, ja tästä syystä se ei ole ihanteellinen tapa arkistoida sähköposteja. Sähköpostien arkistoinnissa ongelmia aiheuttavat myös sähköpostien liitetiedostot, koska ne voivat olla missä formaatissa tahansa. Digitalialla on kehitteillä sovellus, joka tekee pst-tiedoston sisältämistä sähköpostiviesteistä yksittäisiä arkistokelpoisia PDF/A-tiedostoja. Sovellusta voi testata osoitteessa <https://digitalia.xamk.fi/email-converter#!/>.

Digiarkiston muodostaminen

Arkistoa suunniteltaessa voidaan hieman soveltaen noudattaa esimerkiksi Marc Freskon neljää ohjetta (Rännäli 2015):

- Selvitetään ensin, mitä säilytetään, eli aineistolle suoritetaan **seulonta**.
- Selvitetään, kuinka kauan aineistoa säilytetään.
- Selvitetään, kuinka paljon resursseja käytetään aineiston säilyttämiseen ja kuinka paljon siitä maksetaan.
- Päätetään, mitkä ovat ne toimintatavat, joilla nämä tavoitteet saavutetaan.

Jos organisaation tiedonhallinta on retuperällä, niin aineistoa saattaa olla tallennettuna useaan eri pilvipalveluun, kuten Google Drive ja Office 365, eikä aineistoa ole mitenkään

jäsennely tai luokiteltu. Tällaisten aineistojen arkistointi voi osoittautua haastavaksi, koska ei ole tietoa mistä ja kuinka pitäisi lähteä liikkeelle. Kaikkea aineistoa ei välttämättä tarvitse tai edes haluta arkistoida, joten ensimmäisenä on lähdeettävä seulomaan aineistoa, jotta voidaan päättää, mitä halutaan arkistoida. Mikäli samaa tietoa on tallennettuna useina eri kopioina, valitaan aina se tiedosto, joka on laadultaan paras.

Eri aineistoille voi olla vaikeaa määritellä arvoa, mikä tekee seulonnasta haastavaa.

Yrityksille arvonnäilytyksessä on hyvä noudattaa yleistä ohjetta, jonka mukaan hävitetäväksi kelpaavaa aineistoa on sellainen aineisto, johon ei sisälly todistusarvoa, ja arkistoidaan vain oman toiminnan tuloksena syntynyt aineisto. Lisäksi on hyvä muistaa, että samasta dokumentista ei tarvitse arkistoida useita kappaleita.

Varsinkin kuvia saattaa olla useita samankaltaisia, koska kännykkäkameroiden yleisyys ja helppo käytettävyys on lisännyt valokuvaamisen helppoutta, joten kuvista kannattaa valita arkistoitavaksi vain parhaiten onnistuneet otokset. Seulonnan tuloksena aineiston koko pienenee huomattavasti, mikä puolestaan säästää aikaa ja kustannuksia. Kannattaa kuitenkin pohtia myös aineiston tulevaa käyttötarkoitusta. Jos esimerkiksi kuvia tai dokumentteja aiotaan jossain vaiheessa muokata, niin kuvista kannattaa arkistoida mahdollisimman korkealaatuiset ja pakkaamattomat versiot, minkä lisäksi luodaan niin sanottu käyttökopio, joka on tiedostokooltaan pienempi. Käyttökopiot on tarkoitettu tiedoston jakamiseen ja muuhun käyttöön. Dokumenteista on järkevää ottaa talteen sekä arkistokelpoinen PDF-tiedosto että muokattavissa oleva alkuperäinen tekstitiedosto, joka voi olla vaikkapa perinteinen Word-tiedosto.

Seulonnassa kannattaa käyttää apuna esimerkiksi Yksityisten Keskusarkistojen laatimaa erinomaista listaa, joka löytyy osoitteesta www.yksityisetkeskusarkistot.fi/.

YKA:n listan mukaan säilytettävää aineistoa ovat:

- pöytäkirjat/muistiot sekä pöytäkirjojen liitteet, esimerkiksi tilinpäätökset ja talousarviot
- toimintakertomukset ja -suunnitelmat
- lähteneiden kirjeiden kopiot
- puheet, esitelmät, artikkelit, luennot
- itse tuotetut tutkimukset, raportit, yhteenvedot, tilastot
- luettelot (esimerkiksi jäsen- ja toimihenkilöluettelot)
- säännöt, rekisteröinti- ja organisaatiouudistuksiin liittyvät asiakirjat
- sopimukset
- omien kurssien, koulutusten, juhlien, tapahtumien asiakirjat

- valokuvat, diat, filmit, videot ja muut tallenteet
- itse tuotetut painotuotteet (lehdet, kalenterit, julisteet, mainokset, esitteet, kartat, piirustukset)
- saapuneet kirjeet ja asiakirjat, jotka ovat aiheuttaneet toimenpiteitä

Seuraavia aineistoja ei säilytetä pysyvästi:

- Saapuneet kiertokirjeet, yleiskirjeet, tiedotteet ja "jakelussa mainitut" -kirjeet
- Tiedoksi saapuneet pöytäkirjat tai muut asiakirjat
- Lähetekirjeet, joissa ei ole olennaista tietoa
- Lyhyet viestit, ilmoitukset, mainokset
- Tiliotteet ja tositteet

Oma arkisto ei välttämättä tarkoita sitä, että aineiston pitäisi sijaita jossakin ulkoisen palveluntarjoajan omistamassa arkistojärjestelmässä. Yksityishenkilön, pienen yrityksen tai yhdistyksen oma digiarkisto voi sijaita esimerkiksi tietokoneen kovalevyllä tai jossakin pilvipalvelussa johdonmukaisesti järjestetyissä kansioissa. Olennaista ja tärkeintä on se, että tiedostojen pitkäaikainen säilyminen turvataan. On kuitenkin hyvä tiedostaa, että omalla tietokoneella tai pilvipalvelussa sijaitseva digiarkisto ei täytä kaikkia "oikeiden arkistojen" kriteereitä. Esimerkiksi tietoturva ja tiedostojen eheyden valvonta on puutteellista oikeisiin arkistoihin verrattuna. Arkistosta pitäisi myös olla varmuuskopio tallennettuna johonkin toiseen järjestelmään, jotta vältetään laitteiston fyysisten vahinkojen aiheuttama tietojen katoaminen tai vaurioituminen.

Arkistolle suunnitellaan ja toteutetaan loogisesti ja kuvaavasti nimetty kansiorakenne. Mikäli fyysinen arkisto on jo entuudestaan olemassa, niin digiarkiston muodostaminen helpottuu, koska voidaan käyttää samaa arkistokaavaa kuin alkuperäisessä arkistossa. Jos digiarkisto muodostetaan tyhjästä, niin apuna voidaan käyttää erilaisia arkistokaavoja. Helpoimmillaan kuitenkin arkiston kansiorakenne voidaan järjestellä esimerkiksi asiakirjojen tyyppin mukaan. Toinen tapa on luoda kansiorakenne niiden tehtävien mukaan, joihin asiakirjat liittyvät.

Tiedostot pitää myös nimetä järkevästi, joten tiedostojen nimissä kannattaa käyttää esimerkiksi päivämääriä ja sisältöä kuvailevia termejä. Tiedostojen nimissä ei kannata käyttää välilyöntejä, ääkkösiä (Å, Ä ja Ö) eikä erikoismerkkejä. Välilyönnit voi korvata alaviivalla (_) ja esimerkiksi ä-kirjaimet korvataan a-kirjaimilla. Metatiedot määritellään tiedoston luomisen yhteydessä.

Jos arkisto halutaan luovuttaa jollekin ulkopuoliselle palveluntarjoajalle, niin luovuttajan ja vastaanottajan kesken sovitaan siirtopaketin normista. Suomessa julkisen sektorin määrittäminen on Sähke 2. Tätä arkistolle luovutettavaa pakettia kutsutaan luovutuspaketiksi (SIP, Submission Information Package).

OAIS-standardissa (Open Archival Information System) määritellään siirtopaketin lisäksi myös säilytyspaketti (AIP, Archival Information Package) ja jakelupaketti (DIP, Dissemination Information Package). OAIS-mallista voi lukea lisää esimerkiksi osoitteessa <https://www.fsd.uta.fi/lehti/fi/30/OAIS.html>. Käytännössä yksityishenkilön tai yrityksen tarvitsee huolehtia vain siitä, että luovutuspaketti on arkiston määrittämisen mukainen.

Digitointia voi tehdä itse kotona tai ostaa palveluna

Tavanomaisella skannerilla on mahdollista digitoida paperisia dokumentteja. Myös kirjastot tarjoavat ilmaiseksi digitointipalveluja.

Kameralla pystyy digitoimaan dokumentteja ja kuvia

Aina ei tarvita kalliita laitteita, ja perinteiset älypuhelimet ovat usein riittävän hyviä yksityisen henkilön digitointitarpeisiin. Valokuvaaminen on kuitenkin hieman skannaamista työläämpi menetelmä.

Kuvatiedostoista täytyy erikseen tunnistaa ja erotella teksti

Skannaus ja valokuvaaminen tuottavat kuvatiedostoja, joista tietokone ei automaattisesti havaitse tekstiä. Tekstintunnistaminen eli OCR-tekniikka (Optical Character Recognition) tutkii kuvatiedostoja ja etsii niistä tekstiä.

Tiedostomuodot ja tallennusmediat

Tiedostomuodot voidaan jakaa säilytyskelppoihin (eli arkistokelpoihin) ja siirtokelpoihin tiedostomuotoihin. Molempien tiedostomuotojen ominaisuuksiin kuuluu se, että ne ovat laajasti käytössä, minkä lisäksi säilytyskelppoiset tiedostomuodot ovat myös vapaasti käytettävissä, eli niiden käyttämiseen ei tarvita kaupallisia oikeuksia (toisin sanoen niiden käyttö on ilmaista). Siirtokelpoisista tiedostomuodoista voidaan tuottaa myös säilytyskelppoinen tiedostomuoto.

Oikein valittujen tiedostomuotojen tärkeyttä ei voi aliarvioida luotettavan pitkäaikaissäilytyksen näkökulmasta. Kaikille aineistotyypeille ei ole vielä vakiintunutta arkistokelpoista tiedostomuotoa, ja eri arkistoilla on erilaisia suosituksia.

Esimerkiksi Kansalliset pitkäaikaissäilytyspalvelut (PAS) asettaa tiedostomuodoille seuraavia vaatimuksia:

- Tiedostoissa EI SAA käyttää salasanasuojauksia eikä mitään muita salaustekniikoita. Näiden käyttäminen hankaloittaa aineistojen pitkäaikaista säilyttämistä, ja joissakin tapauksissa jopa estää sen.
- Tiedostoissa EI SAA käyttää DRM (Digital Rights Management) -tekniikoita. Nämä voivat aiheuttaa ongelmia jopa bittitason säilyttämiselle.
- Tiedostoja EI SAA pakata. Monet tiedostomuodot ovat pakattuja ja tällainen pakkaaminen on sallittua, mutta esimerkiksi erillinen pakkaaminen (zip jne.) on kiellettyä.
- Tiedostosta EI SAA puuttua sen esittämiseen tarvittavia ulkoisia komponentteja. Tällaisia ovat esimerkiksi fontit ja muut linkitetyt aineistot, jotka vaikuttavat tiedoston esittämiseen.

Arkistokelpoiset tiedostomuodot vaihtelevat tiedoston tyyppin mukaan. Tässä on esitelty muutamia suositeltuja tiedostomuotoja erityyppisille tiedostoille, kuten kuville ja dokumenteille.

PDF-tiedostoja varten on määritelty arkistointiin tarkoitettu PDF/A-standardi (PDF-Archive). PDF/A:sta on useita eri versioita, joista uusin on PDF/A-3, ja esimerkiksi PAS-palvelun suositus on PDF/A-1a ja PDF/A-1b. PDF/A-tiedostossa ei ole sellaisia tavallisen PDF-tiedoston ominaisuuksia, jotka ovat haitallisia tiedon luotettavalle pitkäaikaissäilyttämiselle, kuten salaus, äänet ja tekstiin upotetut hyperlinkit. Alla olevassa taulukossa on vertailtu eri PDF-tiedostojen ominaisuuksia.

Tiedostomuoto	Standardi	Ominaisuudet	Arkistokelpoisuus
PDF	ISO 32000-2	Ei arkistokelpoinen. Tavallinen PDF-tiedosto sisältää esimerkiksi upotettuja linkkejä ja tekstin muotoiluja, jotka eivät sovellu arkistointiin.	Ei
PDF/A-1	ISO 19005-1	PAS-palvelun suosittelu standardi, josta on poistettu kaikki sellaiset ominaisuudet, jotka eivät sovi pitkäaikaissäilyttämiseen.	Kyllä
PDF/A-2	ISO 19005-2	Sisältää mahdollisuuden upottaa muita PDF/A-tiedostoja, minkä avulla voidaan muodostaa useista dokumenteista koostuva kokoelma yhteen tiedostoon.	Kyllä
PDF/A-3	ISO 19005-3	Vastaa PDF/A-2 -standardia, minkä lisäksi dokumenttiin voidaan upottaa muita tiedostoja (kuten tekstitiedostoja ja taulukkotiedostoja).	Kyllä

Tavalliset txt-muotoiset tekstitiedostot soveltuvat myös arkistointiin yksinkertaisuutensa vuoksi, eli ne eivät sisällä esimerkiksi kuvia tai tekstin muotoilua. Tekstiä voidaan myös arkistoida CSV-muodossa (Comma Separated Values). CSV-tiedostoissa teksti on

tallennettu taulukkomuodossa, ja rivit ja kentät on eroteltu pilkuilla ja rivinvaihdoilla. PDF/A on kuitenkin monipuolisin vaihtoehto, ja yleisenä ohjenuorana sitä voidaan pitää parhaana vaihtoehtona tekstiä sisältävien dokumenttien arkistointiin. PDF/A-tiedostoja pystytään myös vaivatta avaamaan verkkoselaimilla, mikä ei onnistu esimerkiksi TIFF-kuvatiedostoilla. PDF/A on myös aineistojen julkaisuun vaivattomasti soveltuva formaatti.

Ääntä voidaan tallentaa useissa eri muodoissa. Vaikka MP3 on kenties suosituin tiedostomuoto musiikin tallentamiseen, niin häviöllisenä formaattina se ei kuitenkaan ole ihanteellinen valinta pitkäaikaissäilytykseen. WAV (Waveform Audio Format) on suosittu pakkaamaton ja häviötön muoto, ja tarkemmin sanottuna PCM-koodattu WAV on hyvä valinta arkistointia varten. FLAC (Free Lossless Audio Codec) on myös kelvollinen häviötön formaatti. Äänen ja audiovisuaalisen arkistoinnin kansainvälinen yhdistys (IASA) kuitenkin suosittelee äänitiedostojen arkistointiin BWF-muotoa (Broadcast Wave Format). BWF on korkealaatuinen ääniformaatti ja se on myös yhteensopiva WAV-tiedostojen kanssa. BWF on myös PAS-palveluiden suositeltujen ääniformaattien listalla.

Videotiedostot muodostuvat niin sanotusta säiliömuodosta (container) ja koodekista (codec). Säiliömuoto on käytännössä kääre, jonka sisälle on tallennettu eri bittivirtoja, joista olennaisimpia ovat videon liikkuva kuva ja ääni. Muita bittivirtoja ovat esimerkiksi metatiedot tai vaikkapa elokuvien tekstitykset. Koodekin tehtävä puolestaan on ohjelmallisesti valmistella säiliömuodon sisältämät bittivirrät eri käyttötarkoituksiin, kuten videon toistaminen tai editointi. Koodekit voivat olla joko häviöllisiä tai häviöttömiä. Häviölliset menetelmät huonontavat videon ja äänen laatua, vaikka niillä saadaan kuitenkin pienennettyä tiedoston kokoa. Häviöttömät koodekit tuottavat parempaa laatua, mutta vastaavasti tiedostokoot ovat suurempia.

Suosituimpia ja yleisimpiä säiliömuotoja ovat AVI ja MOV. AVI-tiedostoissa usein käytetty koodekki on MPEG-4 AVC (H.264), jota käytetään esimerkiksi HD-televisiolähetysissä ja Blu-ray-levyillä. Laajan levinneisyyden takia H.264 onkin hyvä valinta, jos halutaan varmistaa mahdollisimman suuri yhteensopivuus eri ohjelmien ja laitteiden kanssa. Myös esimerkiksi YouTube tukee näitä tiedostomuotoja, mikä kannattaa huomioida, jos vaikkapa digitoituja VHS-kotivideoita on tarkoitus jakaa verkossa. H.264 on kuitenkin häviöllisesti pakattua, joten laadullisesti on olemassa parempia vaihtoehtoja. Esimerkiksi JPEG 2000 on häviötön tiedostomuoto, ja sen säiliömuotona voi toimia Material Exchange Format (MXF) tai Motion JPEG 2000 (MJ2). Useat suuret organisaatiot suosivatkin MXF-kääreeseen tallennettua JPEG2000-videota. Suosittuihin säiliömuotoihin lukeutuu myös MOV (Quicktime), joka on laajalti käytössä pakkaamattoman videon pitkäaikaissäilytyksessä, ja sille löytyy myös tuki useista eri ohjelmista.

Vanhojen filmielokuvien digitointiin käytetään myös Digital Picture Exchange –formaattia (DPX), joka tallentaa jokaisen filmin kuvan yksittäisenä tiedostona. Tästä syystä DPX ei siis ole kovin käytännöllinen vaihtoehto omien kotifilmien digitointiin ja säilytykseen.

Videotiedostoille ei oikeastaan ole mitään täysin vakiintunutta arkistointikelpoista tiedostomuotoa, joka olisi jokaisen arkiston käytössä, vaan eri arkistot ovat valinneet käyttämänsä tiedostomuodot omien kriteerien perusteella. Kansalliset pitkäaikaissäilytyspalvelut (PAS-palvelut) hyväksyvät videotiedostoista MXF- tai MJ2-kääreeseen tallennettua JPEG 2000 –videota, MPEG-4-videoita ja DPX-videoita.

BBC on ainoa arkisto maailmassa, joka käyttää videoiden arkistointiin MXF-kääreessä pakkaamatonta ja häviötöntä v210-muotoa.

Tällainen videotiedosto vaatii kuitenkin valtavasti tallennustilaa: yksi tunti tällä menetelmällä tallennettua liikkuvaa kuvaa vaatii noin 100 gigatavua tallennustilaa. Videotiedostot vaativatkin yleensä runsaasti tallennustilaa, ja esimerkiksi tunnin mittainen video normaalilla DVD-laadulla vaatii lähes 4 gigatavua tallennustilaa. Videotiedoston kokoon pystyy kuitenkin vaikuttamaan huomattavasti laskemalla videon laatua. Arkistointiin soveltuvana tiedostomuotona pidetään myös pakkaamatonta AVI-formaattia. MP4/M4V-muodot ovat pakattuja ja laadultaan AVI-tiedostoja heikompia, vaikka niiden etuna on tiedostojen pienempi koko.

Kuvatiedostoja arkistoidaan useimmiten TIFF-muodossa, koska se on häviötön ja korkealaatuinen tiedostomuoto. 300 DPI:n tarkkuudella skannatut TIFF-kuvat ovat useimpiin tarkoituksiin riittävän korkealaatuisia. Suurin osa valmiiksi saatavilla olevista kuvista on kuitenkin yleensä JPEG-muotoisia, koska esimerkiksi kännykkäkamerat tuottavat oletuksena JPEG-muotoisia digitaalisia kuvia. JPEG tosin on häviöllinen formaatti ja siksi TIFF-kuvissa on parempi laatu. JPEG-muodosta on myös olemassa laadukkaampi JPEG 2000 -muoto (JP2), joka mahdollistaa sekä häviöttömän että häviöllisen kuvan pakkaamisen. JPEG 2000 onkin JPEG:tä parempi vaihtoehto kuvien pitkäaikaissäilyttämiseen, koska paremmasta laadusta huolimatta myös tiedostot ovat pienempiä.

PNG on myös korkealaatuinen kuvaformaatti, jota voidaan käyttää kuvien arkistointiin. PNG löytyy myös PAS-palveluiden suositeltujen tiedostomuotojen listalta.

Seuraavalla sivulla on vertailtu pakatun JPEG-kuvan ja pakkaamattoman PNG-kuvan laatua. Kuvista voi havaita, kuinka paljon pakkaaminen saattaa vaikuttaa kuvanlaatuun. Varsinkin taivaan väreissä näkyy selvästi rakeisuutta. Laadun kuitenkin huomaa myös tiedostojen koossa, nimittäin heikkolaatuinen JPEG-kuva on kooltaan vain 257 kilotavua, kun laadukas PNG-kuva puolestaan vie hulppeat 30 megatavua.



Kuva 1 Korkealaatuinen PNG-kuva



Kuva 2 Heikkolaatuinen ja pakattu JPG-kuva

Kiteytettynä voitaisiin sanoa, että tiedosto on säilytyskelpoinen silloin kun sen sisältämä tieto säilyy ja pysyy muuttumattomana ajankulusta huolimatta. Tiedostomuotoina on hyvä käyttää formaatteja, joiden tekniset tiedot ovat julkisesti saatavilla. Avoimiin standardeihin perustuvat tiedostomuodot ovat arkistointia silmällä pitäen parempi vaihtoehto kuin yksityisomistukselliset formaatit. Kysynnän ja tarjonnan lakeja mukailten suosituimmille tiedostomuodoille löytyy myös laajemmin yhteensopivia ohjelmistoja verrattuna sellaisiin tiedostomuotoihin, joilla on vain pieni käyttäjäryhmä. Arkistoitavan formaatin tärkeimpiä ominaisuuksia on pysyvyys, eli formaatin tulisi pysyä muuttumattomana pitkiäkin aikoja. Mikäli formaattiin tulee muutoksia, sen uudempien versioiden pitäisi olla yhteensopivia vanhojen versioiden kanssa.

Yksityisomistuksellisia tiedostoja hallinnoivat organisaatiot ja kaupalliset tahot, jotka pitävät tiedostomuotojen tarkat tiedot piilossa yleisöltä. Tällaisia tiedostoja voi käsitellä vain tietyillä ohjelmilla ja ne vaativat käyttöoikeuksia johonkin ohjelmistoon. Avoimet tiedostomuodot sen sijaan ovat täysin julkisia, hyvin dokumentoituja ja niitä voi käsitellä laajasti eri ohjelmilla ja alustoilla. Avointen tiedostojen saatavuus on siis parempi, ja tästä syystä niitä pidetään yleisesti paremmin pitkäaikaissäilytykseen sopivina.

Tallennusmediat

Digitaaliset tiedostot pitää tietenkin tallentaa johonkin, missä niitä voidaan säilyttää pitkiäkin aikoja, ja henkilökohtaiseen digiarkistointiin on monia säilytysvaihtoehtoja. Yleisimpiä tallennusmedioita ovat

- tietokoneen kovalevyt
- USB-muistit
- ulkoiset kovalevyt
- optiset mediat (esim. DVD-levyt)
- LTO-nauhat

Ulkoiset kovalevyt ovat helposti siirreltävissä paikasta toiseen, ja lisäksi niiden tallennuskapasiteetti on riittävä esimerkiksi kuvien tallentamista varten. Tavanomaiset USB-muistit ovat myös helposti siirreltävissä, vaikka niiden tallennuskapasiteetti ja kestävyys ei ole ulkoisten kovalevyjen kanssa samalla viivalla.

Tiedostojen, varsinkin videoiden ja kuvien, tallentaminen CD- ja DVD-levyille on melko suosittu digisäilytysmenetelmä. Näiden optisten tallennusmenetelmien huono puoli on kuitenkin kömpelö käytettävyys, ja esimerkiksi ulkoiselta kovalevyltä tiedostot on nopeampi avata. Lisäksi heikkolaatuisilta optisilta levyiltä tieto saattaa alkaa kadota jo kymmenessä vuodessa.

Kannattaa myös muistaa, että arkisto ei ole sama asia kuin varmuuskopio, vaikka se voikin sellaisena myös toimia. Oman digitaalisen aineiston säilömisessä on hyvä muistaa niin sanottu ”3-2-1” –sääntö: aineistosta on kolme eri kopiota, jotka on tallennettu kahdelle eri alustalle ja yksi kopio sijaitsee maantieteellisesti erillään muista kopioista.

Suurissa arkistoissa käytetään LTO-nauhoja suurten tietomäärien tallentamiseen. Nauhatallennus onkin halvin vaihtoehto runsaasti tallennustilaa vaativien aineistojen

säilyttämiseen, vaikka toisaalta nauha-aseman hinta on korkeampi kuin tavanomaisten tallennusvälineiden kuten kiintolevyjen. Nauhatallennus on taloudellinen vaihtoehto vasta siinä vaiheessa, kun aletaan puhua petatavujen tallennustarpeesta, ja tästä syystä on harvoin taloudellista käyttää nauhatallennusta omaan kotikäyttöön tarkoitetun digiarkiston tallennusmuotona. Esimerkiksi vuonna 2024 valmistuvan maailman suurimman radioteleskooppi Square Kilometer Arrayn päivittäin tuottama data vaatisi yli 300 tavallista kiintolevyä, jos jokaisen kiintolevyn tallennuskapasiteetti olisi kolme teratavua. Nauhat myös säästävät energiaa verrattuna tavanomaisiin kiintolevyihin.

Kaikilla tallennusformaateilla on omat rajoitteensa, eikä mikään formaatti kestä ikuisesti. Mitä enemmän tallennusvälinettä käytetään, sitä suurempi on riski, että se vahingoittuu. Tallennusvälineet pitääkin säilyttää turvallisessa ympäristössä, missä ne eivät altistu ilmankosteudelle tai lämpötilojen vaihtelulle.

Säilytysmenetelmät ja -järjestelmät vanhenevat ajan kuluessa, ja aineisto pitäisi säännöllisesti siirtää uudempiin järjestelmiin. Tällaista siirtämistä kutsutaan **migraatioksi**. Tavanomaiset tallennusmenetelmät myös eroavat hieman arkistointiin soveltuvista tallennusmenetelmistä. Siinä missä tavanomaisilla kovalevyillä ja muistitikuilla säilytettyä tietoa käytetään jatkuvasti, niin arkistoissa säilytetty tieto saattaa pysyä käyttämättömänä pitkiäkin aikoja, ja tämä asettaa erilaisia vaatimuksia arkistointiin tarkoitetuille

tallennusmenetelmille. Tallennusvälineiden suunnittelussa pitää ottaa huomioon pitkä aikaväli ja niiden pitää myös olla kestäviä ja mahdollistaa säilytetyn tiedon eheyden tehokas seuranta.

Sähköinen arkisto ei täysin palvele tarkoitustaan, jos se ei kestä koko suunnitellun elinkaarensa ajan. Sähköisen arkiston hallinta ja ylläpito helpottuu, kun arkistossa ei käytetä useita eri tiedostomuotoja. Ennen arkistointiprojektin aloittamista onkin järkevää päättää, mikä on pienin mahdollinen määrä eri tiedostomuotoja, jolla pystytään toteuttamaan haluttu arkisto.

Arkistokelpoiset tiedostomuodot ovat korkealaatuisia

Pakkaamattomat tiedostomuodot mahdollisimman korkealla laadulla ovat arkistointiin paras vaihtoehto. Näin varmistetaan, että arkistoitavassa tiedostossa ei katoa alkuperäisen dokumentin sisältämää informaatiota.

Kaikille aineistotyypeille ei vielä ole vakiintunutta tiedostomuotoa

Eri arkistoilla on omat suosituksensa arkistoitavista tiedostomuodoista.

Metatieto ja dokumentointi

Metadata eli metatieto on digitaalisesta aineistosta tallennettua tietoa, joka on järjestelty koneelliseen käsittelyyn sopivaan muotoon. Lyhyesti sanottuna metatieto on tietoa tiedosta, eli se kertoo aineiston kontekstista, sisällöstä ja rakenteesta sekä kuvailee niiden hallintaa ja käsittelyä koko aineiston elinkaaren ajan. Metatietoa voi käyttää esimerkiksi aineiston hakuun ja tunnistamiseen (<http://digitalpreservation.fi/specifications/sanasto>).

Metadatan tarkoitus pitkäaikaissäilytyksessä on

- säilyttää tietoa aineistoon kohdistuneista toimenpiteistä
- tukea aineiston löydettävyyttä ja käytettävyyttä
- tukea pitkäaikaissäilytyksen tavoitteita, eli tiedon saatavuuden, ymmärrettävyyden, oikeellisuuden ja eheyden varmistamista.

Lisäksi metadatan keskeisiä tehtäviä on kertoa, kuka aineiston on luonut sekä milloin ja mihin tarkoitukseen aineisto on luotu.

Metatietoa voi havainnollistaa parhaiten alla olevan esimerkin avulla. Avataan ”dokumentti”-niminen pdf-tiedosto ja katsellaan tiedoston ominaisuuksia (kuva 3).

The screenshot shows a window titled "Dokumentin ominaisuudet" (Document Properties) with a close button (X) in the top right corner. Below the title bar are several tabs: "Kuvaus" (Description), "Suojaus" (Security), "Kirjasimet" (Fonts), "Aloituskäytä" (Initial View), "Muokattu" (Modified), and "Lisäasetukset" (Additional Settings). The "Kuvaus" tab is selected, and the main area contains the following metadata fields:

- Tiedosto: dokumentti
- Otsikko: Metadatan on tietoa tiedosta
- Tekijä: Matti Meikäläinen
- Aihe: Metadatan
- Avainsanat: metadatan, dokumentointi, digitointi
- Luotu: 22-Jan-19 11:51:39
- Muokattu: 22-Jan-19 11:52:49
- Sovellus: Adobe Acrobat 11.0

At the bottom right of the main area, there is a button labeled "Muut metatiedot..." (More Metadata...).

Kuva 3 PDF-tiedoston sisältämiä metatietoja

Tiedosto sisältää seuraavat metatiedot: otsikko, tekijä, aihe ja avainsanat. Nämä ovat tietoja, joita ei voi heti päätellä pelkästään tiedoston nimen perusteella, vaikka ne tarjoavat tärkeää tietoa dokumentin sisältämästä tiedosta. Esimerkiksi tekijän perusteella dokumentti näkyisi hakutuloksissa, jos käyttäjä hakee sellaista aineistoa, jonka tekijänä on Matti Meikäläinen. ”Avainsanat”-kenttä puolestaan sisältää sanoja, jotka kuvailevat dokumentin sisältöä. Käyttäjä voisi esimerkiksi etsiä aiheeseen ”metadatan” liittyvää

aineistoa, jolloin esimerkkidokumenttimme ilmaantuisi hakutuloksiin. Ilman metadataa aineiston haettavuus olisi siis huomattavasti suppeampaa.

Metatietoja varten on olemassa erilaisia standardeja, jotka määrittelevät tiedostoon liitettävät metatietokentät. Metatietostandardit parantavat tiedon yhtenäisyyttä ja helpottavat tiedostojen siirrettävyyttä ja yhteentoimivuutta eri järjestelmien välillä. Yksi tällaisista standardeista on hyvin yleisesti käytetty Dublin Core, joka määrittelee 15 erilaista metatietokenttää:

- Title (nimeke)
- Creator (tekijä)
- Subject (aihe)
- Description (kuvaus)
- Publisher (julkaisija)
- Contributor (muu tekijä)
- Date (aikamääre)
- Type (laji)
- Format (formaatti)
- Identifier (tunniste)
- Source (lähde)
- Language (kieli)
- Relation (suhde)
- Coverage (kattavuus, esimerkiksi ajallinen tai maantieteellinen)
- Rights (käyttöoikeudet)

Dublin Core -metadatformaatista on olemassa myös suomalainen versio, joka tunnetaan SFS 5895 -standardina. Tästä on edelleen jatkokehitetty SFS 5914 -standardi, joka sisältää lisäksi SFS-ISO 23081-1:2007 -metatiedot:

- Audience (yleisö)
- Location (sijainti)
- Disposition
- Agent
- Function
- RecordType
- Mandate
- Status
- AuditTrail (käsittelyhistoria)

Lisäksi laajalti käytössä olevia muita standardeja ovat esimerkiksi ISO 23081 ja JHS 143. Nämä ovat kuitenkin Dublin Corea ja sen sukulaisia järeämpiä standardeja, ja pienempiin arkistoihin Dublin Coren tarjoamat metatietokentät ovat usein riittävän kattavia.

Metatiedon luominen manuaalisesti vaatii aikaa ja vaivaa, ja se saattaa viedä jopa enemmän aikaa kuin aineiston muuntaminen digitaaliseen muotoon. Metatiedot kannattaa aina lisätä samalla kun digitaalinen tiedosto luodaan. Kuville ja videoille tarpeellisia metatietoja ovat ainakin kuvaaja, kuvassa tai videolla esiintyvät henkilöt, aika ja paikka. Nämä tiedot ovat tärkeitä, koska pelkästään ihmisen muistiin ei voi luottaa, ja vuosien kuluttua, kun digiarkiston uumenista kaivellaan vanhaa kuvamateriaalia eikä kukaan enää

muistakaan henkilöiden tai paikan nimiä ajankohdasta puhumattakaan, niin aineisto on lähes hyödytöntä. Metadata on yhtä tärkeää kuin tiedostojen nimeäminen kuvaavasti.

Sisäinen metatieto sisältyy tiedostoon, kun ulkoinen metatieto sijaitsee erillisessä tiedostossa.

Dokumentointi täydentää metadataa ja tarjoaa tietoa, joka edesauttaa aineiston käytettävyyttä ja saatavuutta kolmannelle osapuolelle. Dokumentointi tarjoaa tarkkaa tietoa aineiston tekijän suorittamista toimenpiteistä ja ohjeistaa muita aineiston käytössä. Dokumentointi on tärkeää, koska digitaaliset aineistot ovat riippuvaisia laitteistosta ja ohjelmistoista, joten dokumentoinnilla varmistetaan, että jälkipolvilla on aineiston säilyttämiseen ja käsittelyyn riittävästi tietoa.

Dokumentointi on metadatan ohella käytännössä ainoa keino varmistaa aineiston oikeellisuus. Digitaalisia aineistoja on usein välttämätöntä kopioida niiden saatavuuden takaamiseksi, ja siksi dokumentoinnilla myös selvennetään aineistoon liittyviä tekijänoikeuksia. Dokumentoinnissa tuodaan myös esille, että aineistoon on kohdistettu seulonta ennen arkistointia, mikä takaa sen, että arkistoidun aineiston arvo on varmistettu.

Kattavan dokumentoinnin laatiminen onnistuu parhaiten, kun dokumentointi aloitetaan samaan aikaan digitoitiprojektin kanssa ja viedään loppuun yhtä aikaa muun projektin kanssa. Hyvä dokumentointi kertoo kattavasti projektin eri vaiheista niin että projektia koskeva tieto on saatavilla, ymmärrettävissä ja käytettävissä. Hyvä dokumentointi myös elää jatkuvasti projektin muutosten mukana, ja sitä päivitetään säännöllisesti, jotta tieto pysyy ajan tasalla.

Metadata on tietoa tiedosta

Metadata eli metatieto kuvailee esimerkiksi aineiston kontekstia, sisältöä ja rakennetta.

Metadata parantaa aineiston käytettävyyttä

Metatietoa voi hyödyntää esimerkiksi aineiston hakuun ja tunnistamiseen.

Dokumentointi täydentää metadataa

Dokumentointi tarjoaa tarkkaa tietoa aineiston tekijän suorittamista toimenpiteistä ja ohjeistaa muita aineiston käytössä.

Dokumentointiin kannattaa panostaa

Hyvä dokumentointi kertoo kattavasti projektin eri vaiheista niin että projektia koskeva tieto on saatavilla, ymmärrettävissä ja käytettävissä.

Kustannukset ja laatukriteerit

Muutamia seikkoja kannattaa pohtia, ennen kuin suin päin rynnätään digitointiprojektin pariin. Tällaisia seikkoja ovat esimerkiksi se, kuinka laajasti aineistoa digitoidaan, mikä on digitoinnin tarkoitus ja mikä on siitä saatava hyöty. Tämä on olennainen kysymys varsinkin silloin kun digitoinnin tarkoitus on kuluja leikkaaminen. Onko sähköisen arkiston ylläpitäminen halvempaa kuin alkuperäisen paperisen arkiston? Mikäli alkuperäinen aineisto aiotaan hävittää digitoinnin jälkeen, niin on huomioitava lakiasiat ja varmistettava, että aineiston hävittäminen on sallittua.

Yksityishenkilölle digitoinnin kilpailutus eri palveluntarjoajien välillä ei välttämättä ole tarpeellista, mutta organisaatioiden digitointiprojektin toteuttaja yleensä haetaan kilpailuttamalla. Ennen tarjouspyynnön jättämistä tiedetään

- mitä digitoidaan
- mitkä ovat laatukriteerit.

Palveluntarjoajan esittämä tarjous on mahdollisimman realistinen, kun edellä mainitut seikat ovat tiedossa. Kun palveluntarjoaja on valittu, niin on järkevää antaa palveluntarjoajan digitoida pieni koe-erä koko aineistosta. Näin pystytään toteamaan vastaako digitointi niitä laatukriteereitä, joita suunnittelussa on asetettu.

Digitointiprojektin hintaan ja keston vaikuttavat useat tekijät, ja kokonaishintaa on hankala määrittää. Hintaan vaikuttaa paljon esimerkiksi se, kuinka paljon työstä joudutaan ulkoistamaan ja kuinka paljon on jo entuudestaan olemassa tarvittavaa IT-infrastruktuuria, osaamista ja laitteistoa. Alkuperäisen aineiston muoto vaikuttaa myös huomattavasti hintaan, koska sen perusteella määräytyy digitointiin tarvittavat laitteistot ja menetelmät. Esimerkiksi tavallisia, hyväkuntoisia kirjoja pystytään digitoimaan tavallisella tasokannerilla noin 200 sivua tunnissa, kun taas valokuvien skannaus käy huomattavasti hitaammin.

Hintaan vaikuttavista tekijöistä voidaan nostaa esille seuraavat kolme tekijää (Poole 2010):

- alkuperäisen aineiston luonne ja kunto
- toiminnan helppo toistettavuus
- digitoidun aineiston tavoiteltu laatu, laajuus ja kestävyys.

Mikäli digitointi on ostettu palveluna, niin palveluntarjoaja antaa projektikohtaisen arvion aikataulusta, mihin luonnollisesti vaikuttavat aineiston kunto ja laajuus, sekä ostajan asettamat laatukriteerit. Yksityishenkilön omaan digitointiprojektiin kuluva aika riippuu samoista tekijöistä sekä henkilön omasta panostuksesta.

Äänitteiden ja videokasettien digitoinnissa kannattaa muistaa, että aineiston siirtäminen digitaaliseen muotoon vie yhtä kauan aikaa kuin alkuperäisellä materiaalilla on kesto. Eli käytännössä kaksi tuntia kestävä kotivideon digitointi vie aikaa kaksi tuntia. Lisäksi tässäkin on hyvä huomioida, että itse digitoinnin lisäksi aikaa kuluu myös projektin muihin tehtäviin, kuten suunnitteluun, dokumentointiin ja alkuperäisen aineiston valmisteluun digitointia varten. Esimerkiksi dokumentit saattavat olla kansioissa tai muovitaskuissa, joista ne täytyy purkaa skannausta varten. Asiakirjoissa on myös usein digitoinnille

haitallisia esineitä, kuten niittejä, klemmareita, teippiä ja liimaa, jotka kaikki täytyy poistaa ennen kuin aineistoa voidaan digitoida.

Digitoidun aineiston laatu vaikuttaa usein suoraan tiedostojen kokoon. Käytännössä siis tiedostot vaativat sitä enemmän tallennustilaa mitä korkeampaa niiden laatu on. Tämä koskee varsinkin kuvia, videoita ja äänitiedostoja. Tästä syystä laatu vaikuttaa siis myös kustannuksiin, koska suurempien tiedostojen vaatima tallennustila myös maksaa enemmän.

Yleensä yritykset veloittavat paperisten aineistojen digitoinnista tuntihinnan perusteella. Tämä johtuu siitä, että aineiston kunto vaikuttaa huomattavasti digitointiin kuluvaan aikaan, eli huonokuntoisen materiaalin digitointi käy huomattavasti hitaammin kuin hyväkuntoisten dokumenttien. Tästä syystä yritysten on vaikea määritellä kiinteää kappalehintaa esimerkiksi yksittäiselle sivulle tai kuvalle. Sen sijaan äänitteille ja videonauhoille käytetään yleensä materiaalin keston perustuvaa veloitusta. Hinnat liikkuvat 20-30 euron välillä yhtä nauhatallennetta kohti.

Kappalehintoja käytetään sellaisille aineistoille, joita voidaan tasaisella nopeudella digitoida, kuten vanhat diakuvat ja helposti käsiteltävissä olevat valokuvat. Usein yritykset veloittavat ensin jonkin kiinteän aloitusmaksun, joka saattaa olla muutamia kymmeniä euroja, minkä jälkeen hintaan lisätään kappalehinta jokaista digitoitua kappaletta kohti. Kappalehinnat liikkuvat muutamissa kymmenissä senteissä.

Esimerkki kustannussäästöistä: Kansallisarkiston massadigitointihanke

Kansallisarkiston massadigitointihankkeessa kaikki viranomaisten hallussa oleva arkistoitava analoginen aineisto tullaan digitoimaan vuoteen 2030 mennessä. Yhteensä digitoitavaa aineistoa on arvioiden mukaan 212 hyllykilometriä. Massadigitoinnin myötä säästöjä saavutetaan kolmella tavalla:

- Nykyisten arkistotilojen kustannukset ovat noin 7 miljoonaa euroa vuodessa, ja massadigitoinnin jälkeen nämä kustannukset laskevat peräti 70%.
- Nykyiset arkistotoimen ja asiakirjahallinnon henkilöstökustannukset ovat noin 50 miljoonaa euroa vuodessa, mistä voidaan säästää 30% kun aineisto digitoidaan.
- Lisäksi analogisessa aineistossa pysyminen tarkoittaisi sitä, että myös arkistotiloja täytyisi tulevaisuudessa rakentaa lisää, mikä tulisi maksamaan 32,8 miljoonaa euroa vuoteen 2039 mennessä.

Laatukriteerit ja laadunvalvonta

Digitoinnin laatukriteereiksi löytyy esimerkiksi Kansallisarkiston laatimat suositukset, joita voi hyödyntää digitointiprojektissa (<https://www.arkisto.fi/uploads/normit/valtiorhallinto/suosituksset/digikriteerit.pdf>).

Digitointiprojektin alussa on järkevää kartoittaa nykyiset taidot ja käytettävissä olevat resurssit, jotta tiedetään, millaista laatua tavoitellaan ja mikä laatu on toteutettavissa nykyisillä resursseilla: tavoite on paras mahdollinen laatu resurssien puitteissa.

Lähtökohtana on siis hyvä ymmärtää organisaation kapasiteetti, tehtävät ja tavoitteet. Arkistoitavasta aineistosta olisi myös hyvä olla jonkinlainen ymmärrys.

Laatukriteerit riippuvat siitä, mihin tarkoitukseen arkistoitavia tiedostoja halutaan käyttää, ja mahdollista jatkokäyttöä varten voi olla myös hyödyllistä arkistoida kaksi eri versiota samasta tiedostosta. Esimerkiksi kuvia ja dokumentteja halutaan kenties jatkossa muokata ja käyttää eri tarkoituksiin, joten kannattaa arkistoida sekä muokattavissa olevat alkuperäiset versiot että arkistokelpoinen versio eri tiedostomuodossa. Esimerkiksi tavanomainen Word-dokumentti voidaan arkistoida sellaisenaan muokkaamista silmällä pitäen, minkä lisäksi samasta dokumentista arkistoidaan myös PDF-versio, jota ei ole tarkoitus muokata.

Laadunvarmistukseen on olemassa kaupallisia tietokoneohjelmia, jotka tarkistavat automaattisesti digitaalisen videon tai äänitteen laadun. Ohjelmat tutkivat esimerkiksi kuvassa näkyviä virheitä ja äänenvoimakkuutta. Dokumenttien laadunvalvonnassa käytetään keinona pistokoemaista aineiston tutkimista, eli aineistosta valitaan satunnaisena otantana materiaalia, jonka laatu varmistetaan manuaalisesti. Jokaisen yksittäisen dokumentin laadun varmistaminen manuaalisesti on käytännössä mahdotonta.

Jos digitoinnin tärkein tavoite on kulujen leikkaaminen, niin joskus valitaan tarkoituksella häviöllisiä tiedostomuotoja, vaikka aineiston laatu kärsiikin. Halutut säästöt saavutetaan kuitenkin tallennustilan määrässä, jota tarvitaan vähemmän kuin korkealaatuiselle aineistolle. Laatukriteerinä voidaan vähintään pitää sitä, että tiedosto on riittävän korkealaatuinen tulevaisuuden käyttöä varten eikä se ole arkistointivaiheessa elinkaarensa lopussa. Tärkeintä on se, että kaikki alkuperäinen informaatio säilyy eikä mitään tietoa katoa digitoinnin aikana.

Laatua valvotaan koko ketjun ajan, koska digitointi ja digiarkistointi koostuvat monista eri prosesseista. Lopullisen digiarkiston laatua ei määrittele ainoastaan se, kuinka korkealaatuisia tiedostot ovat, vaan myös se, kuinka hyvin projekti on alusta asti suunniteltu. Korkealaatuisilla kuvilla ja tärkeillä dokumenteilla ei tee mitään, jos metatiedot ovat puutteelliset tai arkiston kansiorakenne on suunniteltu huiloiden.

Kustannuksiin vaikuttavat monet tekijät

Esimerkiksi alkuperäisen aineiston määrä ja kunto sekä digitoinnissa käytetyt menetelmät vaikuttavat hintaan.

Tavoiteltu laatu päätetään digitointiprojektin alussa

Tavoitteena on saavuttaa paras mahdollinen laatu käytettävissä olevien resurssien puitteissa.

Laadunvalvontaa voidaan osin automatisoida

Suurimmaksi osaksi laatua valvotaan kuitenkin manuaalisesti esimerkiksi tutkimalla satunnaisia otantoja aineistosta. Videon ja äänen laatua on mahdollista tarkkailla koneellisesti.

Tärkeintä on, että alkuperäistä informaatiota ei katoa digitoinnin aikana

Digitoinnin tavoite on tuottaa alkuperäisestä aineistosta sellaisia digitaalisia jäljenteitä, joiden informaatio on täysin sama kuin alkuperäisessä aineistossa.

Aikaa kuluu moniin eri työvaiheisiin

Digitointiprojektissa varsinainen digitointi vie vain murto-osan ajasta, koska lisäksi aikaa kuluu esimerkiksi aineiston valmisteluun ja laitteiston kalibrointiin.

Lakiasiat

Lienee aiheellista perustella, miksi yleiset ohjeistukset ja toimintatavat ovat tärkeitä. Arkistolaki velvoittaa viranomaisia ja julkista tehtävää hoitavia organisaatioita toimimaan lain mukaan, mutta yksityishenkilöitä ja suurinta osaa yrityksissä syntyvästä aineistoista arkistolaki ei koske. Tästä syystä näiden viimeksi mainittujen arkistointia pyritään yhdenmukaistamaan ja säätelemään yleisillä ohjeilla ja käytännöillä.

Joskus aineisto halutaan digitoida, jotta alkuperäinen paperinen aineisto voidaan hävittää. Tällöin alkuperäisen aineiston hävittämisessä pitää ottaa huomioon juridiset seikat ja varmistaa, että aineiston hävittäminen on laillista.

Jokainen arkiston omistava organisaatio on velvollinen noudattamaan tietosuojaan liittyviä lakeja ja käytäntöjä. Henkilökohtaiseen käyttöön tarkoitettu arkisto saa sisältää henkilötietoja, eikä se silti muodosta EU:n tietosuoja-asetuksen piiriin kuuluvaa henkilökisteriä, koska luonnollinen henkilö saa kerätä omaan yksityiseen käyttöönsä henkilötietoja esimerkiksi puhelinluettelo varten. Jos arkiston omistaja kuitenkin jakaa arkiston julkisesti esimerkiksi verkossa, niin tällöin henkilötiedoista muodostuu henkilökisteri ja arkiston omistaja on myös velvoitettu noudattamaan tietosuojaan liittyviä lakeja. EU:n tietosuoja-asetukseen voi perehtyä osoitteessa <https://tietosuoja.fi/gdpr>.

Henkilötietojen käsittelyssä tulee aina noudattaa hyviä käytäntöjä ja huolellisuutta, ja henkilötietojen käsittelylle pitää aina olla perusteltu syy. Tietosuojalaissa kuvataan käsittelyn lainmukaisuutta neljällä kohdalla, joista varsinkin kaksi jälkimmäistä ovat olennaisia arkistoinnin näkökulmasta:

- käsittely on tarpeen tieteellistä tai historiallista tutkimusta taikka tilastointia varten ja se on oikeasuhtaista sillä tavoiteltuun yleisen edun mukaiseen tavoitteeseen nähden; tai
- henkilötietoja sisältävien tutkimusaineistojen, kulttuuriperintöaineistojen sekä näiden kuvailutietoihin liittyvien henkilötietojen käsittely arkistointitarkoituksessa on tarpeen

ja oikeasuhtaista sillä tavoiteltuun yleisen edun mukaiseen tavoitteeseen ja rekisteröidyn oikeuksiin nähden.

Tekijänoikeudella suojatun aineiston kopioiminen omaan käyttöön on sallittua, kunhan kopioitava aineisto on valmistettu laillisesti. Kopioita saa luoda omaan käyttöön muutamia kappaleita, eli puhutaan noin 2-4 kappaleesta. Tekijänoikeus on voimassa koko tekijän eliniän ajan, minkä lisäksi tekijänoikeudella on suoja-aika, joka on voimassa 50 tai 70 vuotta tekijän kuoleman jälkeen. Suoja-aika riippuu teoksen alkuperästä. Mikäli teos on lähtöisin Euroopan Talousalueelta (ETA), niin noudatetaan 70 vuoden suoja-aikaa. ETA-alueen ulkopuolella syntyneiden teosten kohdalla noudatetaan 50 vuoden suoja-aikaa, ellei alkuperämaassa ole erikseen määritelty pidempää suoja-aikaa. Tekijänoikeudella on kuitenkin myös poikkeuksia ja rajoituksia, joita noudatetaan esimerkiksi arkistojen ja opetustoiminnan kohdalla. Tekijänoikeuksiin voi perehtyä lisää esimerkiksi Tekijänoikeus.fi-palvelussa (<https://tekijanoikeus.fi/>).

Lähteet ja kirjallisuutta:

Arkistolaitoksen suositus digitoinnin laatukriteereiksi. 2008. Arkistolaitos. PDF. Saatavissa: <https://www.arkisto.fi/uploads/normit/valtionhallinto/suosituksset/digikriteerit.pdf>.

Digital Preservation Handbook. 2019. Digital Preservation Coalition. Saatavissa: <https://www.dpconline.org/handbook/>.

Digitoinnin suositukset. 2018. Talle ry. Saatavissa: <http://www.talle.fi/digitoinnin-suositukset>.

Digiwiki - digitoinnin ja sähköisen pitkäaikaissäilyttämisen verkkopalvelu. Saatavissa: <http://www.digiwiki.fi>.

Guide to archiving personal data. 2018. The National Archives. PDF. Saatavissa: <http://www.nationalarchives.gov.uk/documents/information-management/guide-to-archiving-personal-data.pdf>.

Hänninen, T. 2019. Avoimilla ohjeistuksilla ja oppailla tuetaan digitaalisen tiedon pitkäaikaissäilyttämistä. Teoksessa Kosonen, M. (toim.) Ratkaisuja digitaalisten aineistojen käytettävyyden parantamiseen. Mikkeli: Kaakkois-Suomen ammattikorkeakoulu, 34–39. PDF. Saatavissa: <https://www.theseus.fi/handle/10024/226884>.

Järvinen, P. 2009. Digiarkistointi. 1. painos. Jyväskylä: WSOYpro.

Laustela, V. 2013. Digitoinnin suunnittelu ja digitoinnin työnkulku. PDF. Saatavissa: <https://www.valokuvataiteenmuseo.fi/>.

Lybeck, J. 2006. ARKISTOT YHTEISKUNNAN TOIMIVA MUISTI - Asiakirjahallinnon ja arkistotoimen oppikirja. PDF. Saatavissa: https://www.arkisto.fi/uploads/Julkaisut/oppaat/asiakirjahallinnon_oppikirja.pdf.

Mäkelä, J. 2017. Kestävän äänen jäljillä 2. Faili 4/2017, 14-17.

Palonen, O. Asiakirjojen pitkäaikaissäilytys. WWW-dokumentti. Saatavissa: <https://docplayer.fi/10622828-Asiakirjojen-pitkaaikaissailytys-osmo-palonen-mikkelin-ammattikorkeakoulu.html>.

Poole, N. 2010. The Cost of Digitising Europe's Cultural Heritage. PDF. Saatavissa: http://nickpoole.org.uk/wp-content/uploads/2011/12/digiti_report.pdf.

Rännäli, J. 2015. Yleispäteviä ratkaisuja ei ole olemassa. Faili 4/2015, 14-16.

Teosten käyttäminen. Tekijanoikeus.fi. WWW-dokumentti. Saatavissa <https://tekijanoikeus.fi/tekijanoikeus/luvallinen-kaytto/>.