Arkistoitavien asiakirjojen digitointia koskevat vaatimukset ja ohjeet

Kansallisarkisto

Ohje
19.12.2024
KA/20919/30.01.00/2024

KANSALLISARKISTO

KA/20919/30.01.00/2024

1 Sisältö ja tarkoitus 2
1.1 Rajaukset 2
1.2 Keskeiset muutokset ja viittaukset muihin asiakirjoihin 3
2 Termit ja käsitteet 6
3 Johdanto 12
3.1 Vaatimusten soveltaminen eri aineistotyyppien kohdalla 13
3.2 Digitoinnin yleiset hyvät käytänteet ja laadunvarmistus 15
4 Pääsääntöisesti A4- tai foliokokoisten aineistojen ja erikoisaineistojen vaatimukset 18
4.1 Vaatimusten mukainen digitointiprosessi ja digitointitulos 18
4.1.1 Kuvatiedostoa ja sen ominaisuuksia koskevat vaatimukset 21
4.1.2 Kuvatiedoston prosessointia kuvaavat MIX-metatiedot ja niiden rakenne 27
4.2 Aineistotyyppikohtaiset vaatimukset 30
4.2.1 Pääsääntöisesti A4- ja foliokokoiset asiakirjat ja sidokset 30
4.2.2 Kartat ja piirustukset 31
4.2.3 Asiakirjat ja sidokset, joiden tietosisältö esitetään aukeamittain 31
4.2.4 Valokuvat 33
4.2.5 Lasinegatiivit, muut negatiivit ja diat 34
4.2.6 Mikrofilmit ja -kortit 35
4.2.7 Röntgenfilmit 36
5 Audiovisuaalisten aineistojen vaatimukset 38
5.1 Ääninauhat 38
5.2 Kuvanauhat 40
6 Digitointiprosessissa muodostettava hakemistorakenne 42
6.1 Kuva-aineiston hakemistorakenne 42
6.2 Audiovisuaalisen aineiston hakemistorakenne 43
6.3 Esimerkkipaketit 45
7 Muutokset 46

KANSALLISARKISTO

KA/20919/30.01.00/2024

1 Sisältö ja tarkoitus

Ohje sisältää arkistoitavien eli arkistolain (831/1994) mukaan pysyvästi säilytettäväksi määrättyjen asiakirjojen digitointia koskevat Kansallisarkiston vaatimukset.

Ohje on tarkoitettu kaikille niille toimijoille, jotka digitoivat arkistolain (831/1994) nojalla arkistoitavaksi määrättyjä asiakirjoja. Vaatimusten mukaisesti digitoidusta asiakirjoista on mahdollista muodostaa Kansallisarkistoon siirrettävä siirtopaketti.

Vaatimusten mukainen digitointiprosessi ja digitointitulos mahdollistavat analogisten asiakirjojen hävittämisen digitoinnin jälkeen, mikäli asiakirjojen analogista säilytysmuotoa ei ole arvioitu kulttuurihistoriallisesti arvokkaaksi. Ohje ei ota kantaa asiakirjojen analogisten ilmentymien hävittämisprosessiin, vaan siinä kuvataan kriteerit, joiden täytyttyä analogisten asiakirjojen hävittäminen on mahdollista. Digitoinnin laadunvarmistusta käsitellään ohjeessa yleisellä tasolla ja suuntaa antavasti. On suositeltavaa, että toimijat määrittelevät monipuolisen laadunvarmistusprosessin lisäksi erillisen validointiprosessin ja toimintatavat digitaalisten ilmentymien todistusvoimaisuuden arviointia varten.

1.1 Rajaukset

Ohjetta sovelletaan, kun toimijalle jo muodostuneita arkistollisia kokonaisuuksia digitoidaan takautuvasti. Vaatimukset eivät koske digitointia, joka toteutetaan asiakirjan laatimisen tai asiankäsittelyn yhteydessä tai kun organisaatiolle saapuvia analogisia asiakirjoja muuten muutetaan digitaaliseen muotoon. Linjaus noudattaa Kansallisarkiston aiempien digitointimääritysten rajausta[1].

Ohje ei ota kantaa siihen, toteutetaanko digitointi toimijan oman organisaation työnä vai hankintaanko se ostopalveluna. Ohjetta sovelletaan kummassakin tilanteessa, kun kyse on arkistoitavista asiakirjoista. Ohje ei koske määräajan säilytettäviä asiakirjoja, joiden käsittelystä ohjeistaa julkisen hallinnon tiedonhallinnasta annetun lain (906/2019) perusteella Julkisen hallinnon tiedonhallintalautakunta.

[1] Kansallisarkiston vaatimukset hävittämiseen tähtäävään digitointiin (KA/3357/07.01.02.04.02/2019), Kansallisarkiston määritykset arkistoitavien erikoisaineistojen digitointiin (KA/15780/07.01.02.04.02/2020).

KANSALLISARKISTO

KA/20919/30.01.00/2024

1.2 Keskeiset muutokset ja viittaukset muihin asiakirjoihin

Tämä ohje (20919/30.01.00/2024) korvaa ja kumoaa Kansallisarkiston aiemmat vaatimukset hävittämiseen tähtäävään digitointiin ja erikoisaineistojen digitointiin.[2] Ne on tässä asiakirjassa yhdistetty yhdeksi ohjeeksi.

Ohjeeseen on lisätty audiovisuaalisten asiakirjojen digitointivaatimukset. Ohje on lisäksi ajantasaistettu ja yhdenmukaistettu Kansallisarkiston digitointia koskevien keskeisten määräysten ja päätösten kanssa.

Päivittämisen yhteydessä on parannettu vaatimusten ymmärrettävyyttä ja luettavuutta. Digitointia koskevat tekniset vaatimukset ovat säilyneet pääosin ennallaan. Muutokset suhteessa kumottuihin ohjeisiin ja vaatimuksiin on dokumentoitu luvussa 7. Toimijoiden on hyvä varmistaa, että sen käyttämät vaatimukset ovat tämän ohjeen mukaisia

Ohjeen suhdetta muihin Kansallisarkiston antamiin keskeisiin määräyksiin, seulontapäätöksiin ja suosituksiin on avattu taulukossa 1.

Säädökset, joihin ohjeen antaminen perustuu

Laki Kansallisarkistosta (1145/2016, 2 §)

Arkistolaki (831/1994, 11 §)

Voimassaoloaika

Toistaiseksi, ohjeen päiväyksestä eteenpäin

Tämä ohje korvaa ja kumoaa ohjeet

- Kansallisarkiston vaatimukset hävittämiseen tähtäävään digitointiin 12.4.2019 (KA/3357/07.01.02.04.02/2019)
- Kansallisarkiston määritykset arkistoitavien erikoisaineistojen digitointiin 19.1.2021 (KA/15780/07.01.02.04.02/2020)

[2] Kansallisarkiston vaatimukset hävittämiseen tähtäävään digitointiin (KA/3357/07.01.02.04.02/2019), Kansallisarkiston määritykset arkistoitavien erikoisaineistojen digitointiin (KA/15780/07.01.02.04.02/2020).

KANSALLISARKISTO

KA/20919/30.01.00/2024

Taulukko 1. Viittaukset muihin asiakirjoihin

ASIAKIRJA	SISÄLTÖ
Arkistolaitoksen vaatimukset digitoidulle aineistolle 24.8.2016 (AL/11130/07.01.02.04.02/2008) Kumottu 19.1.2021 ohjeella Kansallisarkiston määritykset arkistoitavien erikoisaineistojen digitointiin (KA/15780/07.01.02/04.02/2020)	Arkistolaitoksen tekniset vaatimukset digitoidulle aineistolle.
Kansallisarkiston vaatimukset hävittämisen tähtäävään digitointiin 12.4.2019 (KA/3357/07.01.02.04.02/2019) Kumotaan nyt annettavalla ohjeella (KA/20919/30.01.00/2024).	Kansallisarkiston vaatimukset hävittämiseen tähtäävään digitointiin.
Kansallisarkiston määritykset arkistoitavien erikoisaineistojen digitointiin 19.1.2021 (KA/15780/07.01.02.04.02/2020) Kumotaan nyt annettavalla ohjeella (KA/20919/30.01.00/2024).	Kansallisarkiston määritykset arkistoitavien erikoisaineistojen digitointiin. Kumosi ohjeen Arkistolaitoksen vaatimukset digitoidulle aineistolle 24.8.2016 (AL/11130/07.01.02.04.02/2008
Määräys valtion viranomaisten arkistoitavien asiakirjojen siirtämisestä Kansallisarkistoon 18.5.2021 (KA/26492/03.099/2020)	Määräyksessä määrätään Kansallisarkistoon siirrettävien arkistoitavien asiakirjojen muodosta ja siirtämisestä valtionhallinnon osalta. Määräyksen mukaan arkistoitavat asiakirjat siirretään Kansallisarkistoon digitaalisessa muodossa, ellei Kansallisarkisto ole toisin määrännyt.
Määräys arkistoitavien asiakirjojen muodosta 22.12.2021 (KA/15906/07.01.01.00/2021)	Määräyksen mukaan julkishallinnon toimijoille (pl. tuomioistuimet ja muut lainkäyttöelimet) 1.1.2022 alkaen muodostuville asiakirjoille ei erikseen haeta sähköistä arkistointia mahdollistavaa seulontapäätöstä, vaan se mahdollistetaan kyseisessä määräyksessä 1.1.2022.
Julkishallinnon pysyvästi säilytettävien digitoitujen ja digitoitavien asiakirjojen säilytysmuoto ja analogisen säilytysmuodon kulttuurihistoriallisen arvon arviointitietojen pysyvä säilytys 8.12.2023 (KA/15927/28.02.01/2023)	Päätöksessä määrätään ennen 1.1.2022 muodostuneet pysyvästi säilytettävät asiakirjat arkistoitaviksi yksinomaan digitaalisessa muodossa (lukuun ottamatta analogiselta säilytysmuodoltaan kulttuurihistoriallisesti arvokkaita asiakirjoja). Päätös mahdollistaa ennen 1.1.2022 muodostuneiden arkistoitavien analogisten asiakirjojen hävittämisen sil-

	loin, kun digitointi on jo tehty ennen 1.1.2022 Kansallisarkiston ajantasaisten digitointivaatimusten mukaisesti tai arkistolaitoksen digitointivaatimusten 24.8.2016 (AL/11130/07.01.02.04.02/2008) perusteella (arkistolaitoksen 24.6.2016 digitointivaatimukset on sittemmin kumottu, eikä vaatimuksia voi enää soveltaa 1.1.2022 jälkeen tehtyyn tai tehtävään digitointiin). Lisäksi päätöksessä määrätään kulttuurihistoriallisen arvon kriteereistä. Toimijat vastaavat 1.1.2024 alkaen kulttuurihistoriallisen arvon arvioinnista noudattaen päätöksen kriteereitä ja dokumentoivat arviointitiedot. Arkistoitavat analogiset asiakirjat digitoidaan päätöksessä esitettyjen reunaehtojen mukaisesti, mikä mahdollistaa niiden korvaamisen digitaalisilla ilmentymillä.
Arkistoitavien tietoaineistojen perusmetatiedot -suositus 20.02.2024 (KA/20011/30.00.00/2023)	Suositus koskee julkisen hallinnon arkistoitavien tietoaineistojen arkistometatietoa. Arkistometatiedolla viitataan arkistoinnin edellyttämään metatietokokonaisuuteen. Suosituksen perusmetatiedot on määritelty osoittamaan tämän metatiedon vähimmäistasoa. Organisaation on suunniteltava arkistometatiedon toteutustapa ja hyödyntäminen suhteessa tiedonhallinnassa ja tiedon elinkaaren hallinnassa vaadittavien muiden toimintatapojen, metatietojen, tietojen ja dokumentoinnin kokonaisuuteen. Kun tietoaineistoja siirretään Kansallisarkistoon, noudatetaan Kansallisarkiston voimassa olevia siirto-ohjeita. Siirto-ohjeiden metatietovaatimukset ovat yhteensopivat suosituksessa määritettyjen arkistoinnin perusmetatietojen kanssa.

2 Termit ja käsitteet

Ohjeessa on käytetty vaatimusten merkitsemiseen Internet Engineering Task Forcen määrittelemää termistöä[3]. Termien suomenkieliset vastineet on esitetty taulukossa 2.

Taulukko 2. Termit ja niiden käännökset

Englanti	Suomi
Must	**Pitää**
Must not	**Ei saa**
Required	**Pakollinen**
Should	**Pitäisi**
Should not	**Ei pitäisi**
May	**Saa**
Optional	**Vapaaehtoinen**

Asiakirjassa käytetyt käsitteet selitteineen on koottu taulukkoon 3.

Taulukko 3. Käsitteistö

Käsite	Selite
Analoginen asiakirja	Paperisessa tai muussa käsin kosketeltavassa muodossa laadittu, säilytetty ja käytettävä asiakirja (vrt. digitaalinen tieto tai asiakirja, joka muodostuu bittijonoista). Ks. **Digitaalinen asiakirja.**
Analogisen asiakirjan säilytysmuodon kulttuurihistoriallinen arvo	Arkistoidulle tai arkistoitavalle analogiselle asiakirjalle annettu arvo, joka on sidoksissa asiakirjan aineettomiin tai aineellisiin ominaisuuksiin. Kulttuurihistoriallisesti arvokkaaksi arvioitu asiakirja on sellaisenaan korvaamaton ja pitää siksi säilyttää ja arkistoida myös digitoinnin jälkeen sekä analogisessa että digitaalisessa muodossa. Asiakirjan analogista ilmentymää ei siis saa hävittää. Ks. **Digitaalinen asiakirja, Digitointi, Hävittäminen, Ilmentymä.**
Arkistointi	Arkistointi tarkoittaa osaksi asiakirjallista kulttuuriperintöä määriteltyjen asiakirjojen ja tietoaineistojen säilyttämistä sekä niiden säilyvyyden ja käytettävyyden turvaamista. Arkistoituja asiakirjoja ja tietoaineistoja säilytetään pysyvästi eli käytännössä ilman aikarajaa. Arkistointi perustuu Kansallisarkiston seulontapäätöksiin tai lainsäädäntöön. Arkistointi tapahtuu säilytysajan päättymisen jälkeen (Tiedonhallintalaki 906/2019 21 §). Arkistointi tarkoittaa myös arkistolain mukaista pysyvää säilyttämistä (8 § 3 momentti).

[3] RFC 2119, https://www.ietf.org/rfc/rfc2119.txt, viitattu 9.9.2024.

Käsite	Selite
Arkistoyksikkö	Sarja jakautuu pienempiin osiin eli arkistoyksiköihin. Analogisissa aineistoissa arkistoyksikön tulisi olla fyysisesti hallittava kokonaisuus Se voi olla yksittäinen asiakirja, asiakirjanippu, sidos, kartta jne. Ks. myös **Sarja.**
Asiakirja/asiakirjatieto	Asiakirja on todistusvoimaista tietoa. Toimija on tuottanut tai vastaanottanut tiedon osana tehtäviään ja säilyttää sitä tietovarantona sekä todisteena. Asiakirja on sisällön, rakenteen/ilmiasun ja kontekstin kokonaisuus. Metatiedot sisältyvät tai liittyvät asiakirjatietoihin. (SFS-ISO15489-1: Tieto ja dokumentointi. Asiakirjahallinto. Osa 1. Käsitteet ja periaatteet). Lainsäädännössä asiakirjalla tarkoitetaan kirjallisen ja kuvallisen esityksen lisäksi sellaista käyttönsä vuoksi yhteen kuuluviksi tarkoitetuista merkeistä muodostuvaa tiettyä kohdetta tai asiaa koskevaa viestiä, joka on saatavissa selville vain automaattisen tietojenkäsittelyn tai äänen- ja kuvantoistolaitteiden taikka muiden apuvälineiden avulla. (Julkisuuslaki 621/1999, 5 §).
Audiovisuaaliset asiakirjat/aineistot	Näillä tarkoitetaan tässä ohjeessa magneettinauhatekniikkaan pohjautuvia aineistoja, joiden analoginen käyttäminen edellyttää kyseiselle nauhatyypille (kuva tai ääni) soveltuvia laitteita. Niiden käyttö on analogisina rajoitettua myös materiaalien fyysisen haurauden vuoksi. Ne ovat ns. itsetuhoutuvia aineistoja, joiden elinkaari on huomattavasti lyhyempi kuin arkistokelpoisille materiaaleille tuotettujen asiakirjojen. Ks. **Analoginen asiakirja**, **Digitaalinen asiakirja.**
Digitaalinen asiakirja	Tässä ohjeessa digitaalisella asiakirjalla tarkoitetaan Kansallisarkiston digitointiohjeiden mukaan analogisesta asiakirjasta tuotettua samansisältöistä digitaalista asiakirjaa, eli analogisen asiakirjan digitaalista ilmentymää. Digitaalinen asiakirja voi koostua useista tiedostoista. Ks. **Analoginen asiakirja, Ilmentymä, Tiedosto.**
Digitointi	Tässä ohjeessa digitoinnilla tarkoitetaan analogisessa muodossa säilytettävien tai arkistoitujen asiakirjojen tietosisällön muuntamista digitaaliseen muotoon. Digitointi on prosessi eli joukko toimenpiteitä, joita tehdään manuaalisesti, laittein sekä tietojärjestelmäympäristössä. Esi-

Käsite	Selite
	merkiksi skannaaminen on yksi digitointiprosessin toimenpiteistä. Digitointiprosessissa tuotetaan digitoinnin tuloksena tallekappaletiedosto.[4] Ks. **Analoginen asiakirja**. **Ilmentymä, Digitointilaite, Skannaus, Tallekappale.**
Digitointilaite	Laite, jolla analoginen asiakirja muunnetaan digitaaliseen muotoon. Ks. **Digitointi.**
Erikoisaineisto	Tässä ohjeessa erikoisaineistoilla tarkoitetaan karttoja, piirustuksia, valokuvia, negatiiveja, mikrofilmejä, mikrokortteja, röntgenfilmejä sekä asiakirjoja ja sidoksia, joissa sidosten tietosisältö esitetään aukeamittain. Erikoisaineistojen digitointi vaatii esimerkiksi niiden koon, sisällön, materiaalien tai muiden ominaisuuksien vuoksi erilaista esivalmistelua tai erilaisia digitointilaitteita ja -määrityksiä kuin määrämuotoisten A4- tai foliokokoisten aineistojen digitointi. Erikoisaineistot on usein arkistoitu omina kokonaisuuksinaan tai kokoelminaan. Erikoisaineistoksi ei tässä ohjeessa luokitella yksittäisiä asiakirjoja, jotka on arkistoitu pääasiassa A4- tai foliokokoisten asiakirjojen joukkoon.
Fotometrinen tulkinta	Fotometrinen tulkinta (engl. photometric interpretation) on exif-tunniste, joka määrittää kuvan pikselien esitystavan, eli sen, miten pikseliarvoja tulkitaan kuvassa (esim. RGB tai BlackIsZero). Tunniste auttaa ohjelmistoja ymmärtämään kuvan väritilan ja kirkkausdatan rakenteen. **Huom.** Tässä ohjeessa suomenkielinen termi on käytännössä suora käännös exif-tunnisteen englanninkielisestä nimestä.
Hakemistorakenne	Tässä ohjeessa digitoinnissa muodostettujen tiedostojen järjestämiseen ja tallentamiseen käytetty rakenne. Kun tämän ohjeen mukaan digitoituja asiakirjoja siirretään Kansallisarkistoon, hakemistorakenteesta pitää pystyä muodostamaan Kansallisarkiston tietojärjestelmiin siirrettävä siirtopaketti. Ks. **Siirtopaketti**.
Hävittäminen	Määräajan säilytettävän asiakirjan fyysinen (ml. tekninen) tuhoaminen. Arkistoitavan asiakirjan analogisen ilmentymän tuhoaminen. Arkistolaki (831/1994, 7§) käyttää käsitettä hävittäminen: "Arkistotoimen tehtävänä on varmistaa asiakirjojen käytettävyys ja säilyminen,

[4] Yleisemmän määritelmän mukaan digitointi on prosessi, jossa analogisessa muodossa oleva objekti muunnetaan digitaaliseen muotoon. Teknisemmin ilmaistuna kyse on analogisen signaalin tallentamisesta digitaaliseen muotoon: digitoitavasta objektista (valosta) peräisin olevat analogiset signaalitiedot muunnetaan digitaalisesti koodattuun muotoon.

Käsite	Selite
	huolehtia asiakirjoihin liittyvästä tietopalvelusta, määritellä asiakirjojen säilytysarvo ja hävittää tarpeeton aineisto." Tiedonhallintalaki (906/2019, 21 §) käyttää käsitettä tuhoaminen: "Säilytysajan päättymisen jälkeen tietoaineistot on arkistoitava tai tuhottava viipymättä tietoturvallisella tavalla.". Ks. **Ilmentymä, Tuhoaminen.**
ICC-väriprofiilli	Ks. Väriprofiili
Ilmentymä	Asiakirjan digitaalinen tai analoginen ilmenemismuoto. Asiakirjalla voi olla elinkaarensa aikana useita ilmentymiä. Perinteinen esimerkki eri ilmentymien välisestä suhteesta on aineiston analoginen ilmentymä ja siitä digitoimalla tuotettu digitaalinen ilmentymä. Ks. **Analoginen asiakirja, Digitaalinen asiakirja, Digitointi.**
Jakelukappale[5]	Ks. **Käyttökappale.**
Kuva-ala	Kuvatiedostosta silmämääräisellä tarkastelulla näkyvä informaatio.
Kuvatiedosto	Digitointiprosessissa tuotettu bittikarttakuva. Ks. **Tallekappale**.
Käyttökappale[6]	Käyttökappaleet ovat tallekappaleesta erilaisia käyttötarkoituksia ja käyttöliittymiä varten luotoja tiedostoja. Käytetään myös käsitettä jakelukappale. Ks. **Tallekappale**.
Mittataulu	Apuväline, jolla voidaan mitata kuvanlaatua. Mittataulu on yleensä valokuvapaperille tulostettu arkki, jonka raja-arvot ovat tiedossa. Sitä voidaan käyttää apuna sekä silmämääräisessä että koneellisessa laadun arvioinnissa. Mittataulu voi olla esimerkiksi Universal Test Target (UTT) -mittataulu tai IT8-mittataulu. Niiden avulla kuvasta voidaan mitata sen todellista resoluutiota, värejä, vääristymiä, sävyntoistoa jne. Puhekielessä mittatauluista käytetään usein termiä targetti.
Nauhakokonaisuus	Ääninauhan yksi puoli kokonaisuudessaan.
Referenssiarvo	Referenssiarvoilla viitataan muun muassa niihin arvoihin, joihin mittaustulosta verrataan. Usein myös mittataulut voivat olla värimitattuja, jolloin saadaan mittataulukohtaiset arvot. Analysointiohjelmisto voi olla

[5] FADGI, käyttökappale: http://www.digitizationguidelines.gov/term.php?term=derivativefile Viitattu 17.10.2024.
[6] FADGI, käyttökappale: http://www.digitizationguidelines.gov/term.php?term=derivativefile Viitattu 17.10.2024.

Käsite	Selite
	automaattinen ohjelma, joka osaa lukea mittataulusta skannatun kuvan ja siitä eri mitta-alueet. Ks. **Mittataulu.**
Sarja	Arkistoinnissa aineistokokonaisuus voidaan jakaa useisiin pienempiin osiin, joita kutsutaan sarjoiksi. Sarja on muodoltaan, tehtävältään ja/tai sisällöltään yhteenkuuluvien aineistojen muodostama aineiston luokka tai osakokonaisuus. Sarja jakautuu pienempiin osiin eli arkistoyksiköihin. Ks. **Arkistoyksikkö.**
Seulontapäätös	Kansallisarkiston päätös arkistolain 8, 11 ja 14 a §:ien nojalla pysyvästi säilytettävistä (arkistoitavista) asiakirjoista ja niiden säilytysmuodosta.
Siirtopaketti	Kansallisarkisto ottaa digitoitua tietoaineistoa vastaan määrämuotoisissa siirtorakenteissa. Mikäli tässä asiakirjassa esitettyjen vaatimusten mukaan digitoitua tietoaineistoa siirretään Kansallisarkistoon, on se tallennettava siirtoa varten Kansallisarkiston siirto-ohjeiden mukaiseen hakemistorakenteeseen ja paketoitava yhdeksi tiedostoksi. Ks. myös **Hakemistorakenne.**
Skannaus	Skannaus on digitoinnin vaihe, joka toteutetaan joko skannereilla tai kameroilla kuvaamalla. Ks. **Digitointi.**
Tallekappale[7]	Teknisiltä ominaispiirteiltään laadukkain digitoinnissa tuotettu tiedosto. Tallekappale viedään yleensä pitkäaikaissäilytysjärjestelmiin. Ks. **Digitointi, Käyttökappale, Tiedosto.**
Tiedosto	Tiedosto muodostaa joko yksin tai muiden tiedostojen kanssa digitaalisen asiakirjan. Tiedostot voivat olla monentyyppisiä. Tässä ohjeessa esimerkiksi bittikarttakuva, XML-tiedosto tai video-/äänitiedosto. Ks. **Ilmentymä, Digitaalinen asiakirja, Digitointi, Tallekappale.**
Todistusvoimaisuus	Todistusvoimaisten asiakirjojen ominaispiirteet ovat autenttisuus, luotettavuus, eheys ja käytettävyys. Asiakirja on autenttinen eli alkuperäinen, kun sen voidaan todentaa olevan sitä, mitä sen pitää olla ja sen, laatimis-/lähettämisajankohta voidaan todentaa. Asiakirja on luotettava, kun, siihen voidaan uskoa kattavana ja virheettömän esityksenä liiketoimista/tehtävistä/tosiasioista, joista se on osoitus ja siihen voidaan luottaa myöhempien liiketoimien/tehtävien pohjana. Asiakirja on eheä, jos se on täydellinen ja säilynyt muuttumattomana. Käytettävyys tarkoittaa

[7] FADGI, tallekappale: http://www.digitizationguidelines.gov/term.php?term=productionmasterfile. Viitattu 17.10.2024.

Käsite	Selite
	sitä, että asiakirja kyetään paikallistamaan, hakemaan, esittämään ja tulkitsemaan. Lähde: SFS-ISO 15489-1:2017.
Tuhoaminen	Ks. **Hävittäminen.**
Tuotantovuorokausi	Vuorokausi, jonka aikana digitointilaitteella tuotetaan tiedostoja. Ks. **Digitointilaite, Tiedosto.**
Väriavaruus	Joukko värejä, joiden sijainti tietyssä värijärjestelmässä (esim. sRGB tai Adobe RGB) on määritelty. Väriavaruus tukee yhdessä väriprofiilin kanssa sitä, että värit pystytään toistamaan yhdenmukaisesti eri laitteilla ja ohjelmistoilla. Ks. myös **Väriprofiili.**
Väriprofiili	Väriprofiili määrittää, miten värit toistetaan tietyssä laitteessa tai ohjelmistossa. Se tulkitsee kuvan värintoisto-ominaisuudet ja linkittää ne tiettyyn väriavaruuteen. Väriprofiili varmistaa, että värit näkyvät johdonmukaisesti ja oikein riippumatta käyttöympäristöstä. Ks. **Väriavaruus.**
Väritila	Väritila on joukko värejä, jotka voidaan esittää tietyssä värijärjestelmässä (esim. RGB tai Grayscale). Se määrittää, miten värit järjestetään ja esitetään tiedostossa.

3 Johdanto

Kansallisarkiston digitointivaatimusten tavoite on varmistaa asiakirjalliseen kulttuuriperintöön kuuluvien arkistoitavien asiakirjojen tietosisällön säilyminen ja käytettävyys digitoinnin jälkeisessä digitaalisessa muodossa. Vaatimuksissa on huomioitu arkistosektorilla[8] yleisesti käytössä olevat standardit sekä muiden kansallisarkistojen laatuvaatimukset digitoinnille sekä pitkäaikaissäilytyspalveluiden (PAS-palvelut) kansalliset vaatimukset.

Ohjeen vaatimusten velvoittavuus perustuu Kansallisarkiston yleispäätökseen 8.12.2023 (KA/15927/28.02.01/2023) ja määräykseen 22.12.2021 (KA/15906/07.01.01.00/2021), joissa määrätään arkistoitavien asiakirjojen digitaaliseen muotoon muuttamisen vaatimuksista ja hävittämisestä. Ennen digitoinnin aloittamista tulisi aina varmistaa, että asiakirjoilla on Kansallisarkiston seulontapäätös, jossa on määrätty niiden arkistoinnista. Tällöin arkistoitavien asiakirjojen digitointia koskevat vaatimukset pystytään huomioimaan digitoinnin suunnittelussa ja toteutuksessa.

Vaatimusten mukainen digitointiprosessi ja digitointitulos mahdollistavat analogisten asiakirjojen hävittämisen digitoinnin jälkeen, mikäli asiakirjojen analogista säilytysmuotoa ei ole arvioitu kulttuurihistoriallisesti arvokkaaksi. Analogisten asiakirjojen tietosisällön arkistointi yksinomaan digitaalisessa muodossa edellyttää sitä, etteivät digitaaliseen muotoon muuttamisessa käytetyt menetelmät heikennä asiakirjan todistusvoimaisuutta. Hävittäminen on peruuttamaton toimenpide, jota ennen toimijan on varmistettava digitaalisten ilmentymien todistusvoimaisuus.

Tämän ohjeen vaatimuksia pitää noudattaa myös silloin, kun digitoidaan asiakirjoja, joiden analogisella säilytysmuodolla julkishallinnon toimija on arvioinut olevan kulttuurihistoriallista arvoa tai joille Kansallisarkisto on antanut seulontapäätöksen säilytysmuodon kulttuurihistoriallisesta arvosta.[9] Asiakirjat, joiden analogisella ilmentymällä on kulttuurihistoriallista arvoa, arkistoidaan digitoinnin jälkeen sekä analogisessa että digitaalisessa muodossa.

[8] Kansallisarkiston ohjeet koskevat arkistolain (831/1994) nojalla arkistoitavaksi määrättyjen analogisten asiakirjojen digitointia. Esimerkiksi kirjasto- ja museosektorin toimijoiden tuottamat digitointiohjeet voivat poiketa Kansallisarkiston ohjeistuksesta.

[9] Kulttuurihistoriallisen arvon arviointi kohdistuu vain niihin asiakirjoihin, jotka on määrätty arkistoitavaksi. Vuoden 2024 alusta lähtien julkishallinnon toimijat arvioivat analogisten asiakirjojen kulttuurihistoriallisen arvon ja arkistoitavan säilytysmuodon itsenäisesti noudattaen Kansallisarkiston antamaa säilytysmuotoa koskevaa yleispäätöstä (KA/15927/28.02.01/2023, 8.12.2023) ja siinä määriteltyjä arviointikriteereitä. Kansallisarkisto on antanut aiemmin seulontapäätöksiä analogisten asiakirjojen kulttuurihistoriallisesta arvosta ja säilytysmuodosta. Tästä toimintatavasta luovuttiin yleispäätöksen (KA/15927/28.02.01/2023, 8.12.2023) myötä.

3.1 Vaatimusten soveltaminen eri aineistotyyppien kohdalla

Tämä ohjeen vaatimukset on jaoteltu seuraavasti:

1. **Pääsääntöisesti A4- tai foliokokoiset asiakirjat ja sidokset**
   - A4- tai foliokoiset asiakirjat ovat kooltaan pääosin määrämuotoista aineistoa. Foliokoossa arkinkoko voi vaihdella ja on yleensä standardoitua A4-kokoa hieman suurempi.
2. **Erikoisaineistot**
   - Erikoisaineistot koostuvat kartoista, piirustuksista, valokuvista, negatiiveista, röntgenfilmeistä, mikrofilmeistä ja mikrokorteista sekä sidoksista ja asiakirjoista, joiden tietosisältö esitetään aukeamittain. Erikoisaineistot on usein arkistoitu omina kokonaisuuksinaan/kokoelminaan. Erikoisaineistoksi ei tässä ohjeessa luokitella yksittäisiä asiakirjoja, jotka on arkistoitu pääsääntöisesti A4- tai foliokokoisten asiakirjojen joukkoon.
3. **Audiovisuaaliset aineistot (AV-aineistot)**
   - Audiovisuaaliset aineistot jaetaan ohjeessa ääninauhoihin ja kuvanauhoihin

Asiakirjojen sijoittuminen eri aineistotyyppeihin on jäsennelty taulukkoon 4.

Aineistotyyppikohtaiset vaatimukset ja niiden tarkennukset on koottu lukuun 4 (Pääsääntöisesti A4- tai foliokoiset sekä sidokset ja erikoisaineistot) ja lukuun 5 (Audiovisuaaliset aineistot).

Kansallisarkiston digitointivaatimusten mukaiset tallekappaleiden tiedostomuotovaatimukset on esitetty aineistotyyppikohtaisten vaatimusten yhteydessä.[10] Muulla tavoin digitoidut tai muihin tiedostomuotoihin digitoidut asiakirjat[11] ovat arkistoinnin näkökulmasta toimijan omaan käyttöönsä tuottamia käyttökappaleita. Arkistoitavien asiakirjojen analogisia ilmentymiä ei tällöin voi digitoinnin jälkeen hävittää tai digitaaliseen muotoon muutettuja asiakirjoja siirtää Kansallisarkistoon.

Toimija voi tallentaa vaatimusten mukaisten tallekappaleiden lisäksi esimerkiksi PDF-muotoiset käyttökappaleet omiin järjestelmiinsä haluamassaan tietorakenteessa.

[10] Pääsääntöisesti A4- tai foliokokoiset asiakirjat ja sidokset sekä erikoisaineistot: aineistotyypin mukaan joko TIFF tai JPEG. AV-aineistot: FFV1 versio 3 [FFV1] fmt/569 (kuvanauhat) tai LPCM-koodattu WAV (ääninauhat)
[11] Ks. luku 1.1 Rajaukset soveltamisen osalta.

Taulukko 4. Vaatimusten soveltaminen eri aineistojen yhteydessä

Aineisto	Pääsääntöisesti A4- tai foliokokoiset ja sidokset	Erikoisaineistot	Audiovisuaaliset aineistot
A4- ja foliokokoiset asiakirjat	**X**		
A4- ja foliokokoisten asiakirjojen joukossa olevat yksittäiset valokuvat, piirustukset ja kartat	**X**		
A4-kokoisia asiakirjoja pienemmät asiakirjat (esim. kortistot)	**X**		
Asiakirjat, joiden tietosisältö esitetään aukeamittain		**X**	
Kartat		**X**	
Kuvanauhat (VHS-, H8- ja Betacam-videonauhat)			**X**
Lasinegatiivit, muut negatiivit ja diat		**X**	
Mikrofilmit ja -kortit		**X**	
Piirustukset		**X**	
Röntgenfilmit		**X**	
Sidokset	**X**		
Sidokset, joiden tietosisältö esitetään aukeamittain		**X**	
Valokuvat		**X**	

Aineisto	Pääsääntöisesti A4- tai foliokokoiset ja sidokset	Erikoisaineistot	Audiovisuaaliset aineistot
Ääninauhat (äänikelanauhat, C-kasetti ja minikasetti)			**X**

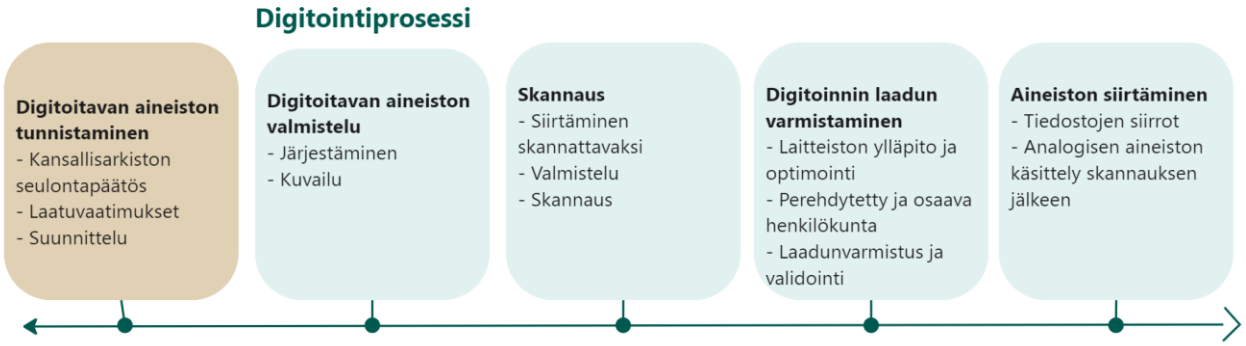
3.2 Digitoinnin yleiset hyvät käytänteet ja laadunvarmistus

Digitointi ymmärretään yleisesti prosessiksi. Esimerkiksi skannaus on yksi digitointiprosessin vaiheista. Tässä luvussa kuvataan digitointiprosessin ja digitoinnin laadunvarmistuksen yleisiä hyviä käytänteitä. Vaatimusten mukaista digitointia voidaan käytännössä tehdä monin eri tavoin ja menetelmin. Prosessia kuvataan luvussa siksi vain pääpiirteittäin, eikä siinä käytetä taulukon 2 termejä.

Ennen digitoinnin aloittamista ja skannausta pitää varmistaa, että asiakirjat on luetteloitu ja kuvailtu. Digitointiprosessissa muodostuvat digitaaliset ilmentymät pitää pystyä yhdistämään luettelointi- ja kuvailutietoihin. Myös analogisen ilmentymän käsittelyketju sekä digitointivalmistelua koskevat tiedot tulisi aina dokumentoida.

Digitointi toteutetaan käytännössä arkistoyksiköittäin.

Ennen asiakirjan skannaamista tai kuvaamista digitoitava aineisto pitää tunnistaa, valmistella ja siirtää digitoitavaksi. Varsinaisen digitoinnin aikana aineiston alkuperäistä järjestystä ei korjata. Digitoijan pitäisi voida olettaa, että aineisto on ennen digitointia järjestetty riittävällä tarkkuudella. Digitoinnissa havaitut puutteet tai epäjärjestys tulisi kuitenkin dokumentoida ja kuvata metatietoina. Skannaamisen jälkeen huolehditaan sekä digitoidun analogisen aineiston käsittelystä että digitoinnissa tuotettujen tiedostojen ja tietojen laadusta, tallentamisesta, säilymisestä sekä saattamisesta käyttöön.



Kuva 1. Digitaliseen muotoon muuntamisen yleisen prosessin vaiheita ovat muun muassa digitoitavan aineiston tunnistaminen, digitoitavan aineiston valmistelu, skannaus, digitoinnin laadun varmistaminen eri menetelmin ja aineiston siirtäminen. Edellä mainitut vaiheet voivat limittyä toisiinsa.

Audiovisuaalisista asiakirjoista esimerkiksi ääninauhat digitoidaan laadukkailla studiokäyttöön tarkoitetuilla laitteistoilla. Niiden osalta pitää ennen digitointia varmistaa muun muassa nauhojen eheys sekä niin sanotun kuivatuskäsittelyn tarve.[12]

Digitointilaitteet ja ohjelmistot muodostavat digitointiprosessin yhteydessä tiedostokohtaisia metatietoja, jotka dokumentoivat digitointiprosessia ja digitoinnissa muodostuneen tiedoston ominaisuuksia. Skannauksen lopputuloksena muodostunutta digitaalista tiedostoa eli tallekappaletta pitää usein vielä rikastaa erilaisin metodein. Skannausohjelmistot ovat yleensä kaupallisia, mutta digitoinnin jälkiprosessointi on esimerkiksi Kansallisarkistossa toteutettu hyödyntämällä myös avoimen lähdekoodin ohjelmia sekä omaa sovelluskehitystä.

Digitoinnin laadunvarmistusta ei voi kohdentaa digitointiprosessin tiettyihin kohtiin, vaan sitä on tehtävä prosessin kaikissa vaiheissa. Esimerkiksi skannausvaiheen laadunvarmistus voidaan jakaa vielä ennen skannausta tehtäviin toimenpiteisiin ja skannauksen jälkeiseen laadunvarmistukseen.

Skannaus voidaan tehdä joko kameralla kuvaamalla tai skannerilla skannaamalla. Luettavuuden takia tässä ohjeessa käytetään molemmista tavoista termiä skannaus. Skannausinfrastruktuurin tuottama kuvatiedosto ei saa olla epätarkka, vääristynyt tai muuten mittasuhteiltaan muuttunut, jotta tietosisällön tulkinta ei vaarannu. Kuvan laadun lisäksi laiteinfrastruktuurissa on varmistettava, että analogiset asiakirjat muunnetaan digitaaliseen muotoon tietosisällöltään täydellisinä.

Skannausprosesseissa pitää välttää esimerkiksi kahden tai useamman päällekkäisen analogisen asiakirjan menemistä laitteen läpi. Tämä on vaarana, kun käytetään läpisyöttöskannereita, avora-

[12] Kelanauhojen kuivatuskäsittelyä käytetään tahmaavien nauhojen käsittelyyn. Lämpökäsittelyn jälkeen nauha yleensä saadaan toistettua ilman toistopään tukkeentumista. Käsittely pitää aina toteuttaa ammattilaisilla. Englanniksi aiheesta löytyy tieto termillä ”Sticky Shed Syndrome”.

taskannereita tai muita skannausratkaisuja, joissa asiakirjoja skannataan ”massana”. Tämän tyyppisissä skannereissa pitää olla ns. tuplasyötön tunnistus (engl. multifeed detection), jonka toiminta on varmistettu. Yleisin tuplasyötön tunnistuksen tapa on ultraäänisensorit, jotka havaitsevat hyvin herkästi päällekkäisten asiakirjojen arkit.

Ennen skannauksen aloittamista optimoidaan skannausinfrastruktuurin suorituskyky. Tavoite on, että skannauksessa tuotetaan aina niin laadukas tiedosto kuin käytössä olevalla teknisellä kokoonpanolla on mahdollista tuottaa. Optimoinnin jälkeen infrastruktuurin suorituskykyä tulisi tarkkailla suunnitellusti, jotta prosessissa tuotettavien digitaalisten ilmentymien laatu säilyy tasaisena. Laadun tarkkailua varten tarvitaan yleensä mittataulu[13], mittataulun referenssiarvot ja analysointiohjelmisto.

Skannauksen jälkeisen validoinnin tavoitteena on varmistua siitä, että digitointi ja digitointitulos täyttävät niille määritellyt vaatimukset. Arkistoitavaksi määrättyjen asiakirjojen kohdalla kyse on tämän ohjeen vaatimuksista. Validointi voidaan toteuttaa otannoilla, joiden määrä ja koko riippuvat skannausprosessin luotettavuudesta ja laadunvalvonnan kokonaistoteutuksesta eli siitä, missä prosessin vaiheissa ja millä tavoilla valvontaa toteutetaan ja miten luotettavaa se kokonaisuutena arvioiden on. Laatupoikkeamat pyritään havaitsemaan silmämääräisellä tarkastelulla.

[13] Mittataululla tarkoitetaan esimerkiksi Universal Test Target (UTT) -mittataulua tai IT8-mittataulua, joiden avulla kuvasta voidaan mitata muun muassa sen todellista resoluutiota, värejä, vääristymiä, sävyntoistoa jne.

4 Pääsääntöisesti A4- tai foliokokoisten aineistojen ja erikoisaineistojen vaatimukset

Tämä luku sisältää digitointiprosessissa tuotettua kuvatiedostoa ja sen prosessointia koskevat vaatimukset silloin, kun digitoidaan pääsääntöisesti A4- tai foliokokoisia asiakirjoja sekä sidoksia tai erikoisaineistoja.

Aineistot on eritelty taulukossa 4 Vaatimusten soveltaminen eri aineistojen yhteydessä.

Luvussa 4.1 esitetään aineistotyyppien yhteiset vaatimukset. Aineistotyyppiriippuvaiset vaatimukset on koottu lukuun 4.2.

4.1 Vaatimusten mukainen digitointiprosessi ja digitointitulos

Digitointiprosessissa tuotetaan tallekappaletiedosto eli kuvatiedosto, joka on aineistotyyppikohtaisten vaatimusten mukaisesti joko häviöttömästi pakatussa TIFF-muodossa tai häviöllisesti pakatussa JPEG-muodossa.[14]

Digitointiprosessissa tuotettu tiedosto ja sen metatiedot dokumentoivat digitointiprosessia. Prosessi **pitää** dokumentoida tässä ohjeessa kuvatuilla tavoilla ja tarkkuudella. Dokumentointi tarkoittaa sitä, että skannauksesta ja kuvatiedostojen mahdollisesta käsittelystä tallennetaan näitä toimenpiteitä dokumentoivat metatiedot. Metatietona **saa** tallentaa myös muita prosessiin kuuluneita toimenpiteitä.[15]

Ennen skannausta digitoinnissa käytetyn infrastruktuurin suorituskyky **pitää** optimoida käyttämällä tarkoitukseen soveltuvia mittatauluja ja ohjelmistoja. Optimoinnin jälkeen suorituskyvyn tasalaatuisuutta **pitää** valvoa. Infrastruktuurin tuottamien kuvatiedostojen laatu **pitäisi** todentaa kerran tuotantovuorokaudessa käyttämällä tähän tarkoitukseen tarkoitettuja mittatauluja.

Esimerkki: Kevyimmillään todentaminen voi olla tiettyjen kuvanlaadullisten elementtien tarkastamista skannatusta mittataulun kuvasta silmämääräisesti. Edistyneemmässä prosessissa tarkastuksen voi hoitaa automaattinen kuvanlaatua analysoiva ohjelma, jolle on määritelty raja-arvot, joiden avulla ohjelma pystyy ilmoittamaan, läpäiseekö kuva testit.

[14] Tallekappale on teknisiltä ominaispiirteiltään laadukkain digitointiprosessissa tuotettu kuvatiedosto. Se viedään yleisesti pitkäaikaissäilytysjärjestelmiin. Vrt. tallekappaleesta muodostetut käyttökappaleet eli erilaisia käyttötarkoituksia varten muodostetut tiedostot.

[15] Digitointia ja muodostuvan kuvatiedoston ominaisuuksia dokumentoivia metatietoja muodostavat pääasiassa digitointiprosessin yhteydessä prosessissa käytettävät laitteet ja ohjelmistot.

Digitointiprosessissa **pitää** varmistua siitä, että digitoitavaksi tarkoitettu kokonaisuus digitoidaan:

- kokonaisuutena
- arkiston alkuperäisen järjestyksen mukaisesti
- sisällöllisesti täydellisenä.

Digitoitavaan kokonaisuuteen kuuluvat analogiset asiakirjat **pitää** digitoida siten, ettei mitään informaatiota[16] jää teknisen tai toiminnallisen virheen takia muuntamatta digitaaliseen muotoon. Jokaisesta digitoitavaan kokonaisuuteen liittyvästä yksittäisestä kuvatiedostosta **pitää** siis silmämääräisellä tarkastelulla olla saatavissa kaikki sama informaatio kuin sen analogisesta vastineesta.

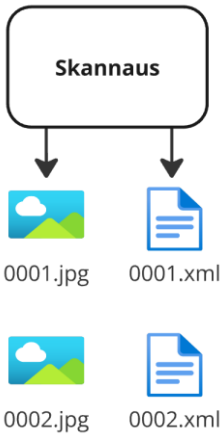
Monisivuisen asiakirjan jokainen sivu **pitäisi** digitoida molemmilta puolilta. Kummastakin puolesta muodostuu oma kuvatiedostonsa. Mikäli toimija on täysin varma, että digitoinnin kohteena olevan tietoaineistokokonaisuuden kääntöpuolet ovat kokonaisuudessaan tyhjiä (eivät sisällä informaatiota), ne voidaan jättää digitoimatta. Ratkaisu vaikuttaa digitointituloksen ja digitoitujen asiakirjojen ymmärrettävyyteen ja tulkintaan, joten se **pitää** dokumentoida tietoaineiston kuvailutietoihin.[17] Kansallisarkisto digitoi asiakirjat lähtökohtaisesti sekä etu- että kääntöpuolilta.

Kuvan metatiedot[18] **pitää** tallentaa kuvatiedostoon. Niistä sekä kuvatiedoston muodostumista ja prosessointia koskevista teknisistä metatiedoista **pitää** lisäksi tuottaa kutakin kuvatiedostoa vastaava tiedostokohtainen MIX-metatietoskeeman (versio 2.0) mukainen XML-tiedosto.

[16] Autenttisuuden, todistusvoimaisuuden ja eheyden edellytys on, että analoginen ilmentymä digitoidaan tietosisällöltään sellaisena kuin se on.

[17] Ks. Arkistoitavien tietoaineistojen perusmetatiedot -suositus, luku 3.3 Aineistojen ja ilmentymien elinkaaritapahtumat ja kohdat 4.1.4 Elinkaaritapahtumat ja 4.1.7 Alkuperäisyys ja eheys ja 4.4.6 Ilmentymän elinkaaritapahtumat

[18] Esimerkiksi EXIF-tiedot ja muut vastaavat tiedot.



Kuva 2. Jokaisesta kuvatiedostosta tuotetaan MIX-metatietoskeeman (versio 2.0) mukainen tiedostokohtainen XML-tiedosto.

Kuvatiedosto **ei saa** sisältää mitään sellaisia elementtejä, joita ei ole analogisessa vastineessa. Poikkeuksen muodostavat samaan kuvaan tarpeen mukaan skannattavat kuvatiedoston värejä, harmaasävyjä, mittasuhteita ja resoluutiota todentavat skannaustekniset mittataulut sekä aineiston käytön kannalta oleelliset tietolaput.

Tietolapuilla voidaan kertoa asiakirjan tietosisältöön tai digitoinnissa tehtyihin ratkaisuihin liittyvistä asioista.

Esimerkki: Jos asiakirja on huomattavasti A0-kokoa suurempi, se voidaan joutua skannaamaan osissa. Esimerkiksi silloin kun suurikokoinen kartta on skannattu kahdessa osassa, kertoo tietolapuilta ilmenevä jaottelu, että kyseessä on osa 1/2, osa 2/2 jne.

Esimerkki: Asiakirjaa ei pystytä digitoimaan, koska sen sisältämä informaatio on täysin tuhoutunut joko ennen digitointia tai sen aikana (paperi on esimerkiksi mennyt silpuksi tai haurastunut jauheeksi). Tällöin tuhoutunut arkki korvataan A4-kokoisella tietolapulla. Tietolapulla kerrotaan, että arkistoyksikössä on tällä paikalla ollut vaurioitunut asiakirja, jonka tietosisältö on tuhoutunut ennen digitointia tai sen aikana. Tietolappu asetetaan vaurioituneen asiakirjan paikalle arkistoyksikköön ja digitoidaan osana muuta aineistoa. Digitoinnin jälkeen vaurioitunut arkki hävitetään tietoturvajätteenä.

Digitointiprosessissa **ei saa** poistaa merkintöjä sisältäviä sivuja.

Digitointiprosessissa tuotettuja kuvatiedostoja **saa** kääntää niiden skannauksen jälkeen vain 90 asteen välein. Jos kuvatiedostoja käsitellään skannauksen jälkeen, yksityiskohtainen kuvankäsittelyhistoria **pitäisi** tallentaa ainakin kuvatiedostojen metatietoihin ja mahdollisuuksien mukaan myös kuvatiedoston muodostumista kuvaileviin XML-muotoisiin tietoihin (MIX-metatiedot).

4.1.1 Kuvatiedostoa ja sen ominaisuuksia koskevat vaatimukset

Digitointiprosessissa tuotettava kuvatiedosto **pitää** tallentaa 24-bittisenä RGB-kuvana. Aineistotyyppikohtainen poikkeus ovat harmaasävyiset mikrofilmit, joiden kuvatiedostot **pitää** tallentaa 8-bittisenä.

Kuvatiedosto **ei saa** missään käsittelyvaiheessa olla tässä luvussa esitettyjä vaatimuksia heikkolaatuisempi. Sen sijaan kuvatiedosto **saa** olla parempilaatuinen. Esimerkiksi asiakirjan tietosisältö saattaa vaatia, että käytetään suurempaa resoluutiota, kuin tässä ohjeessa on määritelty. Resoluutiota nostaessa on kuitenkin syytä muistaa, että Kansallisarkistoon siirrettävien aineistojen kuvakohtainen enimmäiskoko pikseleissä on 65 535 x 65 535, ja sitä **ei saa** ylittää. Kuvatiedosto **pitää** aineistotyyppikohtaiset vaatimukset huomioiden tallentaa joko häviöttömästi pakatussa TIFF-muodossa tai häviöllisesti pakatussa JPEG-muodossa. Sitä **ei saa** tallentaa molemmissa muodoissa.

Mikäli prosessissa ensimmäisenä tuotettu kuvatiedosto on pakattu, lähtökohtaisesti sitä **ei pitäisi** käsitellä ensimmäisen tallennuskerran jälkeen ja tallentaa tämän jälkeen uudelleen. Toisen tallennuksen **saa** tehdä vain, kun se on digitointiprosessissa välttämätöntä, eikä pakkaamatonta formaattia esimerkiksi laitteiston kapasiteetin takia pystytä käyttämään.

Jos toinen tallennuskerta joudutaan tekemään, kuvatiedosto **ei saa** missään vaiheessa olla näissä vaatimuksissa määriteltyä heikkolaatuisempi.[19]

Jos tuotettua kuvatiedostoa käsitellään skannaamisen jälkeen, kuvatiedoston metatietoihin **pitää** dokumentoida ohjelmat, joilla sitä on käsitelty siinä järjestyksessä, missä käsittely on tapahtunut. Itse käsittelyohjelman ei tarvitse osata tätä tietoa merkitä, kunhan tieto lopulta tulee kirjattua oikein. Tiedot merkitään kuvatiedoston EXIF-metatietojen IFD0:ProcessingSoftware (0x000b) -tagiin. Arvot erotetaan toisistaan puolipisteellä (;):

[19] Tiedosto pitäisi toisella tallennuskerralla tallentaa muun muassa ilman subsampling-menetelmää[19], jotta laadun heikkeneminen on mahdollisimman vähäistä. Subsamplaus (engl. jpeg subsampling) tarkoittaa prosessia, jossa kuvaa pakatessa vierekkäisten pikselien värejä (ei luminanssia, joka säilyy entisellään) sulautetaan toisiinsa. Tällöin osa alkuperäisestä väritarkkuudesta katoaa. Esimerkiksi ImageMagick-ohjelmassa subsampling-toiminnon saa pois päältä ”sampling-factor 4:4:4” -parametrin avulla.

Esimerkki: <IFD0:ProcessingSoftware>Photoshop 5.0; ImageMagick 12.4; Kasittelyohjelma v2</IFD0:ProcessingSoftware>.

MIX-tiedoston ChangeHistory-elementtiin pitäisi tällöin lisäksi päivittää kuvatiedoston alkuperäiset tiedot, kuvatiedostolle tehdyt toimenpiteet ja kuvatiedoston nykyiset tiedot. Tämän asiakirjan liitteenä on toimintatapaa avaava esimerkkitiedosto (ChangeHistory.xml).

Kuvatiedostossa **pitää** olla tiedostomuodosta riippuen joko taulukossa 5 tai 6 esitetyt pakolliset tiedot koneluettavassa muodossa. Kuvatiedosto **saa** sisältää myös sellaisia metatietokenttiä, joita ei ole esitetty taulukoissa.

Skannattuihin kuviin **pitää** aina liittää oikea ICC-väriprofiili, joka vastaa käytettyä väriavaruutta. Tämä varmistaa, että kuvat näkyvät oikein eri laitteilla ja ohjelmistoilla. Jos väriprofiili lisätään tai sitä muunnetaan jälkikäteen, **pitää** varmistaa, että tämä tehdään oikein ja värintoisto vastaa alkuperäistä kuvaa.

Migraatiot ja tiedostokonversiot, joissa tiedoston tallennusformaattia muutetaan tai versioidaan, sisältävät aina tiedon säilymiseen liittyviä riskejä. On tärkeää digitoida analoginen asiakirja aina nimenomaan ohjeen mukaiseen tallennusformaattiin, eikä digitointiprosessissa tuotettava tiedosto saa missään käsittelyvaiheessa olla esitettyjä vaatimuksia heikkolaatuisempi. Muunlainen toimintapa on tämän ohjeen vaatimuksissa myös kielletty. Heikompilaatuisesta tallennusformaatista ei voi tuottaa parempilaatuista versiota myöhemmässä käsittelyssä.

Esimerkki 1: Digitointiprosessissa tuotetaan ensin sRGB-väriavaruudessa oleva tiedosto, joka myöhemmin tallennetaan Adobe RGB -väriavaruudessa. Vaikka Adobe RGB kattaa enemmän värisävyjä, siitä ei tässä tapauksessa ole hyötyä, koska kuvassa on edelleen vain ne värisävyt, jotka sRGB kattaa.

Esimerkki 2: Digitointiprosessissa tuotetaan ensin häviöllinen JPEG-kuva, joka myöhemmin tallennetaan häviöttömäksi TIFF-kuvaksi. JPEG-pakkauksen aiheuttama laadun heikkenemä ja pakkausvirheet säilyvät edelleen TIFF-kuvassa.

Yhteenvetona: Molemmissa esimerkeissä toimintatapa aiheuttaa sen, että lopullista kuvaa saatetaan luulla parempilaatuiseksi kuin se on. Oikea tapa toimia on tuottaa kuva jo skannausvaiheessa oikeilla määrityksillä. Parempilaatuisen kuvan saa aina tallennettua huonompilaatuiseksi, mutta ei toisin päin.

Vaadittu arvo -sarakkeen arvojen selitykset:

- **Aineistotyyppiriippuvainen:** Vaadittu arvo tarkistetaan oikean aineistotyypin kohdalta (luku 4.2).
- **Kuvatiedostokohtainen:** Arvot vaihtelevat kuvatiedostoittain. Muut määritykset saattavat kuitenkin heijastua näihin arvoihin. Esimerkiksi ”Kuvan leveys” ja ”Kuvan korkeus” ovat kuvatiedostokohtaisia, vaikka ne ovatkin riippuvaisia käytetystä resoluutiosta.
- **Tuotantoprosessiriippuvainen:** Arvot pysyvät samoina koko prosessin ajan eli ne eivät ole riippuvaisia aineistotyypeistä tai vaihtele kuvatiedostoittain. Esimerkkejä tällaisista arvoista ovat ”Digitaalisen kuvatiedoston tekijä” ja ”Käytetty skannerimalli”**.**

Taulukko 5. JPEG-kuvatiedoston vaatimukset

ELEMENTTI	TARKENNE	VAADITTU ARVO	METATIETOSKEEMA	Metatietokenttä (ID) / Exiftool Group: Tag
Formaatti	JPEG	**image** tai **jpeg**		MIME Type / File: MIMEType
Versio	JPEG part 1 versio	1.00 tai 1.01 tai 1.02		
Kuvan nimi	Kuva nimetään neljällä numerolla (0001.jpg, 0002.jpg jne,). Jos kuvia on 10 000 tai enemmän, käytetään viittä numeroa. Ks. luku 6.1.			
Kuvatiedoston koko	Tiedoston koko tavuina ja kokonaislukuna. Kansallisarkistoon siirrettävien kuvien enimmäiskoko pikseleissä on 65 535 x 65 535.	**Kuvatiedostokohtainen**		FileSize / System:FileSize
Fotometrinen tulkinta	Määrittää kuvan pikselien esitystavan eli sen, miten pikseliarvoja tulkitaan kuvassa.	**Aineistotyyppiriippuvainen**	Exif.Image	PhotometricInterpretation (262) / IFD0: PhotometricInterpretation
ICC-profiili	Kuvatiedoston metatietoihin tallennettu (engl. embedded) väriprofiili.	**Aineistotyyppiriippuvainen**	ICC	profileDescription / ICC_Profile: ProfileDescription
Bittisyvyys	Bittien määrä pikselin kanava-arvossa	8	Exif.Image	BitsPerSample (258) / IFD0: BitsPerSample
Näytteiden määrä	Kanava-arvojen määrä pikselissä	**Aineistotyyppiriippuvainen**	Exif.Image	SamplesPerPixel (277) / IFD0: SamplesPerPixel
Tiedoston pakkaaminen	JPEG	6 = JPEG (old-style) tai 7 = JPEG	Exif.Image	Compression (259) / IFD0: Compression

JPEG-laatu	JPEG-pakkauksen laatu asteikolla 0–100 % (tiedolle ei ole Exif-kenttää, mutta jpeg:n **pitää** olla tallennettu tällä laadulla.)	90 %		
Kuvan leveys	Kertoo kuvan leveyden pikselien määrällä per rivi	**Kuvatiedostokohtainen**	Exif.Image	ImageWidth (256) / IFD0: ImageWidth
Kuvan korkeus	Kertoo kuvan korkeuden pikselirivien määrällä kuvassa	**Kuvatiedostokohtainen**	Exif.Image	ImageLenght (257) / IFD0: ImageHeight
Digitaalisen kuvatiedoston tekijä	Organisaatio, joka on luonut kuvatiedoston analogisesta ilmentymästä	**Tuotantoprosessiriippuvainen**	Exif.Image	Artist (315) / IFD0: Artist
Digitointilaite (skannaus tai kuvaus)	Kertoo minkä valmistajan laitteella analoginen ilmentymä on muutettu digitaaliseen muotoon (valmistajan nimi)	**Tuotantoprosessiriippuvainen**	Exif.Image	Make (271) / IFD0: Make
Digitointilaitteen malli (skannaus tai kuvaus)	Tarkentaa digitointilaitetta kertomalla valmistajan mallin nimen	**Tuotantoprosessiriippuvainen**	Exif.Image	Model (272) / IFD0: Model
Digitoinnissa käytetyn laitteen sarjanumero	Yksilöi käytetyn laitteen	**Tuotantoprosessiriippuvainen**	Exif.Image	CameraSerialNumber (50735) / IFD0: CameraSerialNumber
Kuvatiedoston luomisessa käytetty ohjelma	Sovellus ja versio, jolla digitaalinen tiedosto on luotu	**Tuotantoprosessiriippuvainen**	Exif.Image	Software (305) / IFD0: Software
Kuvatiedoston luontipäivämäärä ja aika	Ilmaistaan muodossa: YYYY:MM:DD HH:MM:SS	**Kuvatiedostokohtainen**	Exif.Image	DateTime (306) / IFD0: ModifyDate
Lukusuunta	Tiedoston lukusuunta. Lukusuunta ei tarkoita kuvan tietosisällön lukusuuntaa, vaan tässä kerrotaan tiedoston lukusuunta.	**Tuotantoprosessiriippuvainen/ Kuvatiedostokohtainen**	Exif.Image	Orientation (274) / IFD0: Orientation
Resoluution mittayksikkö	Mittayksikkö, jota käytetään tulkitessa X- ja Y-resoluutiota	2 = inch	Exif.Image	ResolutionUnit (296) / IFD0: ResolutionUnit
XResoluutio	Pikselien määrä resoluution mittayksikköä kohden kuvan leveyssuunnassa.	**Aineistotyyppiriippuvainen**	Exif.Image	Image.XResolution (282) / IFD0: XResolution
YResoluutio	Pikselien määrä resoluution mittayksikköä	**Aineistotyyppiriippuvainen**	Exif.Image	Image.YResolution (283) / IFD0: YResolution

	kohden kuvan korkeussuunnassa.			
Kuvatiedoston käsittelyohjelma	Mikäli digitointiprosessissa luodaan ensin pakkaamaton tiedosto, jota käsitellään skannauksen jälkeen, tallennetaan käsittelyohjelman nimi ja versio. Jos käsittelyohjelmia on useita, kirjoitetaan ne siinä järjestyksessä, jossa niitä on käytetty. Erotusmerkkinä käytetään puolipistettä (;)	**Tuotantoprosessiriippuvainen**	Exif.Image	Image.Processing-Software (11) / IFD0: ProcessingSoftware

Taulukko 6. TIFF-kuvatiedoston vaatimukset

Elementti	Tarkenne	Vaadittu arvo	Metatieto-skeema	Metatietokenttä (ID) / Exiftool Group: Tag
Formaatti	TIFF	image/tiff		MIME Type / File: MIMEType
Versio	TIFF-tiedoston vaadittu versio	6.0		
Kuvan nimi	Kuva nimetään neljällä numerolla (0001.jpg, 0002.jpg jne,). Jos kuvia on 10 000 tai enemmän, käytetään viittä numeroa. Ks. luku 6.1.			
Kuvatiedoston koko	Tiedoston koko tavuina ja kokonaislukuna. Kansallisarkistoon siirrettävien kuvien enimmäiskoko pikseleissä on 65 535 x 65 535.	**Kuvatiedostokohtainen**		FileSize / System: FileSize
Fotometrinen tulkinta	Määrittää kuvan pikselien esitystavan eli sen, miten pikseliarvoja tulkitaan kuvassa.	**Aineistotyyppiriippuvainen**	Exif.Image	PhotometricInterpretation (262) / IFD0: PhotometricInterpretation
ICC-profiili	Kuvatiedoston metatietoihin tallennettu (engl. embedded) väriprofiili.	**Aineistotyyppiriippuvainen**	TIFF tag, private	ICC Profile (34675) / ICC_Profile
Bittisyvyys	Bittien määrä pikselin kanava-arvossa	8	Exif.Image	BitsPerSample (258) / IFD0: BitsPerSample
Näytteiden määrä	Kanava-arvojen määrä pikselissä	**Aineistotyyppiriippuvainen**	Exif.Image	SamplesPerPixel (277) / IFD0: SamplesPerPixel
Tiedoston pakkaaminen	Käytetty pakkaustyyppi	**Aineistotyyppiriippuvainen**	Exif.Image	Compression (259) / IFD0: Compression

Elementti	Tarkenne	Vaadittu arvo	Metatieto-skeema	Metatietokenttä (ID) / Exiftool Group: Tag
Kuvan leveys	Kertoo kuvan leveyden pikselien määrällä per rivi	**Kuvatiedostokohtainen**	Exif.Image	ImageWidth (256) / IFD0: ImageWidth
Kuvan korkeus	Kertoo kuvan korkeuden pikselirivien määrällä kuvassa	**Kuvatiedostokohtainen**	Exif.Image	ImageLenght (257) / IFD0: ImageHeight
Digitaalisen kuvatiedoston tekijä	Organisaatio, joka on luonut kuvatiedoston analogisesta ilmentymästä	**Tuotantoprosessiriippuvainen**	Exif.Image	Artist (315) / IFD0: Artist
Digitointilaite (skannaus tai kuvaus)	Kertoo minkä valmistajan laitteella analoginen ilmentymä on muutettu digitaaliseen muotoon (valmistajan nimi)	**Tuotantoprosessiriippuvainen**	Exif.Image	Make (271) / IFD0: Make
Digitointilaitteen malli (skannaus tai kuvaus)	Tarkentaa digitointilaitetta kertomalla valmistajan mallin nimen	**Tuotantoprosessiriippuvainen**	Exif.Image	Model (272) / IFD0: Model
Digitoinnissa käytetyn laitteen sarjanumero	Yksilöi käytetyn laitteen	**Tuotantoprosessiriippuvainen**	Exif.Image	CameraSerialNumber (50735) / IFD0: CameraSerialNumber
Kuvatiedoston luomisessa käytetty ohjelma	Sovellus ja versio, jolla digitaalinen tiedosto on luotu	**Tuotantoprosessiriippuvainen**	Exif.Image	Software (305) / IFD0: Software
Kuvatiedoston luontipäivämäärä ja aika (skannauspäivämäärä)	Ilmaistaan muodossa: YYYY:MM:DDTHH:MM:SS	**Kuvatiedostokohtainen**	Exif.Image	DateTimeOriginal (36867) / ExifIFD: DateTimeOriginal
Lukusuunta	Tiedoston lukusuunta. Lukusuunta ei tarkoita kuvan tietosisällön lukusuuntaa, vaan tässä kerrotaan tiedoston lukusuunta.	**Tuotantoprosessiriippuvainen/ Kuvatiedostokohtainen**	Exif.Image	Orientation (274) / IFD0: Orientation
Resoluution mittayksikkö	Mittayksikkö, jota käytetään tulkitessa X- ja Y-resoluutiota	2 = inch	Exif.Image	ResolutionUnit (296) / IFD0: ResolutionUnit
XResoluutio	Pikselien määrä resoluution mittayksikköä kohden kuvan leveyssuunnassa.	**Aineistotyyppiriippuvainen**	Exif.Image	Image.XResolution (282) / IFD0: XResolution
YResoluutio	Pikselien määrä resoluution mittayksikköä kohden kuvan korkeussuunnassa.	**Aineistotyyppiriippuvainen**	Exif.Image	Image.YResolution (283) / IFD0: YResolution

Elementti	Tarkenne	Vaadittu arvo	Metatieto-skeema	Metatietokenttä (ID) / Exiftool Group: Tag
Tavujärjestys	Määrittelee järjestyksen, jossa tavut luetaan, kun tiedosto avataan.	big endian tai little endian		ByteOrder
Kuvatiedoston käsittelyohjelma	Mikäli digitointiprosessissa luodaan ensin pakkaamaton tiedosto, jota käsitellään skannauksen jälkeen, tallennetaan käsittelyohjelman nimi ja versio. Jos käsittelyohjelmia on useita, kirjoitetaan ne siinä järjestyksessä, jossa niitä on käytetty. Erotusmerkkinä käytetään puolipistettä.	**Tuotantoprosessiriippuvainen**	Exif.Image	Image.ProcessingSoftware (11) / IFD0: ProcessingSoftware

4.1.2 Kuvatiedoston prosessointia kuvaavat MIX-metatiedot ja niiden rakenne

Tässä luvussa määritellyt pakolliset tekniset metatiedot dokumentoivat kuvatiedoston syntyhistoriaa, joka todentaa osaltaan myös prosessissa muodostuvan digitaalisen ilmentymän autenttisuutta eli alkuperää ja aitoutta.

Tekniset metatiedot **pitää** tuottaa MIX-metatietoskeeman (versio 2.0) mukaiseksi XML-tiedostoksi.[20] Jokaisella tuotetulla kuvatiedostolla **pitää** olla sitä vastaava XML-tiedosto.

MIX-skeeman kentät ovat joko dataelementtejä, jotka sisältävät tietyn arvon tai säiliöitä, jotka sisältävät yhden tai useamman dataelementin. Säiliöt voivat sisältää myös toisia säiliöitä dataelementteineen. Kaikki tässä ohjeessa vaaditut MIX-metatiedot ovat dataelementtejä eli ne sisältävät vain arvon. Metatiedot on eritelty taulukossa 7, jossa on kerrottu MIX-kentän nimen lisäksi sen tarkoitus vapaasti suomennettuna.

Velvoite-sarake kertoo kentän ja sen skeeman mukaisen pakollisuuden seuraavasti:

- P = pakollinen – tämä tieto **pitää** kuvata
- V = Vapaaehtoinen – tämä tieto **pitäisi** kuvata, mutta se ei ole pakollista

[20] The Library of Congress » Standards » MIX. Kongressin kirjaston verkkosivu http://www.loc.gov/standards/mix/ Viitattu 17.10.2024.

Taulukko 7. Kuvatiedostoa ja sen prosessointia kuvaavat metatiedot

MIX-kentän nimi	Kentän tarkoitus	Velvoite
objectIdentifierType	Dataelementti, joka määrittää sen järjestelmän tai verkkotunnuksen, jossa digitaalisen objektin yksilöivä merkkisarja (ID, ks. objectIdentifierValue) on uniikki.	P
objectIdentifierValue	Digitaalisen objektin identifioiva merkkisarja.	P
fileSize	Tiedoston koko tavuissa, esimerkiksi 72 839.	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: big endian tai little endian.	P
compressionScheme	Käytetty pakkausmenetelmä. Vaadittu arvo: JPEG, kun tallekappale on jpeg-formaatissa. 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 joko: MD5, SHA-1, SHA256, SHA384 tai SHA512	P
messageDigest	messageDigestAlgorithm-kentän määrittämän algoritmin tuottama merkki-sarja, esimerkiksi e8064dc0	P
imageWidth	Kuvan leveys pikseleissä, esimerkiksi 1 330.	P
imageHeight	Kuvan korkeus pikseleissä, esimerkiksi 1 600.	P
colorSpace	Dataelementti, joka määrittää kuvan väriavaruuden. Vaadittu arvo RGB tai WhiteIsZero tai BlackIsZero.	P
iccProfileName	Dataelementti, joka määrittää yleisesti käytetyn ICC-profiilin ni-men. 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
iccProfileURI	Jos ICC-profiili ei ole hyvin dokumentoitu, profiilin URI/URN tallennetaan tähän kenttään.	V

MIX-kentän nimi	Kentän tarkoitus	Velvoite
dateTimeCreated	Dataelementti, joka kertoo kuvatiedoston luontiajan. Ilmaistaan muodossa: YYYY-MM-DDTHH:MM:SS	P
imageProducer	Dataelementti, joka identifioi digitaalisen objektin 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 tarkentaa 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.” Ks. xSamplingFrequency ja ySamplingFrequency	P
xSamplingFrequency	Pikselien määrä resoluution mittayksikköä kohden leveyssuunnassa. Vaadittu arvo on aineistotyyppiriippuvainen. Ks. samplingFrequency unit	P
ySamplingFrequency	Pikselien määrä resoluution mittayksikköä kohden pystysuunnassa. Vaadittu arvo on aineistotyyppiriippuvainen. Ks. samplingFrequency unit	P
bitsPerSampleValue	Dataelementti, joka määrittelee jokaisessa kanavassa olevien bittien määrän. Vaadittu arvo: 8 tai 16	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: 3 tai 1	P

MIX-kentän nimi	Kentän tarkoitus	Velvoite
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 yksilöimän mittataulun mittausdata löytyy.	V

4.2 Aineistotyyppikohtaiset vaatimukset

Tästä luvusta löytyvät taulukoissa 5 ja 6 aineistotyyppiriippuvaiseksi merkityt arvot.

4.2.1 Pääsääntöisesti A4- ja foliokokoiset asiakirjat ja sidokset

Tämän alaluvun määrityksiä pitää noudattaa, kun digitoidaan A4-kokoisia tai foliokokoisia asiakirjoja ja sellaisia sidoksia, joiden tietosisältöä ei esitetä aukeamittain. Nämä aineistot digitoidaan yksisivuisena.

A4- tai foliokoiset asiakirjat ovat kooltaan pääosin määrämuotoista aineistoa. Foliokoossa arkinkoko voi vaihdella ja on yleensä standardoitua A4-kokoa hieman suurempi. Tämän alaluvun määrityksiä noudatetaan myös A4-kokoa pienempiin asiakirjoihin, esimerkiksi kortistoihin.

Mikäli asiakirjan tai sidoksen tietosisältö esitetään aukeamittain ks. alaluku 4.2.3.

Taulukko 8. Pääsääntöisesti A4- ja foliokokoisten asiakirjojen ja sidosten 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-laatutaso 90 %
Resoluutio	300 ppi, tarvittaessa alle A4 kokoisissa suurempi
Fotometrinen tulkinta	RGB
Bittisyvyys	8 bittiä kanava, 3 kanavaa (24 bit)
Tallekappaleen ICC-profiili	sRGB, eciRGB v2, ProPhoto RGB tai AdobeRGB (1 998)

4.2.2 Kartat ja piirustukset

Kartat ja piirustukset **pitää** digitoida kuvatiedostoiksi siten, että syntyvä kuvatiedosto on fyysisiltä mitoiltaan (n cm x n cm) 1:1-koossa analogisen asiakirjan kanssa resoluutiolla 300 ppi.

Digitoitavan analogisen asiakirjan tietosisällön eheyden varmistamiseksi asiakirja **pitää** digitoida siten, että kuva-ala ylittää sivujen reunat. Asiakirjan **pitää** olla kuvaushetkellä tasainen. Mikäli tasaisuuden saavuttamiseksi käytetään lasia tai muuta apuvälinettä, tämä **pitää** huomioida värinhallinnassa.

Kartat ja piirustukset **pitää** digitoida värillisiksi kuvatiedostoiksi (RGB, 24 bit).

Skannauslaitteisto **pitäisi** lähtökohtaisesti mitoittaa aineistolle sopivaksi, siten että kuva-ala on riittävä. Jos skannattava asiakirja on erittäin suuri eikä mahdu kokonaan esimerkiksi A0-kokoiseen skanneriin, sen **saa** skannata useassa osassa. Skannatut osat **pitää** tällöin merkitä tietolapulla, jossa jakotapa kerrotaan selkeästi.[21] Tietolappu **ei saa** peittää tietosisältöä. Jokaiseen osaan **pitää** jättää lisäksi riittävä marginaali, jotta asiakirjaa on helppo lukea myös jakokohdasta.

Osiin jakaminen voi olla tarpeen myös silloin, kun kuvatiedostosta tulee teknisesti liian suuri. Kansallisarkistoon siirrettävien kuvien enimmäiskoko pikseleissä on 65 535 x 65 535.

Karttojen ja piirustusten digitointia koskevat määritykset on koottu taulukkoon 9. Suositus on skannata aineisto 8 bittisenä (per kanava). Mikäli tietosisältö vaatii tarkempia värejä, skannaus voidaan tehdä 16 bittisenä. On tärkeää huomata, että tällöin tiedostokoko kasvaa.

Taulukko 9. Karttojen ja piirustusten määritykset

Elementti	Vaadittu arvo
Tallekappaleen formaatti	TIFF 6.0
Tallekappaleen pakkaaminen	Pakkaamaton tai LZW
Resoluutio	300 ppi
Fotometrinen tulkinta	RGB
Bittisyvyys	8 tai 16 bittiä per kanava, 3 kanavaa (24 tai 48 bit)
Tallekappaleen ICC-profiili	eciRGB v2, ProPhoto RGB, AdobeRGB (1 998)

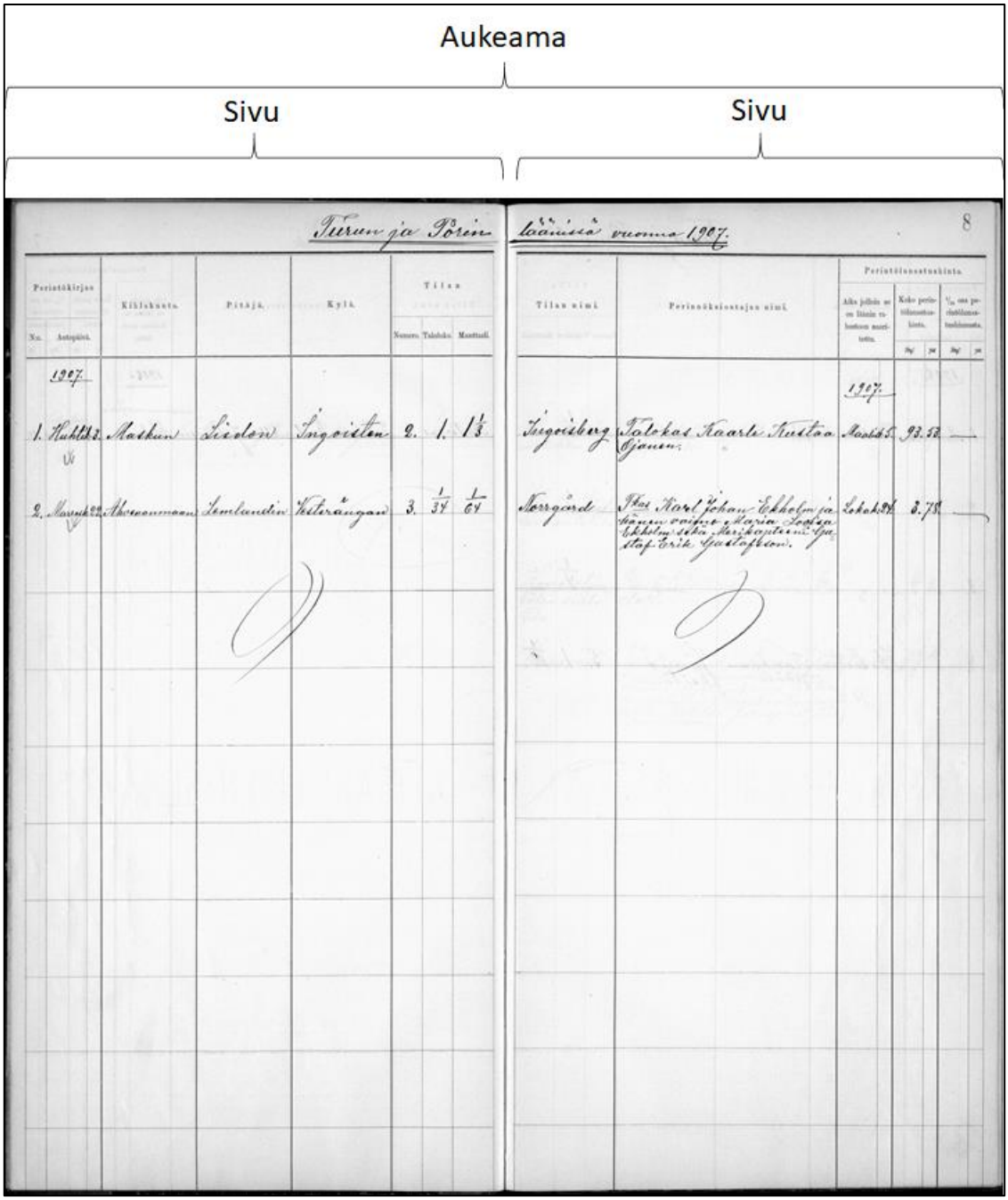
4.2.3 Asiakirjat ja sidokset, joiden tietosisältö esitetään aukeamittain

Mikäli asiakirjojen ja sidosten sisältämä tietosisältö esitetään aukeamittain, **pitää** digitoinnissa soveltaa tässä luvussa esitettyjä vaatimuksia. Tällaisia tietosisältöjä voivat olla esimerkiksi sivun ylittävät taulukot tai muu sisältö, joka on alun perin tuotettu analogiseen asiakirjaan niin, että analogisen

[21] Esimerkiksi: ”Osa 1/2" ja ”Osa 2/2”. Tietolappu asetetaan analogisen aineiston viereen niin, että väliin jää selvä marginaali.

asiakirjan yksittäisiä sivuja on ymmärrettävyyden vuoksi tarkasteltava rinnakkain eli aukeamittain (ks. kuva 3).

Tietosisällön eheyden varmistamiseksi analogiset asiakirjat **pitää** digitoida kokonaisuudessaan siten, että kuva-ala ylittää sivujen reunat.



Kuva 3. Tietosisältö ulottuu sidoksessa koko aukeamalle.

Alla olevassa taulukossa (Taulukko 10) esitetään aukeamittain digitoitavien asiakirjojen ja sidosten digitointia koskevat määritykset.

Taulukko 10. Asiakirjat ja sidokset, joiden tietosisältö esitetään aukeamittain

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-laatutaso 90 %
Resoluutio	300 ppi
Fotometrinen tulkinta	RGB
Bittisyvyys	8 bittiä kanava, 3 kanavaa (24 bit)
Tallekappaleen ICC-profiili	eciRGB v2, ProPhoto RGB, AdobeRGB (1 998), sRGB

4.2.4 Valokuvat

Valokuvalla tarkoitetaan tässä sekä yksittäisiä valokuvia että useamman valokuvan kokonaisuuksia, esimerkiksi valokuva-albumin sivuja.

Valokuvien sisältämän tietosisällön eheyden varmistamiseksi analogiset asiakirjat **pitää** digitoida kokonaisuudessaan[22] siten, että kuva-ala ylittää skannattavan kohteen reunat. Valokuvat **pitää** digitoida värillisiksi kuvatiedostoiksi (RGB, 24 bit).

Jokaisen digitoinnin yhteydessä **pitäisi** kuvata referenssikohde eli mittataulu. Värien ja sävyjen säätö **pitää** tehdä mittataulun avulla.

Alla olevassa taulukossa (Taulukko 11) esitetään valokuvien digitointia koskevat määritykset.

Taulukko 11. Valokuvien määritykset

Elementti	Vaadittu arvo
Tallekappaleen formaatti	TIFF 6.0
Tallekappaleen pakkaaminen	LZW
Resoluutio	Analogisen ilmentymän koon mukaan, ks. Taulukko 12 Resoluutiotaulukko
Fotometrinen tulkinta	RGB
Bittisyvyys	8 tai 16 bittiä per kanava, 3 kanavaa (24 tai 48 bit)
Tallekappaleen ICC-profiili	eciRGB v2, ProPhoto RGB, AdobeRGB (1 998)

Kun digitoitava asiakirja on pienikokoinen, **pitää** resoluutiota nostaa alla olevan resoluutiotaulukon (Taulukko 12) mukaisesti. Mikäli analoginen asiakirja on fyysisiltä mitoiltaan pienempi kuin taulu-

[22] Tarkoittaa myös valokuvien säilytysvälineitä ja kehyksiä, joissa on tietoja tai merkintöjä.

kossa 11 esitetään, **pitää** digitoinnissa käytettävää resoluutiota kasvattaa samaa laskentakaavaa soveltaen. Resoluutiota **ei saa** muuttaa kuvatiedostoa käsiteltäessä, jotta analogisen asiakirjan fyysisten mittojen arviointi on digitaalisen tallekappaleen perusteella mahdollista. Tavoitteena on aina ~A3-kokoinen (cm x cm) tallekappale resoluutiolla 300 ppi.

Taulukko 12. Resoluutiotaulukko

Analogisen asiakirjan koko	Resoluutio
6x9cm	1 600 ppi
9x12cm	1 200 ppi
13x18cm	900 ppi
18x24cm	600 ppi
A3 (tai suurempi)	300 ppi

4.2.5 Lasinegatiivit, muut negatiivit ja diat

Lasinegatiivien, muiden negatiivien ja diojen tietosisällön eheyden varmistamiseksi analogiset asiakirjat **pitää** digitoida kokonaisuudessaan siten, että kuva-ala ylittää skannauksen kohteena olevan analogisen negatiivin tai dian reunat. Mikäli negatiivi tai dia on asetettu kehyksiin, rajauksessa **pitää** noudattaa äärimmäistä huolellisuutta, jotta tietosisältöä ei rajaudu pois kuvatiedostosta. Diat ja negatiivit[23] **pitää** digitoida värillisiksi kuvatiedostoiksi (RGB, 24 bit).

Lasinegatiivien, muiden negatiivien ja diojen digitointia koskevat määritykset on koottu alla olevaan taulukkoon (Taulukko 13).

Taulukko 13. Lasinegatiivien, muiden negatiivien ja diojen määritykset

Elementti	Vaadittu arvo
Tallekappaleen formaatti	TIFF 6.0
Tallekappaleen pakkaaminen	LZW
Resoluutio	Analogisen ilmentymän koon mukaan, ks. Taulukko 12 Resoluutiotaulukko
Fotometrinen tulkinta	RGB
Bittisyvyys	8 tai 16 bittiä per kanava, 3 kanavaa (24 tai 48 bit)
Tallekappaleen ICC-profiili	eciRGB v2, ProPhoto RGB, AdobeRGB (1 998)

[23] Negatiivit digitoidaan suorassa ja oikein päin, ei peilikuvina. Kuvaamalla digitoitaessa negatiivit asetetaan emulsiopuoli alaspäin, kiiltävä puoli kameraa kohden. Lasinegatiivit on suositeltavaa kuvata emulsiopuoli ylöspäin, jolloin ne digitoituvat peilikuvana. Varmista, että digitoitu kuva tallennetaan lopulta oikean lukusuunnan mukaisesti.

Kun digitoitava analoginen asiakirja on pienikokoinen, **pitää** resoluutiota nostaa resoluutiotaulukon (Taulukko 12) mukaisesti. Mikäli analoginen asiakirja on fyysisiltä mitoiltaan pienempi kuin taulukossa 13 esitetään, **pitää** digitoinnissa käytettävää resoluutiota kasvattaa samaa laskentakaavaa soveltaen. Tallekappaletiedoston resoluutiota **ei saa** muuttaa kuvatiedostoa käsiteltäessä, jotta analogisen asiakirjan fyysisten mittojen arviointi on digitaalisen tallekappaleen perusteella mahdollista.

Tavoitteena on aina ~A3-kokoinen (cm x cm) tallekappale resoluutiolla 300 ppi.

4.2.6 Mikrofilmit ja -kortit

Mikrofilmin pienennyskerroin **pitää** tarkistaa filmirullan alusta. Mikäli pienennyskerrointa ei ole ilmoitettu, suurennuskerroin **pitää** asettaa riittävän suureksi, jotta tuotetun digitaalisen asiakirjan tietosisältö on luettavissa. Mikrofilmien kohdalla filmille kuvattu aineisto saadaan suurennuskertoimen avulla toistumaan riittävän luettavana resoluutiolla 300 ppi. Skannauksessa käytettävä resoluutio on suurennuskerroin kertaa (x) haluttu resoluutio (esimerkiksi 12 x 300 = 3 600). Mikrofilmi itsessään on varsin jyrkkä tallenneväline, joten digitoinnissa pitää varoa jyrkkyyden korostamista, jotta luvussa 4 esitetyt määritykset toteutuvat.

Alkuperäisen aineiston oikean koon todentamisessa on voitu käyttää myös muita menetelmiä (esimerkiksi mittanauha). Käytettyjä menetelmiä **pitää** tarkastella tapauskohtaisesti. Samalla **pitää** pyrkiä arvioimaan alkuperäisen, mikrofilmille kuvatun analogisen asiakirjan fyysisiä mittoja, vaikka mitään edellä mainituista menetelmistä ei olisi mikrokuvauksen yhteydessä sovellettu.

Alla olevassa taulukossa (Taulukko 14) esitetään mikrofilmien digitointia koskevat määritykset.

Taulukko 14. 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-laatutaso 90 %
Resoluutio	300 ppi
Fotometrinen tulkinta	BlackIsZero tai WhiteIsZero (Vain Tiff)
Bittisyvyys	8 tai 16 bittiä per kanava, 1 kanava (8 tai 16 bit)
Tallekappaleen ICC-profiili	Gray gamma 2.2

Värilliseen mikrofilmiin **pitää** 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 (1 998) tai sRGB (ks. Taulukko 15).

Taulukko 15. Värillisten 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-laatutaso 90 %
Resoluutio	300 ppi
Fotometrinen tulkinta	RGB
Bittisyvyys	8 tai 16 bittiä per kanava, 3 kanavaa (24 tai 48 bit)
Tallekappaleen ICC-profiili	eciRGB v2, ProPhoto RGB, AdobeRGB (1 998) tai sRGB

4.2.7 Röntgenfilmit

Röntgenfilmien digitointi vaatii siihen erikoistuneita filmiskannereita, sillä kuvien tiheysalue (d-max usein yli 4,0) on erittäin laaja. Normaalit skannerit pystyvät harvoin skannaamaan tällaisen tiheysalueen tarkasti yhdellä skannauskerralla. HDR-digitointi voi auttaa säilyttämään laajan sävyalueen.

Lasia tai muuta tukea **saa** käyttää rullautuneen tai käpristyneen aineiston suoristamiseen, mutta se **ei saa** ei vahingoittaa aineistoa. Käytettävän lasin **pitäisi** olla heijastamatonta ja häiriökuvioita (kuten Newtonin renkaita) vähentävää. Konservaattorin vastuulla on tunnistaa materiaalit ja menetelmät sekä arvioida vauriot ja riskit, mutta aineiston valmistelun ja puhdistuksen **saa** tehdä aineistovalmistelija konservaattorin ohjauksessa. Suoristuksessa ja muussa käsittelyssä **pitää** varmistaa, ettei emulsion halkeamia tai pohjamateriaalin vaurioita synny.

Referenssikohdetta eli mittataulua **pitäisi** käyttää digitoinnissa, ja värien säädöt **pitäisi** tehdä mittataulun avulla.

Tallekappaleet **pitää** skannata niin, että kuvaan jää selvät marginaalit aineiston jokaiselle sivulle. Käyttökappaleisiin voidaan tehdä tarkemmat rajaukset ja mahdolliset korjaukset tai säädöt muun muassa väreihin.

Digitointi **ei saa** aiheuttaa vahinkoa alkuperäiselle aineistolle. Digitointilaitteiden, -menetelmien ja aineiston käsittelyn **pitää** olla sellaisia, että vahingoittumisen riski minimoidaan. Vahinkoriskin ilmetessä työ **pitää** keskeyttää välittömästi.

Aineistoa **pitää** käsitellä tilassa, jossa lämpötila ja kosteus ovat optimaaliset. Altistaminen valolle **pitää** minimoida. Käsittelyssä **pitää** käyttää aina puhtaita, hyvin istuvia nitriilihanskoja. Koskettamista kuva-aloihin, painopintoihin tai valokuvien emulsiopintoihin **pitää** välttää.

Taulukko 16. Röntgenfilmien määritykset

Elementti	Vaadittu arvo
Tallekappaleen formaatti	TIFF 6.0
Tallekappaleen pakkaaminen	Tiff = LZW tai pakkaamaton
Resoluutio	400 ppi
Fotometrinen tulkinta	BlackIsZero tai WhiteIsZero
Bittisyvyys	16 bittiä per kanava, 1 kanava (16 bit)
Tallekappaleen ICC-profiili	Gray Gamma 2.2

5 Audiovisuaalisten aineistojen vaatimukset

Analogiset audiovisuaaliset aineistot ovat ominaisuuksiltaan sekä laiteriippuvaisia että itsetuhoutuvia. Niiden fyysinen säilyvyys ja käyttöikä on huomattavasti lyhyempi kuin muilla asiakirja-aineistoilla. Laiteriippuvuus tuo myös omat käytön haasteensa. Audiovisuaalisten aineistojen sisällön säilyttämisen näkökulmasta digitointi on keskeistä.

Audiovisuaaliset aineistot jaetaan tässä ohjeessa kahteen ryhmään, ääninauhoihin ja kuvanauhoihin. Aineistojen digitointiprosessissa tuotetaan tallekappaletiedosto, joka on aineistotyypin mukaan joko ääni- tai videotiedosto.

Audiovisuaalisten aineistojen digitoinnin tavoitteena on taata aineistojen pysyvä säilyminen digitaalisessa muodossa. Digitoitujen aineistojen **pitää** sisältää mahdollisimman tarkoin sama sisällöllinen kuva- ja ääni-informaatio kuin alkuperäisissä ääninauhoissa tai kuvanauhoissa.

5.1 Ääninauhat

Ääninauhoihin lasketaan muun muassa:

- äänikelanauha (avokelanauha, kelanauha)
- C-kasetti
- minikasetti.

Avokela- ja C-kasettinauhureiden äänipäiden atsimuuttikulma[24] **pitää** säätää nauhakohtaisesti mittaustekniikkaa apuna käyttäen. Nauhat **pitää** digitoida toistamalla jokainen monoraita tai stereoraitapari kokonaisuudessaan alusta loppuun alkuperäisellä nopeudella.

Analoginen äänisignaali muunnetaan digitaaliseksi erillisellä AD-muuntimella. Digitointiparametreina käytetään 24 bitin resoluutiota ja 48 000 hertsin näytteenottotaajuutta, mutta 96 000 hertsin näytteenottotaajuutta voidaan käyttää tapauskohtaisesti. Nauhoissa oleva ääni **pitää** tallentaa mahdollisimman muuttamattomana, joten ääntä **ei saa** parannella digitoinnin yhteydessä. Normalisointia tai kompressointeja **ei saa** tehdä.

Digitoitujen tiedostojen muotona **pitää** käyttää häviötöntä LPCM-koodattua WAV-muotoa.

Äänenvoimakkuuden taso **pitää** säätää digitoitaessa tallennekohtaisesti niin, että huippuarvo asettuu 0- ja 6 dB:n välille. Äänen tasoa **ei saa** säätää ohjelmallisesti digitoinnin jälkeen.

[24] Atsimuuttikulmalla tarkoitetaan äänipään kulmaa suhteessa ääninauhaan. Jos kulma ei ole kohtisuora, on eri kanavien välillä taajuudesta riippuva vaihe-ero.

Jokainen nauhakokonaisuus **pitää** digitoida kokonaisuudessaan pääsääntöisesti samoilla asetuksilla. Nauhakokonaisuudella tarkoitetaan tässä yhteydessä nauhan yhtä puolta kokonaisuudessaan.

Taulukko 17. Ääninauhojen määritykset

Elementti	Vaadittu arvo
Tallekappaleen formaatti	LPCM-koodattu WAV
Tallekappaleen pakkaaminen	Pakkaamaton
Näytteenottotaajuus	48 kHz / 96 kHz
Bittisyvyys	24 bit

Äänitiedostojen pakollisten teknisten metatietojen metatietoskeema on AudioMD[25]. Pakolliset tekniset metatiedot on esitetty alla olevassa taulukossa (Taulukko 18).

Taulukko 18. Äänitiedostojen teknisen metatiedon skeema

Termi	Metatieto	Tuntematon arvo sallittu
Äänikoodauksen tyyppi	audio_data_encoding	Kyllä
Kesto	duration	Kyllä
Datavirran nopeus	data_rate	Kyllä
Datavirran tila	data_rate_mode	Ei
Bittisyvyys	bits_per_sample	Kyllä
Ohjelmisto (koodekki)	codec_creator_app	Kyllä
Versio (koodekki)	codec_creator_app_version	Kyllä
Nimi (koodekki)	codec_name	Kyllä
Pakkauksen laatu (koodekki)	codec_quality	Ei
Näytteenottotaajuus	sampling_frequency	Kyllä
Kanavien määrä	num_channels	Kyllä

[25] The Library of Congress » Standards » AudioMD/VideoMD. Kongressin kirjaston verkkosivu http://www.loc.gov/standards/amdvmd/ Viitattu 19.9.2024.

5.2 Kuvanauhat

Kuvanauhoihin lasketaan muun muassa:

- VHS-nauhat
- Betacam-nauhat
- DVcam-nauhat
- miniDV-nauhat.

Kuvanauhoista digitoitujen videotiedostojen **pitää** sisältää mahdollisimman tarkoin sama sisällöllinen kuva- ja ääni-informaatio kuin alkuperäisissä nauhoissa. Kuvanauhat **pitää** digitoida laadukkailla studiokäyttöön tarkoitetuilla laitteistoilla. Kuvanauhojen määritykset on esitetty alla olevassa taulukossa (Taulukko 19).

Taulukko 19. Kuvanauhojen määritykset

Elementti	Vaadittu arvo
Tallekappaleen formaatti	FFV1 versio 3 [FFV1] fmt/569
Tallekappaleen pakkaaminen	Pakkaamaton

Videotiedostojen pakollisten teknisten metatietojen metatietoskeema on VideoMD[26]. Pakolliset tekniset metatiedot on esitetty alla olevassa taulukossa (Taulukko 20).

Taulukko 20. Kuvanauhatiedostojen teknisen metatiedon skeema

Termi	Metatieto	Tuntematon arvo sallittu
Ääni	sound	Ei
Kesto	duration	Kyllä
Datavirran nopeus	data_rate	Kyllä
Datavirran tila	data_rate_mode	Ei
Bittisyvyys	bits_per_sample	Kyllä
Väritila	color	Ei
Ohjelmisto (koodekki)	codec_creator_app	Kyllä
Versio (koodekki)	codec_creator_app_version	Kyllä
Nimi (koodekki)	codec_name	Kyllä
Pakkauksen laatu (koodekki)	codec_quality	Ei
Kuvataajuus	frame_rate	Kyllä
Kuvan korkeus	height	Kyllä
Kuvan leveys	width	Kyllä

[26] The Library of Congress » Standards » AudioMD/VideoMD. Kongressin kirjaston verkkosivu http://www.loc.gov/standards/amdvmd/ . Viitattu 19.9.2024.

Pikselin kuvasuhde (PAR)	par	Kyllä
Näytön kuvasuhde (DAR)	dar	Kyllä
Värien alinäytteistys	sampling	Kyllä
Signaalin muoto	signal_format	Kyllä

6 Digitointiprosessissa muodostettava hakemistorakenne

Mikäli tässä ohjeessa esitettyjen vaatimusten mukaan digitoitu tietoaineisto siirretään Kansallisarkistoon, se tallennetaan Kansallisarkiston siirto-ohjeiden mukaiseen hakemistorakenteeseen.[27] Kuva-aineistolle ja AV-aineistolle on omat hakemistorakenteensa. Kuhunkin juurihakemistoon tallennetaan pääsääntöisesti yhden arkistoyksikön sisältö. AV-aineistoa sisältävä arkistoyksikkö **saa** sisältää useiden nauhojen tiedostot.

Kuva-aineistojen hakemistorakenteen ja siirtopaketin tarkempi määrittely on ”Digitoidut kuva-aineistot – siirtopaketin muodostaminen” -ohjeessa ja AV-aineistojen hakemistorakenteen ja siirtopaketin tarkempi määrittely ”Muut digitaaliset tietoaineistot – siirtopaketin muodostaminen” -ohjeessa. Ajantasaiset ohjeet ovat Kansallisarkiston ohjepankissa.

Yllä mainituissa siirto-ohjeissa määritellyt ja tässä luvussa tiivistetysti kuvatut hakemistorakenteet koskevat Kansallisarkistoon siirrettäviä tietoaineistoja eivätkä sulje pois muiden tallennusrakenteiden käyttöä. Organisaatio voi tallentaa digitoidusta kuva-aineistosta aina esimerkiksi myös PDF-muotoiset käyttökappaleet omiin järjestelmiinsä haluamassaan tietorakenteessa.

6.1 Kuva-aineiston hakemistorakenne

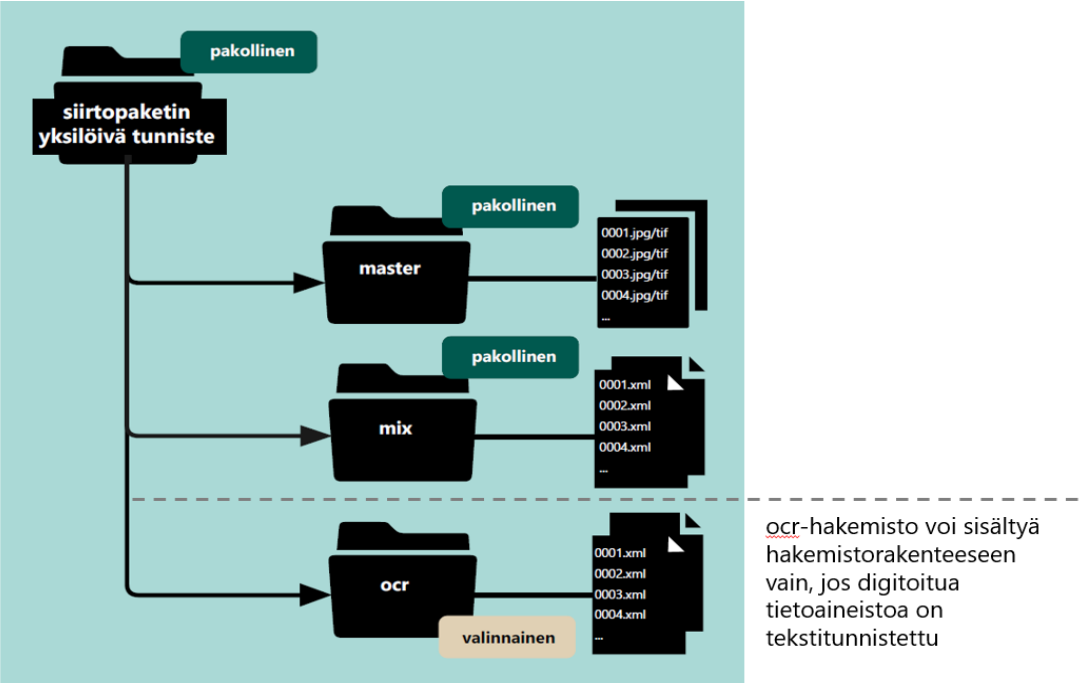
Digitointiprosessissa tuotetut kuvatiedostot, niiden MIX-metatietotiedostot ja tekstitunnistuksessa mahdollisesti tuotetut tiedostot **pitää** tallentaa kuvan 4 mukaiseen hakemistorakenteeseen, kun aineisto siirretään Kansallisarkistoon.

- Juurihakemiston nimen **pitää** olla yksilöivä, nimen pituus saa olla enintään 256 merkkiä[28] ja se saa sisältää ainoastaan seuraavia merkkejä: a-z, A-Z ja 0–9.
- Kuvatiedostot, niiden tekniset metatiedot ja mahdolliset tekstitunnistuksessa luodut tiedostot tallennetaan omiin alihakemistoihinsa.
- Alihakemistot **pitää** nimetä seuraavasti:
    - kuvatiedostojen hakemisto: **master**
    - MIX-metatietotiedostojen hakemisto: **mix**
    - tekstintunnistuksessa luotujen tiedostojen hakemisto: **ocr**

[27] On hyvä huomioida, että hakemistorakenteen luominen digitoiduille kuva-aineistoille jälkikäteen saattaa olla työlästä, ellei sitä ole tuotettu digitointiprosessissa.

[28] Unix/Linux-järjestelmissä nimen ja sen polun yhteispituus saa olla enintään 4096 merkkiä, MacOS-järjestelmissä ei ole rajoituksia. Windowsin rajoituksista lisää https://learn.microsoft.com/en-us/windows/win32/fileio/maximum-file-path-limitation?tabs=registry Viitattu 14.11.2024

Kaikkien alihakemistojen tiedostot **pitää** nimetä lähtökohtaisesti nelinumeroisella lukujonolla. Lukujono on juokseva niin, että ensimmäinen tiedosto saa nimen 0001, toinen 0002 jne. Mix- ja ocr-hakemistojen sisältämien tiedostojen tiedostonimien **pitää** vastata master-hakemiston kuvatiedostojen nimiä. Kuvatiedoston nimeä vastaava lukujono siis linkittää digitoinnissa luodun kuvatiedoston ja siihen liittyvät mix-metatiedot sekä mahdollisen tekstitunnistuksella tuotetun sisällön toisiinsa.



Kuva 4. Tuotettava hakemistorakenne

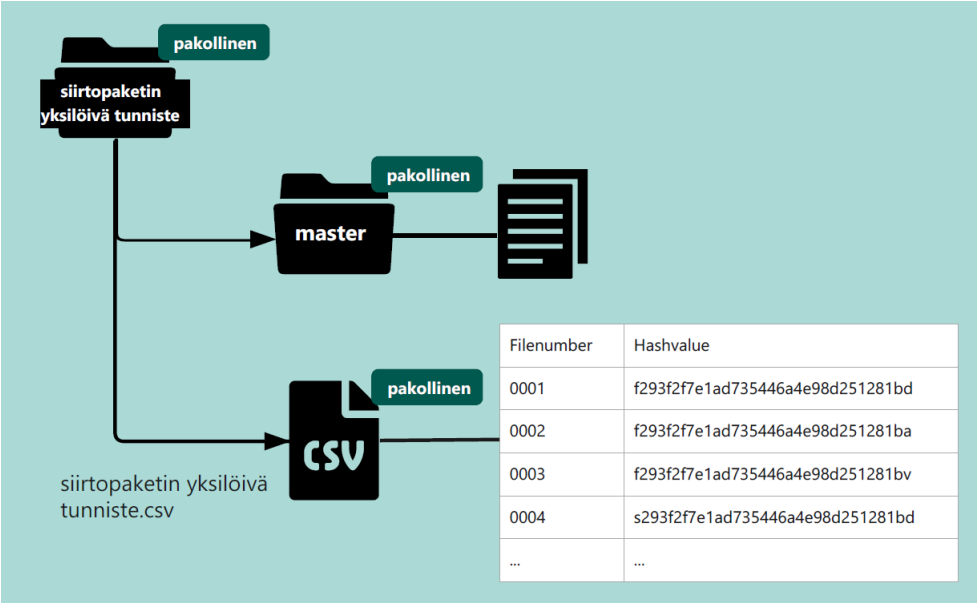
Juurihakemisto **ei saa** sisältää muita kuin kuvan 4 mukaisia alihakemistoja. Kun aineisto siirretään Kansallisarkistoon, jokainen juurihakemisto **pitää** paketoida TAR-tiedostoksi. Tätä TAR-tiedostoa kutsutaan siirtopaketiksi. TAR-tiedoston **saa** lisäksi pakata häviöttömään GZIP- (. tar.gz) tai BZIP2- (.tar.bz2) muotoon.

Hakemistorakenteen ja siirtopaketin tarkempi kuvaus on Kansallisarkiston ajantasaisissa ohjeissa.

6.2 Audiovisuaalisen aineiston hakemistorakenne

AV-aineiston digitointiprosessissa tuotetut tiedostot **pitää** tallentaa Sähköisen arkistoinnin palvelun ”Muut digitaaliset tietoaineistot – siirtopaketin muodostaminen” -ohjeessa kuvattuun hakemistorakenteeseen, kun ne siirretään Kansallisarkistoon.

- Juurihakemiston nimen **pitää** olla yksilöivä, nimen pituus saa olla enintään 256 merkkiä[29] ja se saa sisältää ainoastaan seuraavia merkkejä: a-z, A-Z ja 0–9.
- Digitoinnissa tuotetut tiedostot **pitää** tallentaa master-nimiseen alihakemistoon.
    - Tiedostot **pitää** nimetä lähtökohtaisesti nelinumeroisella lukujonolla. Lukujono on juokseva niin, että ensimmäinen tiedosto saa nimen 0001, toinen 0002 jne.
- Juurihakemistoon tallennettavan CSV-tiedoston pitää sisältää master-hakemiston sisältämien tiedostojen tiedostonimet (sarake Filenumber) ja tiedostoille lasketut tarkistussummat (sarake Hashvalue). Filenumber-sarakkeeseen ei saa sisällyttää tiedostopäätettä. Tarkistussumma **pitää** antaa muodossa MD5[30].
    - Tiedosto **pitää** nimetä siirtopaketin yksilöivällä tunnisteella, eli täsmälleen samalla merkkijonolla kuin juurihakemisto.
    - Tiedoston merkistön **pitää** olla UTF-8[31].
    - Kenttien ympärillä **ei saa** olla lainausmerkkejä.



Kuva 5. AV-aineistoista tuotettava hakemistorakenne

[29] Unix/Linux-järjestelmissä nimen ja sen polun yhteispituus saa olla enintään 4096 merkkiä, MacOS-järjestelmissä ei ole rajoituksia. Windowsin rajoituksista lisää https://learn.microsoft.com/en-us/windows/win32/fileio/maximum-file-path-limitation?tabs=registry Viitattu 14.11.2024

[30] MD5 on hajautusalgoritmi, joka tuottaa 128-bittisen hajautusarvon syötteestä ja sitä käytetään tiedoston eheyden tarkistamiseen (eheystarkiste).

[31] UTF-8 on muuttuvan pituinen merkistökoodausstandardi, joka tukee kaikkia Unicode-merkkejä ja on yhteensopiva ASCII:n kanssa.

Juurihakemiston **pitää** sisältää vain kuvan 5 mukainen hakemisto tiedostoineen. Kun aineisto siirretään Kansallisarkistoon, jokainen juurihakemisto **pitää** paketoida TAR-tiedostoksi. Tätä TAR-tiedostoa kutsutaan siirtopaketiksi. TAR-tiedoston **saa** lisäksi pakata häviöttömään GZIP- (.tar.gz) tai BZIP2- (.tar.bz2) muotoon.

Hakemistorakenteen ja siirtopaketin tarkempi kuvaus on Kansallisarkiston ajantasaisissa ohjeissa.

Huomaa, että myös esimerkiksi ääni- tai kuvanauhan tekstimuotoon litteroitu sisältö tai muu vastaava sisältö voidaan siirtää Kansallisarkistoon. Siirron toteutustapa määritellään tarkemmin siirron yhteydessä.

6.3 Esimerkkipaketit

Kuvatiedostojen siirrossa käytettävistä siirtopaketeista on muodostettu ohjeen liitteeksi kaksi esimerkkipakettia:

1. EsimerkkipakettiJPEG2020.tar
2. EsimerkkipakettiTIF2020.tar

Master-hakemistojen sisältämät kuvatiedostot eivät esimerkkipaketeissa ole kuvanlaadullisia referenssejä. Kuvatiedostot sisältävät tässä asiakirjassa pakollisiksi määritellyt metatiedot. Mix-hakemistojen tiedostot ovat esimerkkejä liitteen paketissa olevista master-tiedostoista, ellei elementtien kohdalla todeta toisin.

7 Muutokset

Taulukko 21. Ohjeeseen tehdyt muutokset

Muutos selitettynä	**Koskee ohjetta:** Kansallisarkiston vaatimukset hävittämiseen tähtäävään digitointiin (KA/3357/07.01.02.04.02/2019)	**Koskee ohjetta:** Kansallisarkiston määritykset arkistoitavien erikoisaineistojen digitointiin (KA/15780/07.01.02.04.02/2020)
Kansallisarkiston seulontapäätösmenettelyyn liittyvä ohjeistus on päivitetty.	**Kyllä**	**Kyllä**
Termejä ja käsitteitä koskevat taulukot on siirretty samaan lukuun. Käsitteitä on lisätty ja joitain käsitteitä on tarkennettu.	**Kyllä**	**Kyllä**
Yleiset luvut (esimerkiksi johdanto) on yhdistetty.	**Kyllä**	**Kyllä**
Aiemmista ohjeista puuttuneet AV-aineistojen ja röntgenfilmien digitointia koskevat ohjeet ja vaatimukset on lisätty.	**Kyllä**	**Kyllä**
Kuvatiedostoa koskevat pakolliset metatiedot on jaoteltu kahteen taulukkoon (taulukko 5, JEPG ja taulukko 6, TIFF). Lisäksi taulukoissa esitettyjä vaatimuksia on pyritty avaamaan aiempaa selvemmin.	**Kyllä**	**Kyllä**
Alaluvun 4.1.1 taulukoiden 5 ja 6. ”Käsittelyohjelma”-metatietoon on lisätty tarkenne siitä, että siihen tallennetaan kaikki käytetyt käsittelyohjelmat.	**Kyllä**	**Kyllä**
Taulukossa 5 ”Kuvatiedoston luontipäivämäärä ja aika” -metatiedon vaatimus DateTimeOriginal on muutettu muotoon DateTime (306).	**Ei**	**Kyllä**

Aineistotyyppikohtaiset vaatimukset on koottu lukuun 4.2 (sekä pääsääntöisesti A4- tai foliokokoiset asiakirjat ja sidokset että erikoisaineistot). Taulukoissa 5 ja 6 "Vaadittu arvo" -sarakkeen arvo "Aineistotyyppiriippuvainen" tarkoittaa sitä, että vaadittu metatietoelementti riippuu aineistotyypistä ja se on tarkistettava luvusta 4.2.	**Kyllä**	**Kyllä**
Taulukot kuvatiedostoa ja kuvatiedoston prosessointia kuvaavista metatiedoista ja rakenteesta on yhdistetty samaan taulukkoon (taulukko 7).	**Kyllä**	**Kyllä**
Taulukko 7: MIX-metatietokentän nimi iccProfile URL on korjattu muotoon iccProfileURI. MIX-metatietojen kansainvälisten määritysten mukaisessa skeemassa (versio 2) on iccProfileURI-niminen elementti. Skeeman kuvauksessa ja Kansallisarkiston digitointimäärityksissä nimi on ollut aiemmin version 1 mukaisessa muodossa iccProfileURL. Kansallisarkistoon siirrettävissä digitoiduissa aineistossa tulee käyttää skeeman mukaista muotoa iccProfileURI.	**Kyllä**	**Kyllä**
Värillisten mikrofilmien vaatimukset on taulukoitu omaan taulukkoonsa (taulukko 15).	**Ei**	**Kyllä**
Alaluku 4.1.1. Lisätty poikkeus siitä, milloin pakattu kuvatiedosto voidaan tallentaa toisen kerran. Lisäksi on määritelty, miten ProcessingSoftware-kenttään tulee kirjata käsittelyohjelmat. Määritelty myös MIX-tiedoston ChangeHistory-elementin	**Kyllä**	**Kyllä**

päivitys, kun kuvatiedostoa muokataan.		
Lisätty tarkennuksia digitoidun kuva-aineiston siirtopaketin muodostamisen ohjeeseen (Luku 6). Lisätty kuvaus siirtopakettirakenteesta, jossa digitoitu AV-aineisto siirretään Kansallisarkistoon.	**Kyllä**	**Kyllä**
Tarkennettu ohjetta suurista kuvatiedostoista sekä lisätty kuville tarkat maksimimitat pikseleissä (Luku 4.1.1.)	**Kyllä**	**Kyllä**
Poistettu luku "5.2 Tunnistetun tekstin tallennusformaatti."	**Kyllä**	