

## RIKSARKIVETS SPECIFIKATIONER FÖR DIGITALISERING SOM SYFTAR TILL FÖRSTÖRING AV DET ANALOGA UTTRYCKET

<b>Innehåll</b>	Riksarkivets specifikationer för digitalisering som gör det möjligt att förstöra det analoga uttrycket.
<b>Syfte</b>	Att säkerställa att informationsinnehållet i handlingar som ingår i det nationella kulturarvet bevaras och att de kan användas efter att det analoga uttrycket har förstörts. Syftet med detta dokument är att specificera en digitaliseringsprocess som gör det möjligt att förstöra det analoga uttrycket.
<b>Målgrupp</b>	Dokumentet är avsett för Riksarkivet och andra aktörer inom den offentliga förvaltningen som med stöd av arkivlagen (831/1994) avser att förstöra de analoga uttrycken av handlingar som ska förvaras varaktigt efter att handlingarna digitaliserats. Förstöring av analoga handlingar förutsätter Riksarkivets gallringsbeslut.
<b>Avgränsningar</b>	<p>Detta dokument tillämpas när materialhelheter eller arkiverade material som bildats i samband med ärendehantering eller annan verksamhet digitaliseras i efterhand. Specifikationerna gäller inte digitalisering som genomförs i samband med att en handling upprättas eller ett ärende behandlas eller när det gäller omvandling till digital form av analoga handlingar som inkommit till en organisation.</p> <p>Dokumentet beskriver digitaliseringsprocessen och dess resultat. I dokumentet beskrivs inte det egentliga förvaringspaketet som sparas i systemet för långtidsförvaring. Av resultatet av den digitaliseringsprocess som beskrivs i detta dokument är det möjligt att bilda ett paket som sparas i systemet för långtidsförvaring. Detta dokument tar inte ställning till hur det analoga uttrycket ska förstöras, utan det beskriver de kriterier som ska uppfyllas för att det ska vara möjligt att förstöra analoga handlingar.</p>

**Tillämpliga bestämmelser** Arkivlagen (831/1994) 14 a §

**Giltighetstid** Tills vidare, från dokumentets datum

## Innehåll

<b>1</b>	<b>Termer och begrepp</b>	<b>1</b>
<b>2</b>	<b>Inledning</b>	<b>2</b>
<b>3</b>	<b>De allmänna specifikationerna för digitaliseringsprocessen</b>	<b>3</b>
<b>4</b>	<b>Allmänna rekommendationer om digitaliseringsprocessen och god praxis vid digitalisering</b>	<b>4</b>
<b>5</b>	<b>Godkända format</b>	<b>5</b>
5.1	Bildfil	5
5.2	Format för sparande av identifierad text	8
5.3	Metadata och struktur som beskriver bildfilen och behandlingen av bildfilen	8
<b>6</b>	<b>Överföringspaket som bildas genom digitaliseringsprocessen</b>	<b>11</b>
<b>7</b>	<b>Exempelpaket</b>	<b>13</b>
<b>8</b>	<b>Underskrifter</b>	<b>13</b>

## 1 Termer och begrepp

Den terminologi som används i dokumentet grundar sig på en definition [RFC 2119] som sammanställts av Internet Engineering Task Force.<sup>1</sup> Tabell 1 meddelar översättningarna av de termer som används i detta dokument.

Tabell 1: Översättningar som används i detta dokument

ENGELSKA	SVENSKA
MUST	SKA
MUST NOT	FÅR INTE
REQUIRED	OBLIGATORISK
SHOULD	BÖR
SHOULD NOT	BÖR INTE
MAY	FÅR
OPTIONAL	VALFRI

I tabellen nedan (Tabell 2: Begrepp) anges vad som avses med begreppen i tabellen i detta dokument:

Tabell 2: Begrepp

BEGREPP	FÖRKLARING
Analog handling	En handling som upprättats, förvarats och använts på papper eller i något annat konkret format.

<sup>1</sup> <https://www.ietf.org/rfc/rfc2119.txt> Hämtad 6.3.2019

BEGREPP	FÖRKLARING
<b>Analogt uttryck</b>	Det analoga uttrycket av en helhet av analoga handlingar som man beslutat digitalisera. Med analogt uttryck avses i detta dokument en helhet av handlingar som i regel består av pappershandlingar (analoga handlingar) i formatet A4/folio. Helheten kan också innehålla handlingar som är större eller mindre än detta.
<b>Digitalt uttryck</b>	Det digitala uttrycket av ett analogt verk/en helhet av analoga handlingar som man beslutat digitalisera.
<b>Digitalt objekt</b>	En digital fil som antingen ensam eller tillsammans med andra digitala filer bildar en digitaliserad handling. Med digitalt objekt avses i detta dokument en bildfil eller en XML-fil.
<b>Digitaliseringsprocess</b>	Ett antal funktioner med vilka ett analogt uttryck konverteras till ett digitalt uttryck.
<b>Digitaliserad handling</b>	En elektronisk version av en analog handling som producerats genom digitaliseringsprocessen. Kan bestå av n stycken digitala objekt.
<b>Bildfil</b>	En bildfil är en bitmap-bild som producerats genom digitaliseringsprocessen (ett digitalt objekt). Den bildfil som visas i detta dokument kan vara antingen i TIFF-format eller i JPEG-format och fungerar som arkivexemplar (master). <sup>2</sup>
<b>Huvudläsriktning</b>	Gör det möjligt att tolka handlingens informationsinnehåll utan att vända bildfilen. Om det finns informationsinnehåll med flera läsriktningar i handlingen, avses med huvudläsriktningen den riktning i vilken största delen av handlingens informationsinnehåll kan läsas.
<b>Överföringspaket</b>	En helhet som bildas i digitaliseringsprocessen och som kan överföras till Riksarkivets informationssystem.
<b>Produktionsdygn</b>	Dygn under vilket digitala objekt produceras med utrustningen.

## 2 Inledning

Digitalisering som syftar till förstöring av det analoga uttrycket betyder att digitaliseringsprocessen avslutas med att det analoga uttrycket förstörs. Det är inte fråga om att de analoga handlingarnas informationsinnehåll förstörs utan om att informationsinnehållet konverteras till en annan bevarandeform. Förstöring av det analoga uttrycket förutsätter att konverteringen till digitalt format, dvs. digitaliseringen, har utförts med metoder som inte försämrar handlingens bevisvärde, integritet och autenticitet.

Analogt uttryck som är föremål för digitalisering SKA ha Riksarkivets gallringsbeslut som bestämmer om det format i vilket informationen i handlingen ska förvaras. Beslutet fastslår huruvida det analoga uttrycket har ett kulturhistoriskt värde, varför det INTE FÅR förstöras efter digitaliseringen av uttrycket. Allmänt taget BÖR ett gallringsbeslut finnas redan innan digitaliseringen av handlingar inleds. Då är det lättare att planera enligt vilka specifikationer handlingarna digitaliseras. Om ett gallringsbeslut inte finns, FÅR handlingarna INTE förstöras efter digitaliseringen, även om digitaliseringen utförts i enlighet med de specifikationer som meddelas i detta dokument.

<sup>2</sup> FADGI, arkivexemplar: <http://www.digitizationguidelines.gov/term.php?term=productionmasterfile> Hämtad 6.3.2018

De kriterier som meddelas i detta dokument SKA iakttas när en myndighet med stöd av arkivlagen digitaliserar analoga handlingar som ska förvaras varaktigt och vilkas analoga exemplar utifrån Riksarkivets gallringsbeslut ska förstöras efter digitaliseringen. Mottagningen av digitala uttryck i Riksarkivets informationssystem förutsätter att de digitala uttrycken uppfyller de specifikationer som meddelas i detta dokument.

Vid uppgörandet av denna handling har de allmänna standarderna som används inom arkivsektorn och de övriga riksarkivens kvalitetskrav på digitaliseringen beaktats. I kapitlen Godkända format och Överföringspaket som bildas genom digitaliseringsprocessen beaktas dessutom långtidsförvaringstjänsternas (PAS-tjänsterna) nationella specifikationer för material som ska förvaras.<sup>3</sup>

Detta dokument är inriktat på digitalisering av analoga handlingar till bildfiler och på behandling och registrering av innehåll som identifierats i dessa handlingar med olika tekniker. Dokumentet behandlar inte digitaliseringen till exempel av ljud- eller filminspelningar.

### 3 De allmänna specifikationerna för digitaliseringsprocessen

Konverteringen av analoga uttryck till digitala uttryck är en process (digitaliseringsprocess) som SKA dokumenteras på de sätt och med den precision som meddelas i detta dokument. Med dokumentering av processen avses att metadata som dokumenterar skanningen samt eventuell behandling av bilderna sparas. Utöver dessa metadata FÅR även andra åtgärder sparas som metadata för den digitala konverteringsprocessen.

I anslutning till digitaliseringsprocessen SKA man försäkra sig om att den helhet som avses bli digitaliserad blir digitaliserad som en helhet och att det kompletta innehållet blir digitaliserat. Detta innebär i praktiken att alla analoga handlingar i den helhet som det bestämts att ska digitaliseras SKA digitaliseras på ett sådant sätt att ingen information blir okonverterad till digitalt format på grund av ett tekniskt eller funktionellt fel.

Det SKA vara möjligt att vid visuell granskning av varje enskild bildfil som hör till den digitaliserade helheten få samma information som från bildfilens analoga motsvarighet. En bildfil FÅR INTE innehålla några element som inte finns i den analoga motsvarigheten. Ett undantag till detta utgörs av eventuella skanningstekniska måttavlor som skannas/fotograferas till samma bildfil och som verifierar färger, gråtoner, proportioner och upplösning för digitala objekt. Dessa måttavlor SKA placeras så att de inte täcker det objekt som digitaliseras.

Sidor som innehåller anteckningar FÅR INTE avlägsnas i digitaliseringsprocessen. De bildfiler som produceras i digitaliseringsprocessen SKA vara vända i huvudläsriktningen. Bildfiler som producerats genom digitaliseringsprocessen FÅR vändas endast i steg om 90 grader efter att de skannats in.

---

<sup>3</sup>Kansalliset pitkäaikaissäilytyspalvelut -> Määrittelyt -> Säilytys- ja siirtokelpoiset tiedostomuodot  
<http://digitalpreservation.fi/files/PAS-tiedostomuodot-1.6.1.pdf> Hämtad 6.3.2018

Prestationsförmågan hos den infrastruktur som används vid digitaliseringen BÖR optimeras före skanningsåtgärden. Efter optimeringen BÖR kvaliteten på de bildfiler som infrastrukturen producerar verifieras med hjälp av måttavlor som utvecklats för detta ändamål. Kvaliteten BÖR verifieras en gång per produktionsdygn.

#### 4 Allmänna rekommendationer om digitaliseringsprocessen och god praxis vid digitalisering

I detta kapitel beskrivs god praxis i anslutning till skanning och kvalitetssäkringen av skanningen.

Digitaliseringen uppfattas på det allmänna planet som en process som omfattar de faser som presenteras på bilden nedan (Bild 1, Process för konvertering till elektroniskt format – allmän).

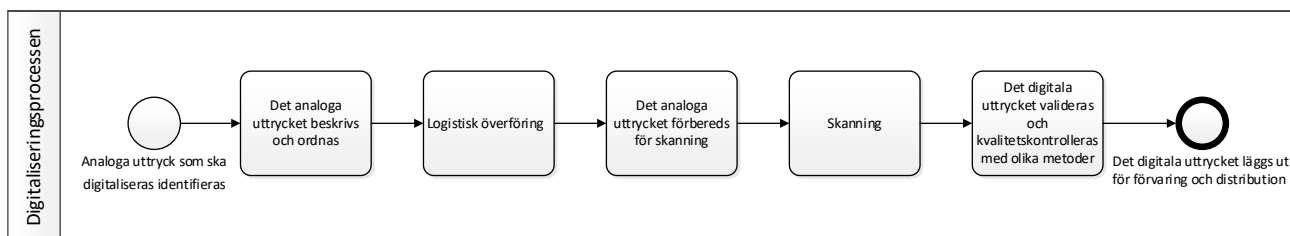


Bild 1, Process för konvertering till elektroniskt format – allmän

Det är bra att betona att det inte är möjligt att inrikta kvalitetssäkringen på en bestämd fas i digitaliseringsprocessen. Digitaliseringens slutresultat är pålitligt då kvalitetssäkring utförs i processens olika faser. I detta dokument ligger fokus särskilt på kvalitetssäkring i anslutning till skanningen.

Innan konverteringen av de handlingar som ska digitaliseras inleds är det bra att beskriva handlingarna i metadatasystemet. Då är det möjligt att dokumentera såväl det analoga uttryckets behandlingskedja mer heltäckande som till exempel information om hur materialet förberetts. Efter skanningen kan metadata om materialet kompletteras antingen för hand eller med automatiska metoder.

Kvalitetssäkringen av skanningen kan grovt taget delas in i åtgärder före skanningsåtgärden och i kvalitetssäkring efter skanningsåtgärden, dvs. validering.

Såsom det meddelas i kapitel 3, bör skanningsinfrastrukturens prestanda optimeras före skanningsåtgärden på så sätt att det digitala uttryck som åtgärden producerat representerar bästa möjliga uttryck som kan produceras med den tekniska sammansättningen i fråga. Efter optimeringen bör infrastrukturens prestanda kontrolleras planmässigt så att kvaliteten på de digitala uttryck som produceras i processen förblir jämn. För kontrollen behövs i allmänhet en måttavla, referensvärden för måttavlan och en analysprogramvara. Utöver bildkvaliteten ska man i utrustningsinfrastrukturen säkerställa att de analoga handlingarna konverteras till digitalt format med komplett informationsinnehåll. Detta innebär bland annat att det vid anskaffning av utrustningen är nödvändigt att fästa särskild vikt vid apparatens förmåga att skilja handlingar från varandra, så att man kan undvika att två handlingar går på varandra genom apparaten (skanner

med genommatning eller med öppen bana och andra skanningslösningar som används för "massvis" skanning av handlingar).

Valideringen efter skanningen kan utföras med urval. Urvalsstorleken beror på skannerns pålitlighet. Ett stort antal allmänna referensvärden och rekommendationer har publicerats. Syftet med valideringen är att säkerställa att de specifikationer som meddelas i kapitel 3 uppfylls. Om materialet är maskinskrivet kan informationsinnehållet i de bildfiler som skapats i digitaliseringsprocessen identifieras med olika metoder (t.ex. med OCR = Optical Character Recognition). Denna fas kan också användas som en indikator för hur väl skanningen lyckats, om det är möjligt att ställa in gränser för hur identifieringen lyckas.

Om bildfilerna behandlas efter skanningen, bör en detaljerad bildhanteringshistorik sparas åtminstone i bildfilernas metadata, om möjligt även i de XML-uppgifter som beskriver det digitala objektets uppkomst.

## 5 Godkända format

Avsnittet om format har indelats i tre underkategorier:

1. Bildfil
2. Informationsinnehåll som identifierats i en bildfil
3. Metadata och struktur som beskriver bildfilen och behandlingen av bildfilen

### 5.1 Bildfil

Den bildfil som produceras i digitaliseringsprocessen SKA sparas som en 24 bitars RGB-bild. Bildfilen FÅR INTE i något skede av behandlingen ha en kvalitet som underskrider kraven i det här kapitlet. Bildfilen ska sparas antingen i förlustfritt komprimerat TIFF-format eller i JPEG-format som medför förlust (inte i båda).<sup>4</sup> Om den bildfil som skapats först i processen är komprimerad, FÅR den INTE behandlas efter den första lagringsgången och därefter sparas på nytt.

I tabellerna Tabell 3 och Tabell 4 nedan anges de obligatoriska uppgifter som ska finnas i maskinläsbart format i bildfilen. I tabell 3 presenteras obligatoriska metadata, om bildfilen sparas i TIFF-format. I Tabell 4 presenteras i sin tur obligatoriska metadata, om bildfilen sparas i JPEG-format. Om kolumnen "Element" inte preciseras ska uppgiften meddelas, men i detta sammanhang har ett obligatoriskt fält inte specificerats för uppgiften. Utöver de uppgifter som meddelas i tabellen FÅR bildfilen innehålla också andra metadatafält.

Tabell 3: Obligatoriska metadata för bildfil (TIFF)

Element	Identifierare	Specificerat värde, om det kan uttryckas entydigt	Schema för metadata	Fält för metadata
Format	TIFF	image/tiff		MIME Type
Version		6.0		
Bildens namn				

<sup>4</sup> JPEG: <https://jpeg.org/jpeg/index.html> Hämtad 6.3.2019

TIFF: <https://www.itu.int/itudoc/itu-t/com16/tiff-fx/docs/tiff6.pdf> Hämtad 6.3.2019

Element	Identifierare	Specifiserat värde, om det kan uttryckas entydigt	Schema för metadata	Fält för metadata
<b>Bildfilens storlek</b>				
<b>Färgrymd</b>	Bildfilens färgrymd	RGB	Exif.Image	PhotometricInterpretation (262)
<b>ICC-profil</b>		sRGB, eciRGB v2, ProPhoto RGB, AdobeRGB (1998)	TIFF tag, private	ICC Profile (34675)
<b>Bitdjup</b>	Antalet bitar i pixelns kanalvärde	8 8 8	Exif.Image	BitsPerSample (258)
	Antalet kanalvärden per pixel	3	Exif.Image	SamplesPerPixel (277)
<b>Komprimering av filen</b>		5 = LZW	Exif.Image	Compression (259)
<b>Bildens bredd</b>	Anger bildens bredd i antal pixlar per rad		Exif.Image	ImageWidth (256)
<b>Bildens höjd</b>	Anger bildens höjd med antalet pixelrader på bilden		Exif.Image	ImageLength (257)
<b>Den digitala bildfilens producent</b>	Den organisation som skapat bildfilen av det analoga uttrycket		Exif.Image	Artist (315)
<b>Digitaliseringsapparat (skanning eller fotografering)</b>	Anger med vilken tillverkarens apparat det analoga uttrycket har konverterats till digitalt format (tillverkarens namn)		Exif.Image	Make (271)
<b>Digitaliseringsapparatens modell (skanning eller fotografering)</b>	Specificerar digitaliseringsapparatens genom att ange namnet på tillverkarens modell		Exif.Image	Model (272)
<b>Serienumret för den apparat som använts vid digitaliseringen</b>	Identifierar den använda apparaten		Exif.Image	CameraSerialNumber (50735)
<b>Program som användes för att skapa bildfilen</b>	Applikation och version med vilken den digitala filen skapades		Exif.Image	Software (305)
<b>Datum och klockslag när bildfilen skapades (skanningsdatum)</b>	Anges i formatet: YYYY:MM:DDTHH:MM:SS		Exif.Image	DateTimeOriginal (36867)
<b>Läsriktning</b>	Läsriktning för filen (liggande eller stående). Läsriktningen tar inte ställning till läsriktningen för informationsinnehållet i bilden, utan den anger i vilken riktning filen är.		Exif.Image	Orientation (274)
<b>Mätenheten för upplösning</b>	Mätenhet för tolkning av X- och Y-upplösningen	2 = inch	Exif.Image	ResolutionUnit (296)
<b>XUpplösning</b>	Antalet pixlar per upplösningens måtenhet i bildens breddriktning.	300	Exif.Image	XResolution (282)



Element	Identifierare	Specificerat värde, om det kan uttryckas entydigt	Schema för metadata	Fält för metadata
<b>Yupplösning</b>	Antalet pixlar per upplösningens måtenhet i bildens höjdriktning.	300	Exif.Image	YResolution (283)
<b>Byteordning</b>		big endian eller little endian		ByteOrder
<b>Program för behandling av bildfilen</b>	Om man i digitaliseringsprocessen först skapar en okomprimerad fil som behandlas efter skanningen, sparas behandlingsprogrammets namn och version.		Exif.Image	Image.ProcessingSoftware (11)

Tabell 4: Obligatoriska metadata för bildfil (JPEG)

Element	Identifierare	Specificerat värde	Schema för metadata	Fält för metadata
<b>Format</b>	JPEG	image/jpeg		MIME Type
<b>Version</b>	JPEG part 1 version	1.00 eller 1.01 eller 1.02		JFIF Version
<b>Bildens namn</b>				
<b>Bildfilens storlek</b>				
<b>Färgrymd</b>	Bildfilens färgrymd	RGB	Exif.Image	PhotometricInterpretation (262)
<b>ICC-profil</b>	Färgprofil som sparats i bildfilens metadata.	sRGB	ICC	profileDescription
<b>Bitdjup</b>	Antalet bitar i pixelns kanälvärde	8 8 8	Exif.Image	BitsPerSample (258)
	Antalet kanälvärden per pixel	3	Exif.Image	SamplesPerPixel (277)
<b>JPEG-kvalitet</b>	JPEG-komprimeringens kvalitet på skalan 0–100%	90%		
<b>Bildfilens producent</b>	Den organisation som skapat bildfilen av det analoga uttrycket		Exif.Image	Artist (315)
<b>Bildens höjd</b>	Anger bildens höjd med antalet pixelrader på bilden		Exif.Image	ImageLength(257)
<b>Bildens bredd</b>	Anger bildens bredd med antalet pixelrader på bilden		Exif.Image	ImageWidth(256)
<b>Digitaliseringsapparat</b>	Anger med vilken tillverkarens apparat det analoga uttrycket har konverterats till digitalt format (tillverkarens namn)		Exif.Image	Make (271)
<b>Modell på digitaliseringsapparaten</b>	Specificerar digitaliseringsapparaten genom att ange namnet på tillverkarens modell		Exif.Image	Model (272)
<b>Serienumret för den apparat som använts vid digitaliseringen</b>	Specificerar modellen och individualiserar apparaten med		Exif.Image	CameraSerialNumber (50735)



Element	Identifierare	Specificerat värde	Schema för metadata	Fält för metadata
	vilken det analoga objektet har konverterats till digitalt format			
<b>Program som användes för att skapa bildfilen</b>	Applikation och version med vilken den digitala filen skapades		Exif.Image	Software (305)
<b>Datum och klockslag när bildfilen skapades</b>	Anges i formatet: YYYY:MM:DD HH:MM:SS		Exif.Image	DateTime (306)
<b>Läsriktning</b>	Filens läsriktning (horisontell eller vertikal)		Exif.Image	Orientation (274)
<b>Mätenheten för upplösning</b>	Mätenhet för tolkning av X- och Y-upplösningen	2 = inch	Exif.Image	Image.Resolution Unit (296)
<b>XUpplösning</b>	Antalet pixlar per upplösningens måtenhet i bildens breddriktning	300	Exif.Image	Image.XResolution (282)
<b>YUpplösning</b>	Antalet pixlar per upplösningens måtenhet i bildens höjdriktning	300	Exif.Image	Image.YResolution (283)
<b>Program för behandling av bildfilen</b>	Om man i digitaliseringsprocessen först skapar en okomprimerad fil som behandlas efter skanningen, sparas behandlingsprogrammets namn och version.		Exif.Image	Image.Processing Software (11)

## 5.2 Format för sparande av identifierad text

Om text i bildfilerna identifieras till exempel med metoderna OCR (identifiering av maskinskriven text) eller HTR (identifiering av handskriften text), SKA den sparas i formatet Analyzed Layout and Text Object (ALTO).<sup>5</sup> De godtagbara ALTO-versionerna är 3.0 eller senare. En ALTO-fil SKA sparas för varje bildfil, om informationsinnehållet i bildfilen i regel är maskinskrivet.<sup>6</sup>

## 5.3 Metadata och struktur som beskriver bildfilen och behandlingen av bildfilen

De metadata som beskrivs i detta kapitel beskriver bildfilens tillkomsthistoria, som för sin del också verifierar autenticiteten för det digitala uttryck som uppkommit i processen. Bildfilernas obligatoriska tekniska metadata SKA presenteras i enlighet med metadataformatet MIX version 2.0.<sup>7</sup>

I tabellen nedan (Tabell 5) anges MIX-fältets namn, syftet med fältet i fri översättning samt skyldigheten. Uppgifterna presenteras från vänster till höger. I fältet för skyldighet meddelas huruvida fältet och uppgiften i enlighet med schemat är obligatoriska på följande sätt:

- O = Obligatorisk – denna uppgift SKA beskrivas
- F = Frivillig – denna uppgift BÖR beskrivas, men det är inte obligatoriskt

<sup>5</sup> The Library of Congress » Standards » ALTO. Webbplatsen för kongressens bibliotek <https://www.loc.gov/standards/alto/> Hämtad 6.3.2019

<sup>6</sup> Ett dokument som mestadels är maskinskrivet innehåller både maskinskriven och handskriften text, men mestadels maskinskriven text. Dessutom kan dokumentet innehålla bilder eller annat innehåll.

<sup>7</sup> The Library of Congress » Standards » MIX. Webbplatsen för kongressens bibliotek <http://www.loc.gov/standards/mix/> Hämtad 6.3.2019

I metadataschemat MIX finns fält av två slag: behållare och dataelement. Dataelementen innehåller ett visst värde, medan behållarna innehåller ett eller flera dataelement och kan innehålla andra behållare jämte dataelement. I Tabell 5 meddelas endast fält, dvs. dataelement, som innehåller ett visst värde.

Tabell 5: Metadata som beskriver bildfilen och dess behandling (i tabellen meddelas endast fält som innehåller information som SKA presenteras i strukturen enligt metadataformatet MIX version 2.0)

MIX-fältets namn	Syftet med fältet	Skyldighet
objectIdentifierType	Dataelement som anger det system eller den domän där den digitala handlingens specificerande ID är unikt.	O
objectIdentifierValue	Serie av tecken som identifierar det digitala objektet.	O
fileSize	Filens storlek i byte, t.ex. 72839.	O
formatName	Filens format. Det specificerade värdet är image/jpeg eller image/tiff	O
formatVersion	Filens version, t.ex. 1.01	F
byteOrder	Dataelement som definierar ordningsföljden då byte sparas. Det specificerade värdet är antingen big endian eller little endian.	O
compressionScheme	Den använda komprimeringen. Det specificerade värdet är JPEG (när arkivexemplaret är i jpeg-format) eller LZW (när arkivexemplaret är i TIFF-format).	O
compressionRatio	Dataelement som meddelar den använda komprimeringens nivå. Uppges med numerator "90" och denominator "100".	O (endast JPEG)
messageDigestAlgorithm	Dataelement som identifierar den algoritm med vilken värdet i fältet messageDigest har bildats. Fältet ska ha något av de följande värdena: MD5, SHA-1, SHA256, SHA384, SHA512	O
messageDigest	Den teckenserie som den algoritm som definierar fältet messageDigestAlgorithm producerat, t.ex. e8064dc0	O
imageWidth	Bildens bredd i pixlar, t.ex. 1330.	O
imageHeight	Bildens höjd i pixlar, t.ex. 1600.	O
colorSpace	Dataelement som definierar bildens färgrymd. Det specificerade värdet är RGB.	O
iccProfileName	Dataelement som definierar den allmänt använda ICC-profilens namn. Specificerat värde i JPEG-filen är sRGB. Specificerat värde i TIFF-filen är sRGB, eciRGB v2, ProPhoto RGB eller AdobeRGB (1998)	O
iccProfileVersion	Dataelement som anger den använda ICC-profilens version, t.ex. v4 [t.ex. sRGB v4]	O
iccProfileURL	Om ICC-profilen inte är väl dokumenterad ska profilens URL/URN sparas i detta fält.	F
dateTimeCreated	Dataelement som anger tidpunkten då bildfilen skapades. Anges i formatet: YYYY-MM-DDTHH:MM:SS	O
imageProducer	Dataelement som identifierar den organisation som skapade bildfilen.	O
scannerManufacturer	Dataelement som anger namnet på tillverkaren av den apparat som användes vid skanningen.	O
scannerModelName	Dataelement som anger namnet på modellen av den apparat som användes vid skanningen.	O
scannerModelNumber	Dataelement som preciserar namnet på digitaliseringsapparatens modell med dess typnummer.	O
scannerModelSerialNo	Digitaliseringsapparatens serienummer med vilken apparaten kan individualiseras.	O
scanningSoftwareName	Namnet på den skanningsprogramvara som använts.	O
scanningSoftwareVersionNo	Versionsnumret för den skanningsprogramvara som använts.	O
orientation	Dataelement som anger bildens läsriktning.	O

MIX-fältets namn	Syftet med fältet	Skyldighet
samplingFrequencyUnit	Dataelement som anger den måttenhet som används för att tolka X- och Y-upplösningen. Det specificerade värdet är "in."	O
xSamplingFrequency	Antalet pixlar per upplösningens måttenhet i bildens breddriktning. Det specificerade värdet är 300.	O
ySamplingFrequency	Antalet pixlar per upplösningens måttenhet i bildens höjdriktning. Det specificerade värdet är 300.	O
bitsPerSampleValue	Dataelement som definierar antalet bitar i varje kanal. Det specificerade värdet är 8.	O
bitsPerSampleUnit	Dataelement som definierar hur bitarna tolkas. Värdet är antingen integer eller floating point.	O
samplesPerPixel	Dataelement som definierar antalet kanälvärden per pixel. Det specificerade värdet är 3.	O
targetType	Dataelement som anger huruvida måttavlan som verifierar skanningens kvalitet är en del av bilden eller om den skannats i en separat bild.	F
targetManufacturer	Dataelement där måttavlans tillverkare antecknas.	F
targetName	Dataelement som identifierar den använda måttavlans namn.	F
targetNo	Dataelement som innehåller den använda måttavlans serienummer.	F
externalTarget	Dataelement som anger var den digitala bilden på TargetID-behållarens specificerade måttavla finns.	F
performanceData	Dataelement som anger var mätdatan om TargetID-behållarens specificerade måttavla finns.	F

## 6 Överföringspaket som bildas genom digitaliseringsprocessen

De i kapitel 5 och dess avsnitt nämnda olika digitala objekt som produceras i digitaliseringsprocessen SKA sparas i en sådan katalogstruktur som beskrivs på bilden (Bild 2, Den överföringspaketstruktur som krävs i digitaliseringsprocessen), så att de kan tas emot i Riksarkivet. Ett digitalt uttryck SKA produceras i katalogstrukturen oberoende av när det överförs till Riksarkivet. När materialet överlämnas FÅR det INTE innehålla något annat material än de mappar som anges i Bild 2, Den överföringspaketstruktur som krävs i digitaliseringsprocessen. Om materialet inte kommer att överföras till Riksarkivet i något skede är iakttagandet av katalogstrukturen VALFRITT. Utöver den katalogstruktur som definieras här FÅR organisationen spara till exempel bruksexemplar i sina egna datasystem i den datastruktur som det aktuella systemet förutsätter. Den struktur som fastställs i denna handling utesluter således inte användningen av eventuella andra registreringsstrukturer.

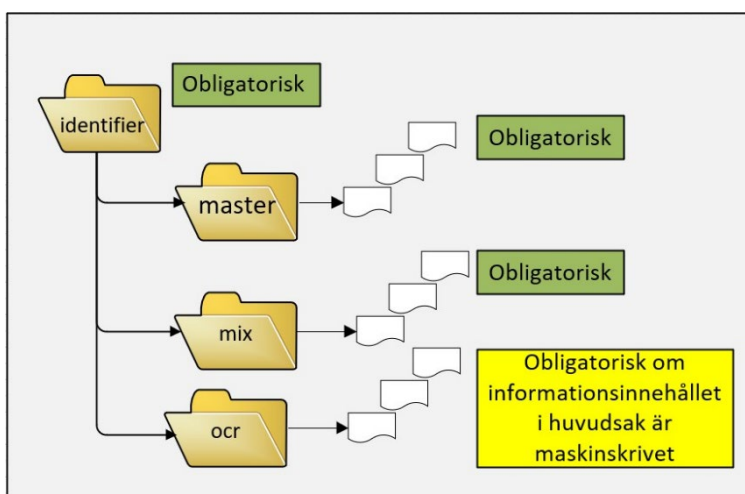


Bild 2, Den överföringspaketstruktur som krävs i digitaliseringsprocessen

Tabell 6 beskriver hur digitala objekt SKA namnges i den katalogstruktur som visas i bild 2. De digitala objekt som produceras i processen SKA möta varandra. Det innebär att AltoXML-filen 0001.xml SKA innehålla det informationsinnehåll som identifierats i bildfilen 0001.jpg eller 0001.tif. Filen 0001.xml som är förenlig med metadataschemat MIX SKA innehålla metadata som beskriver bildfilen 0001.jpg eller 0001.tif.

Tabell 6 Innehållet i katalogerna för överföringspaket

Katalog	Förklaring
identifier	Avser en beteckning som individualiserar ett digitalt uttryck som SKA göra det möjligt att identifiera den aktuella handlingshelheten (till exempel en arkivenhet). <sup>8</sup> Katalogen innehåller "de digitala objektens kataloger".
master	De bildfiler som beskrivs i tabellerna (Tabell 3: Obligatoriska metadataschema för bildfil (TIFF) <b>eller</b> Tabell 4: Obligatoriska metadataschema för bildfil (JPEG)) SKA sparas i katalogen som separata filer. Filerna SKA namnges med fyra siffror med början från 0001.tif eller 0001.jpg.
ocr	Den AltoXML-fil som beskrivs i kapitel 5.2 SKA sparas i katalogen på så sätt att det finns en separat XML-fil för varje bildfil. Filerna SKA ges ett namn med fyra siffror med början i 0001.xml.
mix	De obligatoriska uppgifter som meddelas Tabell 5 och som avser alla bildfiler som ingår i master-katalogen SKA sparas i katalogen. Även de övriga uppgifterna som meddelas i tabellen BÖR sparas i filen. Det ÄR TILLÅTET att spara också andra uppgifter i enlighet med metadataschemat MIX i filen i en struktur som är förenlig med schemat. Filerna SKA ges ett namn med fyra siffror med början i 0001.xml

Om materialet skickas till Riksarkivet SKA varje överföringspaket paketeras i formatet TAR. TAR-paketets innehåll FÅR INTE komprimeras i detta skede. En kontrollsumma i MD5-format SKA beräknas för TAR-paketet och uppgiften om summan lämnas i anslutning till överföringen. När material levereras till Riksarkivet avses med identifier-katalogen materialbeteckningen Ai01 för material i AHAA-systemet (i rollen ahaa teknisk).

<sup>8</sup> Ett analogt uttryck som bestämts att ska digitaliseras bör vara beskrivet (beskrivande metadata har producerats) innan det digitaliseras. Det SKA vara möjligt att med hjälp av identifier koppla de digitala uttryck som skapats genom digitaliseringsprocessen till de metadata som beskrivs ovan.

## 7 Exempelpaket

Det finns två exempel på överföringspaket:

1. Exempelpaket\_JPEG.zip
2. Exempelpaket\_TIF.zip

Bildfilerna i master-katalogerna är inte referenser avseende bildkvaliteten. I filerna finns de metadata som definierats som obligatoriska i detta dokument. Filerna i ocr-katalogerna är exempel på att en AltoXML SKA skapas för varje fil. Filerna i mix-katalogerna är exempel på master-filer i bilagepaket, fränsett element där något annat konstateras. Katalogerna har inte paketerats i TAR-paketet.

## 8 Underskrifter

Generaldirektörens ställföreträdare  
Forskningschef

Päivi Happonen

Utvecklingschef

Mikko Eräkaski