

7.1 Tarvittavat tekniset metatiedot

Kansallisarkiston vaatimukset hävittämiseen tähtäävään digitointiin (Pääasiassa A4-kokoinen aineisto):

Taulukko 4: Kuvatiedoston (JPEG) pakolliset metatiedot

Elementti	Tarkenne	Vaadittu arvo	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				
Kuvatiedoston koko				
Väritila	Kuvatiedoston väritila	RGB	Exif.Image	PhotometricInterpretation (262)
ICC -profiili	Kuvatiedoston metatietoihin tallennettu väriprofiili.	sRGB	ICC	profileDescription
Bittisyvyys	Bittien määrä pikselin kanava-arvossa	8 8 8	Exif.Image	BitsPerSample (258)
	Kanava-arvojen määrä pikselissä	3	Exif.Image	SamplesPerPixel (277)
JPEG-laatu	JPEG-pakkauksen laatu asteikolla 0 -100%	90%		
Kuvatiedoston tekijä	Organisaatio, joka on luonut kuvatiedoston analogisesta ilmentymästä		Exif.Image	Artist (315)
Kuvan korkeus	Kertoo kuvan korkeuden pikselirivien määrällä kuvassa		Exif.Image	ImageLength(257)
Kuvan leveys	Kertoo kuvan leveyden pikselirivien määrällä kuvassa		Exif.Image	ImageWidth(256)
Digitointilaite	Kertoo minkä valmistajan laitteella analoginen objekti on muutettu digitaaliseksi (valmistajan nimi)		Exif.Image	Make (271)
Digitointilaitteen malli	Tarkentaa digitointilaitetta kertomalla valmistajan mallin nimen		Exif.Image	Model (272)
Digitoinnissa käytetyn laitteen sarjanumero	Tarkentaa mallia ja yksilöi laitteen, jonka avulla analoginen objekti on muunnettu digitaaliseksi		Exif.Image	CameraSerialNumber (50735)
Kuvatiedoston luomisessa käytetty ohjelma	Sovellus ja versio, millä digitaalinen tiedosto on luotu		Exif.Image	Software (305)
Kuvatiedoston luontipäivämäärä ja aika	Ilmaistaan muodossa: YYYY:MM:DD HH:MM:SS		Exif.Image	DateTime (306)
Lukusuunta	Tiedoston lukusuunta (horisontaalinen tai vertikaalinen)		Exif.Image	Orientation (274)

Elementti	Tarkenne	Vaadittu arvo	Metatieto- skeema	Metatietokenttä
Resoluution mittaussyksikkö	Mittayksikkö, jota käytetään tulkitessa X ja Y resoluutiota	2 = inch	Exif.Image	Image.Resolution Unit (296)
XResoluutio	Pikselien määrä resoluution mittayksikköä kohden kuvan leveyssuunnassa.	300	Exif.Image	Image.XResoluutio n (282)
YResoluutio	Pikselien määrä resoluution mittayksikköä kohden kuvan korkeussuunnassa.	300	Exif.Image	Image.YResoluutio n (283)
Kuvatiedoston käsittelyohjelma	Mikäli digitointiprosessissa luodaan ensin pakkaamaton tiedosto, jota käsitellään skannauksen jälkeen, tallennetaan käsittelyohjelman nimi ja versio		Exif.Image	Image.Processing Software (11)

Mahdollisesti toteutetaan tulevaisuudessa, jos siirretään Kansallisarkistoon:

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.

Alla olevassa taulukossa 5 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 5 ilmaistaan vain tietyn arvon sisältäviä kenttiä eli dataelementtejä.

Taulukko 5: 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 objektin 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	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.	P
iccProfileName	Dataelementti, joka määrittää yleisesti käytetyn ICC-profiilin nimen. Vaadittu arvo JPEG-tiedostossa sRGB. Vaadittu arvo TIFF-tiedostossa sRGB, eciRGB v2, ProPhoto RGB tai AdobeRGB (1998)	P
iccProfileVersion	Dataelementti, joka kertoo käytetyn ICC-profiilin version, esimerkiksi v4 [eli sRGB v4]	P
iccProfileURL	Jos ICC-profiili ei ole hyvin dokumentoitu, profiilin URL/URN tallennetaan tähän kenttään.	V
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 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

MIX -kentän nimi	Kentän tarkoitus	Velvoite
xSamplingFrequency	Pikselien määrä resoluution mittayksikkö kohden leveysuunnassa. Vaadittu arvo 300	P
ySamplingFrequency	Pikselien määrä resoluution mittayksikkö kohden pystysuunnassa. Vaadittu arvo 300	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 3	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 yksilöimän mittataulun mittausdata löytyy.	V

Tiedostojen suhde toisiinsa

Prosessissa tuotettujen digitaalisten objektien PITÄÄ kohdata keskenään. Toisin sanoen AltoXML -tiedoston 0001.xml PITÄÄ sisältää kuvatiedostosta 0001.jpg tai 0001.tif tunnistettu tietosisältö. MIX-metatietoskeeman mukaisen 0001.xml-tiedoston PITÄÄ sisältää kuvatiedostoa 0001.jpg tai 0001.tif kuvailevia metatietoja.

Kansallisarkiston vaatiman siirtopaketin muodostaminen on prosessissa vapaaehtoista, koska tällä hetkellä kunnat eivät ole siirtämässä aineistojaan Kansallisarkiston pitkäaikaissäilytykseen. Alla olevassa taulukossa näkyy kuitenkin minkä tyyppisiä tiedostojen tulee olla.

Taulukko 6 Siirtopaketin hakemistojen sisältö

Hakemisto	Selite
identifier	Tarkoittaa digitaalisen ilmentymän yksilöivää tunnusta, jonka avulla PITÄÄ pystyä tunnistamaan, mistä asiakirjakokonaisuudesta on kyse (esimerkiksi arkistoyksikkö). ⁸ Hakemisto sisältää "digitaalisten objektien hakemistot".
master	Hakemistoon PITÄÄ tallentaa taulukoissa (Taulukko 3: Kuvatiedoston (TIFF) pakolliset metatiedot tai Taulukko 4: Kuvatiedoston (JPEG) pakolliset metatiedot) esitetyt kuvatiedostot yksittäisinä tiedostoina. Tiedostot PITÄÄ nimetä nelinumeroisina alkaen 0001.tif tai 0001.jpg
ocr	Hakemistoon PITÄÄ tallentaa luvussa 5.2 esitetty AltoXML tiedosto siten, että jokaisesta kuvatiedostosta on oma XML-tiedostonsa. Tiedostot PITÄÄ nimetä nelinumeroisina alkaen 0001.xml.
mix	Hakemistoon PITÄÄ tallentaa taulukossa 5 esitetyt pakolliset tiedot koskien kaikkia master – hakemiston sisällä olevia kuvatiedostoja. Tiedostoon PITÄÄSI tallentaa myös muut taulukossa esitetyt tiedot. Tiedostoon SAA tallentaa myös muita MIX-metatietoskeeman mukaisia tietoja, skeeman mukaisessa rakenteessa. Tiedostot PITÄÄ nimetä nelinumeroisina alkaen 0001.xml