

KANSALLISARKISTON MÄÄRITYKSET ARKISTOITAVIEN ERIKOISAINEISTOJEN DIGITOINTIIN

Sisältö	Kansallisarkiston määräykset arkistoitavan erikoisaineiston digitoinnille. Määritykset mahdollistavat analogisen ilmentymän tietosisällön säilyttämisen yksinomaan digitaalisena ja alkuperäisen analogisen ilmentymän hävittämisen digitoinnin jälkeen.
Tarkoitus	Varmistaa kansalliseen kulttuuriperintöön kuuluvien ja arkistoitavaksi ¹ määrättyjen asiakirjojen sisältämän tietosisällön säilyminen ja niiden käytettävyys analogisen ilmentymän hävittämisen jälkeen. ² Tämän ohjeen tarkoituksena on toimia määrityksenä siitä, minkälainen digitointiprosessi mahdollistaa analogisen ilmentymän hävittämisen. Analogisten asiakirjojen hävittäminen edellyttää Kansallisarkiston seulontapäätöstä. Ohje täydentää vuonna 2019 julkaistua Kansallisarkiston vaatimukset hävittämiseen tähtäävään digitointiin (KA/3357/07.01.02.04.02/2019) asiakirjaa. Tätä ohjetta sovelletaan, kun digitoidaan arkistoitavia tai jo arkistoituja analogisia erikoisaineistoja.
Kohderyhmä	Ohje on tarkoitettu Kansallisarkistolle sekä niille toimijoille, jotka digitoivat arkistolain (831/1994) nojalla arkistoitavaksi määrättyjä analogisia asiakirjoja.
Rajaukset	Ohje ei koske skannausta, joka toteutetaan asiakirjan laatimisen tai asiankäsittelyn yhteydessä taikka organisaatiolle saapuvien analogisten asiakirjojen muuttamiseksi digitaaliseen muotoon. Erikoisaineistot koostuvat karttakokoelmista, piirustuskokoelmista, valokuvakokoelmista, negatiiveista, dioista, mikrofilmeistä sekä sidoksista, joissa sidosten tietosisältö esitetään aukeamittain (Taulukko 1). Ohje ei koske erikoisaineistoiksi luokiteltavia yksittäisiä asiakirjoja, jotka ovat arkistoitu A4- ja foliokokoisten asiakirjojen joukkoon.

¹ Arkistoitavilla asiakirjoilla tarkoitetaan tässä ohjeessa niitä asiakirjatietoja, jotka arkistolain 8 §:n nojalla on määrätty pysyvään säilytykseen.

² Kansallisarkiston päätöksellä säilytysajan päätyttyä asiakirjatiedot voidaan arkistoida yleisen edun mukaisia arkistointitarkoituksia, tieteellisiä tai historiallisia tutkimustarkoituksia sekä tilastollisia tarkoituksia varten.

Taulukko 1, Ohjeiden soveltaminen eri asiakirjatyyppien yhteydessä.

Aineisto	Kansallisarkiston määritykset arkistoitavien erikoisaineistojen digitointiin (2020)	Kansallisarkiston vaatimukset hävittämiseen tähtäävään digitointiin 2019
Karttakokoelma	x	
Piirustuskokoelma	x	
Valokuvakokoelmat	x	
Lasinegatiivit, muut negatiivit ja diat	x	
Mikrofilmit	x	
Sidokset, jossa sidosten tietosisältö esitetään aukeamittain	x	
Sidokset		x
A4- ja foliokokoiset asiakirjat		x
A4- ja foliokokoisten asiakirjojen joukossa olevat yksittäiset valokuvat, piirustukset ja kartat		x

Ohjeessa kuvataan digitointiprosessia ja sen lopputulosta. Ohjeessa ei kuvata varsinaista säilytyspakettia, joka tallennetaan pitkäaikaissäilytysjärjestelmään. Pitkäaikaissäilytysjärjestelmään tallennettava paketti on mahdollista muodostaa tässä ohjeessa esitetystä digitointiprosessin lopputuloksesta. Tässä ohjeessa ei oteta kantaa analogisen ilmentymän hävittämisprosessiin, vaan kuvataan ne kriteerit, joiden täytyttyä analogisen asiakirjan hävittäminen olisi mahdollista.

Säädökset, johon ohjeen antaminen perustuu

Laki Kansallisarkistosta 1145/2016 § 2

Voimassaoloaika

Toistaiseksi, ohjeen päiväyksestä eteenpäin.

Kumoaa ohjeen ja suosituksen

Arkistolaitoksen tekniset vaatimukset digitoidulle aineistolle 24.8.2016 ja 15.8.2008, AL/11130/07.01.02.04.02/2008.



Sisällysluettelo

1	Termit ja käsitteet	1
2	Johdanto.....	3
3	Yleiset digitointiprosessin suositukset ja hyvät käytänteet	3
4	Yleiset digitointiprosessin määritykset	4
5	Aineistotyyppikohtaiset määritykset	5
5.1	Kartta- ja piirustuskokoelmat	5
5.2	Sidokset, jossa sidosten tietosisältö esitetään aukeamittain	6
5.3	Valokuvakokoelmat	7
5.4	Lasinegatiivit, muut negatiivit ja diat.....	8
5.5	Mikrofilmit.....	9
6	Hyväksyttävät formaatit.....	9
6.1	Kuvatiedosto.....	9
6.2	Kuvatiedostoa ja kuvatiedoston prosessointia kuvaavat metatiedot ja rakenne.....	13
7	Digitointiprosessissa muodostettava siirtopaketti	15
8	Esimerkkipaketit	16
9	Allekirjoitukset	17

1 Termit ja käsitteet

Ohjeessa käytetty termistö perustuu Internet Engineering Task Force:n toimesta tehtyyn määrittelyyn [RFC 2119].³ Alla olevassa taulukossa (Taulukko 2) ilmaistaan, mitä käännöksiä termistöstä tässä ohjeessa käytetään.

Taulukko 2, Tässä asiakirjassa käytetyt käännökset

ENGLANTI	SUOMI
MUST	PITÄÄ
MUST NOT	EI SAA
REQUIRED	PAKOLLINEN
SHOULD	PITÄISI
SHOULD NOT	EI PITÄISI
MAY	SAA
OPTIONAL	VAPAAEHTOINEN

Alla olevassa taulukossa (Taulukko 3) ilmaistaan, mitä tässä asiakirjassa taulukossa esitetyillä käsitteillä tarkoitetaan:

Taulukko 3, Käsitteistö

KÄSITE	SELITE
Analoginen asiakirja	Paperisessa tai muussa käsin kosketeltavassa muodossa laadittu, säilytetty ja käytettävä asiakirja.
Analoginen ilmentymä	Digitoitavaksi päätetyn analogisen asiakirjakokonaisuuden analoginen ilmiä, joka koostuu erilaisista analogista asiakirjoista.
Analogisen asiakirjan kulttuurihistoriallinen arvo	Arkistoitavan tai arkistoidun analogisen asiakirjan ominaisuus, kulttuurihistoriallinen arvo, jonka vuoksi analoginen asiakirja säilytetään digitoinnista huolimatta myös analogisessa muodossa.
Arkistointi	Arkistoinnilla tarkoitetaan toimenpiteitä, joilla varmistetaan arkistoidun tiedon autenttisuus, luotettavuus, eheys ja käytettävyys. Arkistointi alkaa säilytysajan päättymisen jälkeen. Arkistointi tarkoittaa myös arkistolain (831/1994) mukaista pysyvää säilyttämistä (8 § 3 momentti).
Asiakirja	Asiakirja on todistusvoimaista tietoa. Toimija on tuottanut tai vastaanottanut tiedon osana tehtäviään ja säilyttää sitä tietovarantona sekä todisteena. Asiakirja on sisällön, rakenteen/ilmiäsun ja kontekstin kokonaisuus.
Digitaalinen ilmentymä	Digitoitavaksi päätetyn analogisen asiakirjakokonaisuuden digitaalinen ilmiäsu.

³ <https://www.ietf.org/rfc/rfc2119.txt> Viitattu 9.6.2020

KÄSITE	SELITE
Tiedosto	Joko yksin tai muiden tiedostojen kanssa muodostaa digitoidun asiakirjan. Tässä asiakirjassa tiedosto on joko bittikarttakuva tai XML-tiedosto.
Digitointiprosessi	Joukko toimintoja, joiden avulla analoginen ilmentymä muunnetaan digitaaliseksi ilmentymäksi.
Digitoitu asiakirja	Analogisesta asiakirjasta digitointiprosessilla tuotettu digitaalinen versio, joka voi koostua n-määrästä tiedostoja.
Erikoisaineisto	Koostuu omina kokonaisuuksina/kokoelmina arkistoiduista kartoista, piirustuksista, valokuvista, negatiiveista, mikrofilmeistä ja mikrokorteista sekä sidoksista, joissa sidosten tietosisältö esitetään aukeamittain.
Kuva-ala	Kuvatiedostosta visuaalisella tarkastelulla näkyvä informaatio.
Kuvatiedosto	Digitointiprosessissa tuotettu bittikarttakuva. Tarkentaa digitaalisen tiedoston muodon bittikarttakuvaksi.
Käyttökappale⁴	Tiedosto, joka tarjotaan käyttöön erilaisiin käyttöliittymiin. Vertaa tallekappale.
Pysyvä säilyttäminen	Tarkoittaa arkistolain (831/1994) 8 §:n mukaista asiakirjatietojen pysyvää säilyttämistä. EU:n yleisessä tietosuojasetuksessa (2016/679) ja tiedonhallintalaissa (906/2019) vastaava käsite on arkistointi.
Päälukusuunta	Mahdollistaa asiakirjan tietosisällön tulkitsemisen kuvatiedostoa kääntämättä. Mikäli asiakirjassa esiintyy tietosisältöä useampaan lukusuuntaan, tarkoittaa päälukusuunta sitä suuntaa, jossa suurin osa asiakirjan tietosisällöstä on luettavissa.
Seulontapäätös	Kansallisarkiston päätös arkistolain 8, 11 ja 14 a §: ien nojalla pysyvästi säilytettävistä (arkistoitavista) asiakirjoista ja niiden säilytysmuodosta.
Siirtopaketti	Digitointiprosessissa muodostettu kokonaisuus, joka on siirrettävissä Kansallisarkiston tietojärjestelmiin.
Digitointi	Analogisessa muodossa säilytettävien tai arkistoitujen asiakirjojen digitaaliseen muotoon muuttaminen.
Tallekappale⁵	Digitointiprosessissa tuotettu tiedosto (bittikarttakuva), joka on teknisiltä ominaispiirteiltään laadukkaita digitointiprosessissa tuotettu kuvatiedosto. Tallekappale yleisesti viedään pitkäaikaissäilytysjärjestelmiin.
Tuotantovuorokausi	Vuorokausi, jonka aikana laitteella tuotetaan digitaalisia tiedostoja.

⁴ FADGI, käyttökappale: <http://www.digitizationguidelines.gov/term.php?term=derivativefile> Viitattu 9.6.2020

⁵ FADGI, tallekappale: <http://www.digitizationguidelines.gov/term.php?term=productionmasterfile> Viitattu 9.6.2020

2 Johdanto

Digitointi edistää omalta osaltaan digitalisaatiota ja on osa kulttuuriperintöaineiston säilyvyyden ja käytettävyyden edistämisen kokonaisuutta. Tässä ohjeessa esitetään määräyksiä, jotka kohdistuvat erikoisaineistojen digitointiprosessiin. Esitettyjä määräyksiä PITÄÄ noudattaa, kun julkishallinnon toimija digitoi arkistoitavaksi määrättyjä analogisia erikoisaineistoja.

Digitoinnin kohteena olevilla asiakirjoilla PITÄÄ olla Kansallisarkiston seulontapäätös. Seulontapäätöksessä määrätään analogisten asiakirjojen kulttuurihistoriallisesta arvosta ja digitoinnin jälkeisestä säilytysmuodosta. Analogisia ilmentymiä EI SAA hävittää ennen kuin seulontapäätös on tehty.

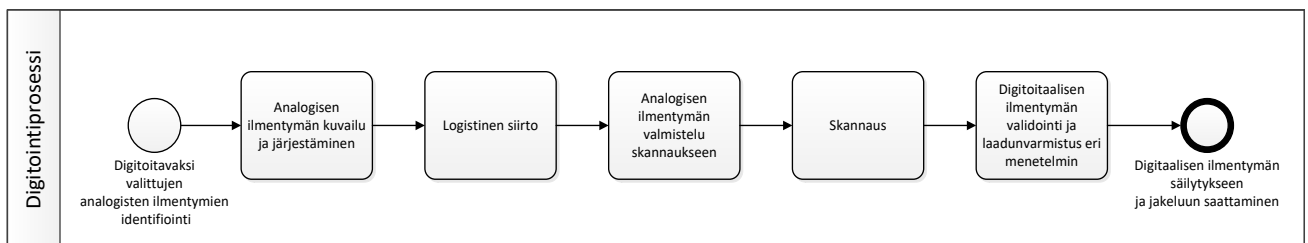
Analogisten ilmentymien tietosisällön arkistointi yksinomaan digitaalisessa muodossa edellyttää, että digitaaliseen muotoon muuttaminen on toteutettu menetelmillä, jotka eivät heikennä asiakirjan todistusvoimaisuutta, eheyttä ja autenttisuutta. Tämän ohjeen noudattaminen digitointiprosessissa osaltaan turvaa digitoinnin kohteena olevien analogisten ilmentymien tietosisällön säilyvyyden yksinomaan digitaalisena.

Digitaalisten ilmentymien vastaanottaminen Kansallisarkiston tietojärjestelmiin edellyttää, että digitaaliset ilmentymät täyttävät tässä ohjeessa esitetyt määräykset. Tämän ohjeen laadinnassa on huomioitu arkistosektorilla yleisesti käytössä olevat standardit sekä muiden Kansallisarkistojen laatuvaatimukset digitoinnille. Lisäksi luvuissa hyväksyttävät formaatit ja digitointiprosessissa muodostettava paketti, on huomioitu kansalliset pitkäaikaissäilytyspalveluiden (PAS-palvelut) asettamat määräykset säilytettävälle aineistoille.⁶

3 Yleiset digitointiprosessin suositukset ja hyvät käytänteet

Tässä luvussa kuvataan yleisiä hyviä käytänteitä, jotka liittyvät skannaukseen ja skannauksen laadunvarmistukseen. Tässä luvussa ei käytetä taulukossa 2 esitettyjä termejä.

Digitointi käsitetään yleisesti prosessina, joka sisältää alla olevassa kuvassa (Kuva 1) esitetyt vaiheet.



Kuva 1, Digitaaliseen muotoon muuntamisen prosessi - yleinen

On hyvä korostaa, että digitointiprosessissa laadunvarmistusta ei voi kohdentaa vain tiettyyn vaiheeseen prosessissa, vaan digitoinnin lopputuloksen luotettavuus muodostuu siitä, että

⁶ <http://digitalpreservation.fi/files/PAS-tiedostomuodot-1.8.0.pdf> Viitattu 9.6.2020

laadunvarmistusta tehdään prosessin eri vaiheissa. Tässä ohjeessa keskitytään erityisesti skannauksen laadunvarmistukseen.

Digitoitavat asiakirjat on hyvä kuvaila aineistometatietojärjestelmään ennen niiden digitaaliseen muotoon muuntamisen aloittamista. Tällöin analogisen ilmentymän käsittelyketjua voidaan dokumentoida kattavammin ja esimerkiksi tietoja aineiston valmistelusta voidaan dokumentoida. Skannauksen jälkeen aineiston metatietoja voidaan rikastaa joko manuaalisesti tai automaattisin menetelmin.

Skannauksen laadunvarmistus voidaan karkeasti jakaa ennen skannaustapahtumaa tapahtuvaan toimintaan ja sen jälkeiseen laadunvarmistamiseen eli validointiin.

Skannausinfrastruktuurin suorituskyky optimoidaan ennen skannaustapahtumaa siten, että sen tuottama digitaalinen ilmentymä edustaa parasta mahdollista ilmentymää, joka kyseisellä teknisellä kokoonpanolla voidaan tuottaa. Optimoinnin jälkeen infrastruktuurin suorituskykyä tulisi tarkkailla suunnitellusti, jotta prosessissa tuotettavien digitaalisten ilmentymien laatu säilyy tasaisena. Tarkkailua varten tarvitaan yleisesti mittataulu, mittataulun referenssiarvot ja analysointiohjelmisto. Kuvanlaadun lisäksi laiteinfrastruktuurissa pitäisi kiinnittää huomiota siihen, että voidaan varmistua analogisten asiakirjojen muuntuvaan digitaaliseen muotoon tietosisällöltään täydellisinä. Tämä tarkoittaa yhtäältä sitä, että skannausprosesseissa voidaan välttyä esimerkiksi kahden päällekkäisen analogisen asiakirjan menemisestä laitteen läpi (läpisyöttöskannerit, avoratskannerit ja muut skannausratkaisut, missä asiakirjoja skannataan ”massana”) ja toisaalta sitä, että infrastruktuurin tuottama kuvatiedosto ei ole vääristynyt tai muutoin mittasuhteiltaan muuttunut, joka aiheuttaa tietosisällön tulkinnan suhteen haasteita.

Skannauksen jälkeinen validointi voidaan toteuttaa otannoilla. Otannan määrä on riippuvainen skannausprosessin luotettavuudesta. Validoinnin tavoitteena on varmistua siitä, että tässä ohjeessa esitetyt määritykset täyttyvät.

Jos kuvatiedostoja käsitellään skannaustapahtuman jälkeen, pitäisi yksityiskohtainen kuvankäsittelyhistoria tallentaa ainakin kuvatiedostojen metatietoihin. Mahdollisuuksien mukaan myös kuvatiedoston syntyä kuvaileviin XML –tietoihin (MIX).

Huomioitavaa on myös se, että jokainen migraatio ja tiedostokonversio (tiedoston tallennusformaatin muuttaminen) sisältää riskejä liittyen tiedon säilymiseen eheänä. Tämän takia olisi tärkeää digitoida analoginen aineisto digitoinnin yhteydessä tässä ohjeessa esitettyyn tallennusformaattiin. Heikompilaatusesta tallennusformaattista ei voi tuottaa parempilaatuista myöhemmässä käsittelyssä ja se on myös tämän ohjeen määrityksissä kielletty.

4 Yleiset digitointiprosessin määritykset

Analogisten ilmentymien digitaaliseksi muuntaminen on prosessi (digitointiprosessi), jonka vaiheita PITÄÄ dokumentoida tässä ohjeessa ilmaistuin tavoin ja tarkkuuksin. Prosessin dokumentoinnilla tarkoitetaan, että skannauksesta sekä kuvien mahdollisesta käsittelystä tallennetaan näitä toimenpiteitä dokumentoivat metatiedot. Näiden metatietojen lisäksi

digitaaliseen muotoon muuttamisen prosessista SAA tallentaa metatiedoiksi myös muita toimenpiteitä.

Digitointiprosessissa PITÄÄ varmistua, että digitoitavaksi tarkoitettu kokonaisuus tulee digitoitua kokonaisuutena ja sisällöllisesti täydellisenä. Tämä tarkoittaa käytännössä sitä, että kaikki digitoitavaksi päätetyn kokonaisuuden analogiset asiakirjat PITÄÄ digitoida siten, että mitään informaatiota ei jää teknisen tai toiminnallisen virheen takia muuntamatta digitaaliseen muotoon.

Jokaisesta digitoitavaan kokonaisuuteen liittyvästä yksittäisestä kuvatiedostosta PITÄÄ olla visuaalisella tarkastelulla saatavissa sama informaatio kuin sen analogisesta vastineesta. Kuvatiedosto EI SAA sisältää mitään elementtejä, joita ei ilmene analogisessa vastineessa. Tästä poikkeuksen muodostavat mahdolliset samaan kuvatiedostoon skannattavat/kuvattavat kuvatiedoston värejä, harmaasävyjä, mittasuhteita ja resoluutiota todentavat skannaustekniset mittataulut. Kyseiset mittataulut PITÄÄ asetella siten, että ne eivät peitä digitoitavaa kohdetta.

Digitointiprosessissa EI SAA poistaa merkintöjä sisältäviä sivuja. Digitointiprosessissa tuotetut kuvatiedostot PITÄÄ olla käännetty päälukusuuntaan. Digitointiprosessissa tuotettuja kuvatiedostoja SAA kääntää niiden skannauksen jälkeen vain 90 asteen välein.

5 Aineistotyyppikohtaiset määritykset

Tässä luvussa ja sen alaluvuissa luetellaan aineistotyyppikohtaisia määrityksiä digitointiprosesseille sekä prosessissa tuotettaville tallekappaleille. Yleisesti määritykset digitaalisiin tiedostoihin liittyen ilmaistaan luvussa 6. Aineistotyyppit on kuvattu taulukossa 1. Luvussa 5 aineistotyyppinä on yhdistelty, mikäli aineistotyyppihin kohdistetut määritykset ovat keskenään samanlaiset.

5.1 Kartta- ja piirustuskokoelmat

Erillisinä kokoelmina olevat kartta- ja piirustuskokoelmat PITÄÄ digitoida kuvatiedostoiksi siten, että syntyvä kuvatiedosto on fyysisiltä mitoiltaan (n cm x n cm) 1:1 koossa analogisen asiakirjan kanssa resoluutiolla 300 ppi. Digitoinnin kohteena olevan analogisen asiakirjan tietosisällön eheyden varmistamiseksi PITÄÄ analoginen asiakirja digitoida siten, että kuva-ala ylittää sivujen reunat. Asiakirjan PITÄÄ olla kuvaushetkellä tasainen. Mikäli tasaisuuden saavuttamiseksi käytetään lasia tai muuta apuvälinettä, PITÄÄ se huomioida värinhallinnassa. Kartat ja piirustukset digitoidaan väritiedostoiksi (RGB, 24 bit).

Ennen skannaustapahtumaa PITÄÄ digitoinnissa käytetyn infrastruktuurin suorituskyky optimoida käyttäen tähän tarkoitukseen soveltuvia mittatauluja ja ohjelmistoja. Optimoinnin jälkeen PITÄISI infrastruktuurin tuottamien kuvatiedostojen laatu todentaa kerran tuotantovuorokaudessa käyttämällä tähän tarkoitukseen tarkoitettuja mittatauluja.

Alla olevassa taulukossa (Taulukko 4) esitetään kartta- ja piirustuskokoelmien digitointiin kohdistuvat määritykset.

Taulukko 5, Sidosten määrittelyt

Elementti	Vaadittu arvo
Tallekappaleen formaatti	TIFF 6.0 tai JPEG part 1 versio
Tallekappaleen pakkaaminen	TIFF = LZW tai JPEG = JPEG -laatuaste 90%
Resoluutio	300 ppi
Väritila	2 = RGB
Bittisyvyys	8 bittiä kanava, 3 kanavaa (24 bit)
Tallekappaleen ICC-profiili	eciRGB v2, ProPhoto RGB, AdobeRGB (1998), sRGB

5.3 Valokuvakokoelmat

Valokuvien sisältämän tietosisällön eheyden varmistamiseksi analogiset asiakirjat PITÄÄ digitoida kokonaisuudessaan siten, että kuva-ala ylittää sivujen reunat. Valokuvat PITÄÄ digitoida joko väritiedostoiksi (RGB, 24 bit) tai harmaasävytiedostoiksi (8 bit) riippuen analogisen asiakirjan sisältämästä informaatiosta. Mikäli analoginen ilmentymä päädytään digitoimaan harmaasävydigitoitina, digitoinnin kohteena oleva analoginen asiakirja EI SAA sisältää värillistä informaatiota, joka on tietosisällön tulkinnan kannalta nyt tai tulevaisuudessa relevanttia.

Alla olevassa taulukossa (Taulukko 6, Valokuvakokoelmien) esitetään valokuvakokoelmien digitointiin kohdistuvat määrittelyt.

Taulukko 6, Valokuvakokoelmien määrittelyt

Elementti	Vaadittu arvo
Tallekappaleen formaatti	TIFF 6.0
Tallekappaleen pakkaaminen	LZW
Resoluutio	Analogisen ilmentymän mukaan, ks. Taulukko7 Resoluutiotaulukko
Väritila	Värikuva: 2 = RGB Harmaasävykuva: 0 = WhitelsZero TAI 1 = BlacksZero
Bittisyvyys	RGB = 8 bittiä kanava, 3 kanavaa (24 bit) Harmaasävy = 8 bittiä kanava, 1 kanava (8 bit)
Tallekappaleen ICC-profiili	RGB: eciRGB v2, ProPhoto RGB, AdobeRGB (1998) Harmaasävy: Gray gamma 2,2

Digitoinnin kohteena olevan analogisen asiakirjan ollessa pienikokoista, PITÄÄ resoluutiota nostaa alla olevan taulukon (Taulukko 7) mukaisesti. Mikäli analoginen asiakirja on fyysisiltä mitoiltaan pienempi mitä alla olevassa taulukossa 7 esitetään, PITÄÄ samaa laskentakaavaa soveltaen digitoinnissa käytettävää resoluutiota kasvattaa. Tallekappaleetiedoston resoluutiota EI SAA muuttaa kuvatiedoston käsittelyprosessissa, jotta analogisen asiakirjan fyysisten mittojen arviointi on digitaalisen tallekappaleen perusteella mahdollista. Tavoitteena on aina ~A3 –kokoinen (cm x cm) tallekappale resoluutiolla 300ppi.

Taulukko 7, Resoluutiotaulukko

Analogisen asiakirjan koko	Resoluutio
6x9cm	1600 ppi
9x12cm	1200 ppi
13x18cm	900 ppi
18x24cm	600 ppi
A3 (tai suurempi)	300 ppi

5.4 Lasinegatiivit, muut negatiivit ja diat

Lasinegatiivien, muiden negatiivien ja diojen tietosisällön eheyden varmistamiseksi analogiset asiakirjat PITÄÄ digitoida kokonaisuudessaan siten, että kuva-ala ylittää digitoitavan kohteen reunat. Lasinegatiivit, muut negatiivit ja diat PITÄÄ digitoida joko väritiedostoiksi (24 bit) tai harmaasävytiedostoiksi (8 bit) riippuen analogisen asiakirjan sisältämästä informaatiosta. Mikäli analogiset asiakirjat päädytään digitoimaan harmaasävydigitoitina, digitoinnin kohteena oleva analoginen asiakirja EI SAA sisältää värillistä informaatiota, joka tietosisällön tulkinnan kannalta on nyt tai tulevaisuudessa relevanttia.

Alla olevassa taulukossa (Taulukko 8) esitetään lasinegatiivien, muiden negatiivien ja diojen digitointiin kohdistuvat määritykset.

Taulukko 8, Lasinegatiivien, muiden negatiivien ja diojen määritykset

Elementti	Vaadittu arvo
Tallekappaleen formaatti	TIFF 6.0
Tallekappaleen pakkaaminen	LZW
Resoluutio	Analogisen ilmentymän mukaan, ks. Taulukko 7 Resoluutiotaulukko
Väritila	Värikuva: 2 = RGB Harmaasävykuva: 0 = WhitelsZero TAI 1 = BlackIsZero
Bittisyvyys	RGB = 8 bittiä kanava, 3 kanavaa (24 bit) Harmaasävy = 8 bittiä kanava, 1 kanava (8 bit)
Tallekappaleen ICC-profiili	RGB: eciRGB v2, ProPhoto RGB, AdobeRGB (1998) Harmaasävy: Gray gamma 2,2

Digitoinnin kohteena olevan analogisen asiakirjan ollessa pienikokoista, PITÄÄ resoluutiota nostaa taulukon 7 mukaisesti. Mikäli analoginen asiakirja on fyysisiltä mitoiltaan pienempi mitä taulukossa 7 esitetään, PITÄÄ samaa laskentakaavaa soveltaen digitoinnissa käytettävää resoluutiota kasvattaa. Tallekappaleetiedoston resoluutiota EI SAA muuttaa kuvatiedoston käsittelyprosessissa, jotta analogisen asiakirjan fyysisten mittojen arviointi on digitaalisen tallekappaleen perusteella mahdollista. Tavoitteena on aina ~A3 –kokoinen (cm x cm) tallekappale resoluutiolla 300ppi.

5.5 Mikrofilmit

Mikrofilmin pienennyskerroin PITÄÄ tarkistaa filmirullan alusta. Mikäli pienennyskerrointa ei ole ilmoitettu, PITÄÄ suurennuskerroin asettaa riittävän suureksi, jotta skannausprosessissa tuotettu digitaalisen asiakirjan sisältämä tietosisältö on luettavissa. Mikrofilmien osalta suurennuskertoimen avulla saadaan mikrofilmille kuvattu aineisto toistumaan riittävän luettavana resoluutiolla 300 ppi. Skannauksessa käytettävä resoluutio on suurennuskerroin x haluttu resoluutio (esimerkiksi 12 x 300 = 3600). Mikrofilmi itsessään on varsin jyrkkä tallenneväline, joten digitoinnissa PITÄÄ varoa jyrkkyyden korostamista, jotta luvussa 4 esitetyt määritykset toteutuvat.

Alkuperäisen aineiston oikean koon todentamisessa voi olla käytetty myös muita menetelmiä (esimerkiksi mittanauha), joita PITÄÄ tarkastella tapauskohtaisesti. Tapauskohtaisessa arvioinnissa PITÄÄ pyrkiä arvioimaan alkuperäisen, mikrofilmille kuvatun analogisen asiakirjan fyysisiä mittoja, vaikka mitään edellä mainituista menetelmistä ei olisi mikrokuvauksen yhteydessä sovellettu.

Alla olevassa taulukossa (Taulukko 9) esitetään mikrofilmien digitointiin kohdistuvat määritykset.

Taulukko 9, Mikrofilmien määritykset

Elementti	Vaadittu arvo
Tallekappaleen formaatti	TIFF 6.0 tai JPEG part 1 versio 1.00 tai 1.01 tai 1.02
Tallekappaleen pakkaaminen	TIFF = LZW tai JPEG = JPEG -laatus 90%
Resoluutio	300 ppi
Väritila	Harmaasävy
Bittisyvyys	8 bittiä kanava, 1 kanava (8 bit)
Tallekappaleen ICC-profiili	Gray gamma 2,2

Mikäli kyseessä on värillinen mikrofilmi, PITÄÄ siihen soveltaa tässä luvussa esitettyjä vaatimuksia, mutta tallekappaleen väritilan PITÄÄ olla RGB, bittisyvyyden PITÄÄ olla 24 bit ja tallekappaleen ICC-profiilin PITÄÄ olla eciRGB v2, ProPhoto RGB, AdobeRGB (1998) tai sRGB.

6 Hyväksyttävät formaatit

Formaattiosio on jaettu kahteen alaluokkaan:

1. Kuvatiedosto
2. Kuvatiedostoa ja kuvatiedoston prosessointia kuvaavat metatiedot ja rakenne

6.1 Kuvatiedosto

Digitointiprosessissa tuotettava kuvatiedosto EI SAA missään käsittelyvaiheessa olla tässä luvussa esitettyjä vaatimuksia heikkolaatuisempi. Tallekappale PITÄÄ tallentaa joko pakkaamattomassa/häviöttömästi pakatussa TIFF-muodossa tai häviöllisesti pakatussa JPEG-muodossa (ei molemmissa).⁷ Tallekappaleen tallennusformaatti on riippuvainen analogisen aineiston olomuodosta. Aineistotyyppikohtaiset määritykset esitetään luvussa 5 ja sen alaluvuissa.

⁷ JPEG: <https://jpeg.org/jpeg/index.html> Viitattu 2.9.2020

TIFF: <https://www.itu.int/itudoc/itu-t/com16/tiff-fx/docs/tiff6.pdf> Viitattu 2.9.2020

Mikäli prosessissa ensimmäisenä tuotettu kuvatiedosto (tallekappale) on pakattu, sitä EI SAA käsitellä sen ensimmäisen tallennuskerran jälkeen ja tallentaa tämän jälkeen uudelleen tallekappaleeksi.

Alla olevissa taulukoissa 10 ja 11 esitetään tiedostoformaateittain pakolliset tiedot, jotka kuvatiedostossa PITÄÄ olla koneymmärrettävässä muodossa. Mikäli taulukoiden ”Elementti” – saraketta ei tarkenneta, PITÄÄ tieto ilmaista, mutta tiedolle ei ole tässä yhteydessä määritelty vaadittavaa kenttää. Taulukossa esitettyjen tietojen lisäksi kuvatiedosto SAA sisältää muita metatietokenttiä.

Taulukoissa 10 ja 11 esitetään kaikki mahdolliset arvot, joita tässä ohjeessa määriteltyjen kuvatiedostojen tallennusformaattien osalta PITÄÄ ilmaista. Toisin sanoen taulukoiden 10 ja 11 arvoja PITÄÄ tulkita yhdessä luvussa 5 Aineistotyyppikohtaiset ja sen alaluvuissa esitettyjen aineistotyyppikohtaisten vaatimusten kanssa. Edellä mainitut kohdat ilmaistaan sarakkeessa ”Vaadittu arvo” arvolla ”Aineistotyyppi-riippuvainen”.

Mikäli sarakkeessa ”Vaadittu arvo” lukee ”Tiedostokohtainen”, on jokaisella tiedostolla yksilöllinen arvo. Mikäli sarakkeessa ilmaistaan ”Tuotantoprosessi-riippuvainen” on arvo riippuvainen prosessista, jonka avulla analoginen ilmentymä muunnetaan digitaaliseksi ilmentymäksi.

Taulukko 10, Kuvatiedoston (TIFF) pakolliset metatiedot

Elementti	Tarkenne	Vaadittu arvo	Metatieto-skeema	Metatietokenttä
Formaatti	TIFF	image/tiff		MIME Type
Versio	TIFF-tiedoston vaadittu versio	6.0		
Kuvan nimi	Kuvan nimi ilmaistaan neljällä numerolla 0001.tif, 0002.tif, 0003.tif jne, Ks. luku 7.			
Kuvatiedoston koko				
Väritila	Kuvatiedoston väritila	Aineistotyyppi-riippuvainen. 0 = WhitesZero tai 1 = BlacksZero tai 2 = RGB	Exif.Image	PhotometricInterpretation (262)
ICC-profiili	Kuvatiedoston metatietoihin tallennettu (embedded) väriprofiili.	Aineistotyyppi-riippuvainen. sRGB, eciRGB v2, ProPhoto RGB, AdobeRGB (1998) tai Gray gamma 2,2	TIFF tag, private	ICC Profile (34675)
Bittisyvyys	Bittien määrä pikselin kanava-arvossa	8	Exif.Image	BitsPerSample (258)
	Kanava-arvojen määrä pikselissä	Aineistotyyppi-riippuvainen. 3 tai 1	Exif.Image	SamplesPerPixel (277)
Tiedoston pakkaaminen		5 = LZW	Exif.Image	Compression (259)
Kuvan leveys	Kertoo kuvan leveyden pikselien määrällä per rivi	Tiedostokohtainen	Exif.Image	ImageWidth (256)

Elementti	Tarkenne	Vaadittu arvo	Metatieto- skeema	Metatietokenttä
Kuvan korkeus	Kertoo kuvan korkeuden pikselirivien määrällä kuvassa	Tiedostokohtainen	Exif.Image	ImageLenght (257)
Digitaalisen kuvatiedoston tekijä	Organisaatio, joka on luonut kuvatiedoston analogisesta ilmentymästä	Tuotantoprosessi-riippuvainen	Exif.Image	Artist (315)
Digitointilaite (skannaus tai kuvaus)	Kertoo minkä valmistajan laitteella analoginen ilmentymä on muutettu sähköiseen muotoon (valmistajan nimi)	Tuotantoprosessi-riippuvainen	Exif.Image	Make (271)
Digitointilaitteen malli (skannaus tai kuvaus)	Tarkentaa digitointilaitetta kertomalla valmistajan mallin nimen	Tuotantoprosessi-riippuvainen	Exif.Image	Model (272)
Digitoinnissa käytetyn laitteen sarjanumero	Yksilöi käytetyn laitteen	Tuotantoprosessi-riippuvainen	Exif.Image	CameraSerialNumber (50735)
Kuvatiedoston luomisessa käytetty ohjelma	Sovellus ja versio, millä digitaalinen tiedosto on luotu	Tuotantoprosessi-riippuvainen	Exif.Image	Software (305)
Kuvatiedoston luontipäivämäärä ja aika (skannauspäivämäärä)	Ilmaistaan muodossa: YYYY:MM:DDTHH:MM:SS	Tiedostokohtainen	Exif.Image	DateTimeOriginal (36867)
Lukusuunta	Tiedoston lukusuunta (vaaka tai pysty). Lukusuunta ei ota kantaa kuvan tietosisällön lukusuuntaan, vaan tässä ilmaistaan tiedoston lukusuunta.	Tuotantoprosessi-riippuvainen/ Tiedostokohtainen	Exif.Image	Orientation (274)
Resoluution mittayksikkö	Mittayksikkö, jota käytetään tulkitessa X ja Y resoluutiota	2 = inch	Exif.Image	ResolutionUnit (296)
XResoluutio	Pikselien määrä resoluution mittayksikköä kohti leveysuunnassa.	Aineistotyyppi-riippuvainen	Exif.Image	XResolution (282)
YResoluutio	Pikselien määrä resoluution mittayksikköä kohti pystysuunnassa.	Aineistotyyppi-riippuvainen	Exif.Image	YResolution (283)
Tavujärjestys		big endian tai little endian		ByteOrder
Kuvatiedoston käsittelyohjelma	Mikäli digitointiprosessissa luodaan ensin pakkaamaton tiedosto, jota käsitellään skannauksen jälkeen, tallennetaan	Tuotantoprosessi-riippuvainen	Exif.Image	Image.ProcessingSoftware (11)

Elementti	Tarkenne	Vaadittu arvo	Metatieto-skeema	Metatietokenttä
	käsittelyohjelman nimi ja versio			

Taulukko 11, Kuvatiedoston (JPEG) pakolliset metatiedot

Elementti	Tarkenne	Vaadittu arvo/selite	Metatieto-skeema	Metatietokenttä
Formaatti	JPEG	image/jpeg		MIME Type
Versio	JPEG part 1 versio	1.00 tai 1.01 tai 1.02		JFIF Version
Kuvan nimi	Kuvan nimi ilmaistaan neljällä numerolla 0001.jpg, 0002.jpg, 0003.jpg jne, Ks. luku 7.			
Kuvatiedoston koko				
Väritila	Kuvatiedoston väritila	Aineistotyyppi-riippuvainen. 0 = WhitesZero tai 1 = BlacksZero tai 2 = RGB	Exif.Image	PhotometricInterpretation (262)
ICC-profiili	Kuvatiedoston metatietoihin tallennettu (embedded) väriprofiili.	Aineistotyyppi-riippuvainen. sRGB, eciRGB v2, ProPhoto RGB, AdobeRGB (1998) tai Gray gamma 2,2	ICC	profileDescription
Bittisyvyys	Bittien määrä pikselin kanava-arvossa	8	Exif.Image	BitsPerSample (258)
	Kanava-arvojen määrä pikselissä	Aineistotyyppi-riippuvainen. 3 tai 1	Exif.Image	SamplesPerPixel (277)
Tiedoston pakkaaminen	JPEG	6 = JPEG (old-style) tai 7 = JPEG	Exif.Image.	Compression (259)
JPEG-laatu	JPEG-pakkauksen laatu asteikolla 0-100 %	90 %		
Kuvatiedoston tekijä	Organisaatio, joka on luonut kuvatiedoston analogisesta ilmentymästä	Tuotantoprosessi-riippuvainen	Exif.Image	Artist (315)
Kuvan korkeus	Kertoo kuvan korkeuden pikselirivien määrällä kuvassa	Tiedostokohtainen	Exif.Image	ImageLength(257)
Kuvan leveys	Kertoo kuvan leveyden pikselirivien määrällä kuvassa	Tiedostokohtainen	Exif.Image	ImageWidth(256)
Digitointilaite	Kertoo minkä valmistajan laitteella analoginen asiakirja on muutettu digitaaliseksi (valmistajan nimi)	Tuotantoprosessi-riippuvainen	Exif.Image	Make (271)

Elementti	Tarkenne	Vaadittu arvo/selite	Metatietoskeema	Metatietokenttä
Digitointilaitteen malli	Tarkentaa digitointilaitetta kertomalla valmistajan mallin nimen	Tuotantoprosessi-riippuvainen	Exif.Image	Model (272)
Digitoinnissa käytetyn laitteen sarjanumero	Tarkentaa mallia ja yksilöi laitteen, jonka avulla analoginen asiakirja on muunnettu digitaaliseksi	Tuotantoprosessi-riippuvainen	Exif.Image	CameraSerialNumber (50735)
Kuvatiedoston luomisessa käytetty ohjelma	Sovellus ja versio, millä digitaalinen tiedosto on luotu	Tuotantoprosessi-riippuvainen	Exif.Image	Software (305)
Kuvatiedoston luontipäivämäärä ja aika	Ilmaistaan muodossa: YYYY:MM:DD HH:MM:SS	Tiedostokohtainen	Exif.Image	DateTime (306)
Lukusuunta	Tiedoston lukusuunta (horisontaalinen tai vertikaalinen)	Tuotantoprosessi-riippuvainen/ Tiedostokohtainen	Exif.Image	Orientation (274)
Resoluution mittausyksikkö	Mittayksikkö, jota käytetään tulkitessa X ja Y resoluutiota	2 = inch	Exif.Image	Image.ResolutionUnit (296)
XResoluutio	Pikselien määrä resoluution mittayksikköä kohden kuvan leveysuunnassa.	Aineistotyyppi-riippuvainen/ Tiedostokohtainen	Exif.Image	Image.XResolution (282)
YResoluutio	Pikselien määrä resoluution mittayksikköä kohden kuvan korkeusuunnassa.	Aineistotyyppi-riippuvainen/ Tiedostokohtainen	Exif.Image	Image.YResolution (283)
Kuvatiedoston käsittelyohjelma	Mikäli digitointiprosessissa luodaan ensin pakkaamaton tiedosto, jota käsitellään skannauksen jälkeen, tallennetaan käsittelyohjelman nimi ja versio	Tuotantoprosessi-riippuvainen	Exif.Image	Image.ProcessingSoftware (11)

6.2 Kuvatiedostoa ja kuvatiedoston prosessointia kuvaavat metatiedot ja rakenne

Tässä luvussa määritellyt metatiedot kuvaavat kuvatiedoston syntyhistoriaa, joka osaltaan todentaa myös prosessissa syntyneen digitaalisen ilmentymän autenttisuutta. Kuvatiedostojen pakolliset tekniset metatiedot PITÄÄ esittää MIX-metatietoskeeman version 2.0 mukaisesti.⁸

⁸ The Library of Congress » Standards » MIX. Kongressin kirjaston verkkosivu <http://www.loc.gov/standards/mix/> JA ANSI/NISO Z39.87-2006 (R2017) Data Dictionary - Technical Metadata for Digital Still Images <https://www.niso.org/publications/ansiniso-z3987-2006-r2017-data-dictionary-technical-metadata-digital-still-images> Viitattu 2.9.2020

Alla olevassa taulukossa (Taulukko 12) ilmaistaan vasemmalta oikealle MIX-kentän nimi, kentän tarkoitus vapaasti käännettynä ja velvoite. Velvoite- kentässä ilmaistaan kyseisen kentän ja sen skeeman mukaisen tiedon pakollisuus seuraavalla tavalla:

- P = pakollinen – tämä tieto PITÄÄ kuvata
- V = Vapaaehtoinen – tämä tieto PITÄISI kuvata, mutta se ei ole pakollista

MIX-metatietoskeemassa on kahdenlaisia kenttiä: säiliöitä ja dataelementtejä. Dataelementit sisältävät tietyn arvon, kun taas säiliöt sisältävät yhden tai useamman dataelementin ja ne voivat sisältää toisia säiliöitä dataelementteineen. Taulukossa 12 ilmaistaan vain tietyn arvon sisältäviä kenttiä eli dataelementtejä.

Taulukko 12, Kuvatiedostoa ja sen prosessointia kuvaavat metatiedot (taulukossa on ilmaistu vain tietoa sisältävät kentät, jotka PITÄÄ esittää MIX-metatietoskeeman version 2.0 mukaisessa rakenteessa)

MIX -kentän nimi	Kentän tarkoitus	Velvoite
objectIdentifierType	Dataelementti, joka määrittää järjestelmän tai verkkotunnuksen, jossa digitaalisen asiakirjan yksilöivä ID on uniikki.	P
objectIdentifierValue	Digitaalisen tiedoston identifioiva merkkisarja.	P
fileSize	Tiedoston koko tavuissa, esimerkiksi 72839.	P
formatName	Tiedoston formaatti. Vaadittu arvo image/jpeg tai image/tiff	P
formatVersion	Tiedoston versio, esimerkiksi 1.01 TAI 6.0	V
byteOrder	Dataelementti, joka määrittää tavujen tallennusjärjestyksen. Vaadittu arvo on joko big endian tai little endian.	P
compressionScheme	Käytetty pakkaus. Vaadittu arvo JPEG (kun tallekappale on jpeg-formaatissa) tai LZW (kun tallekappale on TIFF-formaatissa).	P
compressionRatio	Dataelementti, joka kertoo käytetyn pakkauksen tason. Ilmaistaan käyttäen numerator "90" ja denominator "100".	P (vain JPEG)
messageDigestAlgorithm	Dataelementti, joka identifioi algoritmin, jolla messageDigest-kentän arvo on luotu. Kentän arvo on jokin seuraavista: MD5, SHA-1, SHA256, SHA384, SHA512.	P
messageDigest	messageDigestAlgorithm kentän määrittämän algoritmin tuottama merkki sarja, esimerkiksi e8064dc0.	P
imageWidth	Kuvan leveys pikseleissä, esimerkiksi 1330.	P
imageHeight	Kuvan korkeus pikseleissä, esimerkiksi 1600.	P
colorSpace	Dataelementti, joka määrittää kuvan väriavaruuden. Vaadittu arvo RGB tai WhitesZero tai BlacksZero.	P
iccProfileName	Dataelementti, joka määrittää yleisesti käytetyn ICC-profiilin nimen. Vaadittu arvo sRGB, eciRGB v2, ProPhoto RGB, AdobeRGB (1998) tai Gray gamma 2,2.	P
iccProfileVersion	Dataelementti, joka kertoo käytetyn ICC-profiilin version, esimerkiksi v4 [eli sRGB v4].	P
iccProfileURL	Dataelementti, johon tallennetaan ICC-profiilin URL/URN, mikäli ICC-profiili ei ole hyvin dokumentoitu.	V
dateTimeCreated	Dataelementti, joka kertoo kuvatiedoston luontiajan. Ilmaistaan muodossa: YYYY-MM-DDTHH:MM:SS.	P

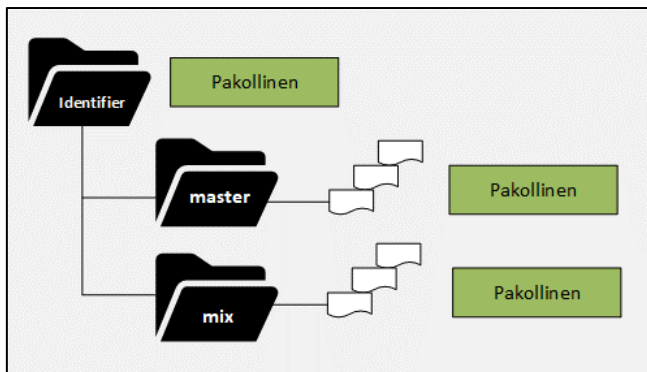
MIX -kentän nimi	Kentän tarkoitus	Velvoite
imageProducer	Dataelementti, joka identifioi digitaalisen tiedoston luoneen organisaation.	P
scannerManufacturer	Dataelementti, joka kertoo skannauksessa käytetyn laitteen valmistajan nimen.	P
scannerModelName	Dataelementti, joka kertoo käytetyn digitointilaitteen mallin nimen.	P
scannerModelNumber	Dataelementti, joka tarkoittaa digitointilaitteen mallin nimeä sen tyyppinumerolla.	P
scannerModelSerialNo	Digitointilaitteen sarjanumero, jonka avulla tietty laite on mahdollista yksilöidä.	P
scanningSoftwareName	Käytetyn skannausohjelmiston nimi.	P
scanningSoftwareVersionNo	Käytetyn skannausohjelmiston version numero.	P
orientation	Dataelementti, joka kertoo kuvan lukusuunnan.	P
samplingFrequencyUnit	Dataelementti, joka kertoo mittayksikön, jota käytetään tulkittaessa X ja Y resoluutiota. Vaadittu arvo "in."	P
xSamplingFrequency	Pikselien määrä resoluution mittayksikkö kohden leveysuunnassa. Vaadittu arvo on aineistotyyppiin riippuvainen.	P
ySamplingFrequency	Pikselien määrä resoluution mittayksikkö kohden pystysuunnassa. Vaadittu arvo on aineistotyyppiin riippuvainen.	P
bitsPerSampleValue	Dataelementti, joka määrittelee jokaisessa kanavassa olevien bittien määrän. Vaadittu arvo 8.	P
bitsPerSampleUnit	Dataelementti, joka määrittää bittien tulkintatavan. Arvo on joko integer tai floating point.	P
samplesPerPixel	Dataelementti, joka määrittää kanava-arvojen määrän pikselissä. Vaadittu arvo joko 3 tai 1.	P
targetType	Dataelementti, joka kertoo onko skannauksen laatua todentava mittataulu osa kuvaa vai skannattu erilliseen kuvaan.	V
targetManufacturer	Dataelementti, johon merkitään mittataulun valmistaja.	V
targetName	Dataelementti, joka identifioi käytetyn mittataulun nimen.	V
targetNo	Dataelementti, joka sisältää käytetyn mittataulun sarjanumeron.	V
externalTarget	Dataelementti, joka kertoo mistä TargetID-säiliön yksilöidyn mittataulun digitaalinen kuva löytyy.	V
performanceData	Dataelementti, joka kertoo mistä TargetID-säiliön yksilöidyn mittataulun mittausdata löytyy.	V

7 Digitointiprosessissa muodostettava siirtopaketti

Luvussa 6 ja sen alaluvuissa mainitut digitointiprosessissa tuotetut digitaaliset tiedostot PITÄÄ tallentaa alla olevassa kuvassa (Kuva 3) esitettyyn hakemistorakenteeseen, jotta ne voidaan ottaa vastaan Kansallisarkistoon. Digitaalinen ilmentymä PITÄÄ tuottaa hakemistorakenteeseen riippumatta siitä, milloin se siirretään Kansallisarkistoon. Aineistoja luovutettaessa siirtopaketti EI SAA sisältää mitään muuta kuin kuvassa 3 esitettyjä hakemistoja.

Tässä määritellyn hakemistorakenteen lisäksi organisaatio SAA tallentaa esimerkiksi käyttökappaleet omiin tietojärjestelmiinsä siinä tietorakenteessa, mitä kyseinen järjestelmä

edellyttää. Tässä asiakirjassa määritelty rakenne ei siis sulje pois muiden mahdollisten tallennusrakenteiden käyttöä.



Kuva 3, Digitointiprosessin vaadittu siirtopakettirakenne

Alla olevassa taulukossa (Taulukko 13) kuvaillaan, miten digitaaliset tiedostot PITÄÄ nimetä esitetyn hakemistorakenteen sisällä. Prosessissa tuotettujen digitaalisten tiedostojen PITÄÄ kohdata keskenään. Toisin sanoen MIX-metatietoskeeman mukaisen 0001.xml-tiedoston PITÄÄ sisältää kuvatiedostoa 0001.jpg tai 0001.tif kuvailevia metatietoja.

Taulukko 13, Siirtopaketin hakemistojen sisältö

Hakemisto	Selite
identifier	Tarkoittaa digitaalisen ilmentymän yksilöivää tunnusta, jonka avulla PITÄÄ pystyä tunnistamaan, mistä asiakirjakokonaisuudesta on kyse (esimerkiksi arkistoyksikkö). ⁹ Hakemisto sisältää ”digitaalisten tiedostojen hakemistot”.
master	Hakemistoon PITÄÄ tallentaa taulukoissa (Taulukko 10, Kuvatiedoston (TIFF) pakolliset metatiedot tai Taulukko 11, Kuvatiedoston (JPEG) pakolliset metatiedot) esitetyt kuvatiedostot yksittäisinä tiedostoina. Tiedostot PITÄÄ nimetä nelinumeroisina alkaen 0001.tif tai 0001.jpg.
mix	Hakemistoon PITÄÄ tallentaa Taulukko 12 esitetyt pakolliset tiedot koskien kaikkia master – hakemiston sisällä olevia kuvatiedostoja. Tiedostoon SAA tallentaa myös muita MIX-metatietoskeeman mukaisia tietoja, skeeman mukaisessa rakenteessa. Tiedostot PITÄÄ nimetä nelinumeroisina alkaen 0001.xml.

Mikäli aineisto toimitetaan Kansallisarkistolle, PITÄÄ jokainen siirtopaketti paketoita TAR-paketiksi. TAR-paketin sisältöä EI SAA tässä vaiheessa pakata. TAR-paketille PITÄÄ laskea tarkistesumma MD5-muodossa ja se PITÄÄ toimittaa siirron yhteydessä. Kansallisarkistoon aineistoa toimitettaessa tarkoitetaan identifier-hakemistolla AHAA-järjestelmän aineiston tunnistetta Ai01 (roolissa ahaa tekninen).

8 Esimerkkipaketit

Esimerkkejä siirtopaketeista on kaksi:

1. EsimerkkipakettiJPEG2020.tar

⁹ Digitoitavaksi päätetyn analogisen ilmentymän pitäisi olla kuvailtuna (kuvaileva metatieto tuotettu) ennen sen digitointia. Identifierin avulla PITÄÄ pystyä yhdistämään digitointiprosessissa syntyneet digitaaliset ilmentymät edellä mainittuun kuvailevaan metatietoon.

2. EsimerkkipakettiTIF2020.tar

Master-hakemistojen sisältämät kuvatiedostot eivät ole kuvanlaadullisia referenssejä. Kuvatiedostot sisältävät tässä asiakirjassa pakollisiksi määritellyt metatiedot. Mix –hakemistojen tiedostot ovat esimerkkejä liitteen paketissa olevista master-tiedostoista lukuun ottamatta mahdollisia elementtejä, joiden kohdalla toisin todetaan.

9 Allekirjoitukset

Pääjohtaja

Jussi Nuorteva

Tutkimusjohtaja

Päivi Happonen