

THE NATIONAL ARCHIVES' SPECIFICATIONS FOR THE DIGITISATION OF SPECIAL ARCHIVAL MATERIALS AND COLLECTIONS

Content	The National Archives' specifications for the digitisation of special archival materials and collections
Purpose	These instructions define the digitisation process and the result that ensure the preservation and usability of the information content of special archival materials and collections in a digital format.
Target group	Instructions are meant for the National Archives of Finland as well as operators who digitise special analogue material pursuant to the Archives Act (831/1994).
Limitations	The instructions do not apply to scanning that is performed when the record is created or processed in transactions and work processes, or when incoming analogue records are converted into a digital format. The special materials consist of map collections, drawing collections, photograph collections, negatives, film slides and microfilms. Also bound volumes where the information content is presented in double-page layout are categorised as special archival material (Table 1). The instructions do not apply to individual records that can be classified as special material if they have been archived together with records of A4 and foil size.

Table 1. Applying the instructions for different types of records.

Material	The National Archives' specifications for the digitisation of special archival materials and collections (2021)	Requirements set by the National Archives of Finland for digitisation process enabling the disposal of the analogue manifestation after it has been converted into digital format (2019)
Map collection	x	
Drawing collection	x	
Photograph collections	x	
Glass and other negatives and slides	x	
Microfilms	x	
Bound volumes where the information content is presented in double-page layout	x	
Bound volumes		x



INSTRUCTIONS

19 January 2021

KA/15780/07.01.02.04.02/2020

Records of A4 and foil size		x
Individual photographs, drawings, and maps when they are archived among A4 and foil size records.		x

A digitisation process conducted following these instructions does not result in a long-term Archival Information Packet (AIP) to be stored in the long-term preservation system. However, an AIP can be created when material has been digitised in accordance with these instructions. The instructions do not comment on the destruction of the analogue manifestation, but rather describe the criteria that must be met before disposing of the analogue record.

The instructions are based on

The Act on the National Archives (1145/2016), section 2.

Validity

Indefinitely from the instructions' date.

Repeals the instructions and recommendation

The National Archives' technical requirements for digitised material 24 August 2016 and 15 August 2008, AL/11130/07.01.02.04.02/2008.

Table of contents

1	Terms and concepts	1
2	Introduction	3
3	General recommendations and good practices for the digitisation process	4
4	General specifications of digitisation processes	5
5	Material type-specific specifications	6
5.1	Map and drawing collections	6
5.2	Bound volumes where the information content is presented in double-page layout.....	7
5.3	Photograph collections	8
5.4	Glass negatives, other negatives, and slides	9
5.5	Microfilms.....	9
6	Accepted formats	10
6.1	Image file	10
6.2	Metadata and structure describing an image file and its processing	14
7	Transfer package generated in the digitisation process.....	16

19 January 2021

KA/15780/07.01.02.04.02/2020

1 Terms and concepts

The terms used in this record are based on the specification [RFC 2119] prepared by the Internet Engineering Task Force.¹ The table below (Table 2) describes terms used in this record.

Table 2. Terms used in this record

TERM	DESCRIPTION
must	This word, or the terms "REQUIRED" or "SHALL", mean that the definition is an absolute requirement of the specification.
must not	This phrase, or the phrase "SHALL NOT", mean that the definition is an absolute prohibition of the specification.
required	This word means that the definition is an absolute requirement of the specification.
should	This word, or the adjective "RECOMMENDED", mean that there may exist valid reasons in particular circumstances to ignore a particular item, but the full implications must be understood and carefully weighed before choosing a different course.
should not	This phrase, or the phrase "NOT RECOMMENDED" mean that there may exist valid reasons in particular circumstances when the particular behaviour is acceptable or even useful, but the full implications should be understood, and the case carefully weighed before implementing any behaviour described with this label.
may	This word, or the adjective "OPTIONAL", mean that an item is truly optional. One vendor may choose to include the item because a particular marketplace requires it or because the vendor feels that it enhances the product while another vendor may omit the same item. An implementation which does not include a particular option MUST be prepared to interoperate with another implementation which does include the option, though perhaps with reduced functionality. In the same vein an implementation which does include a particular option MUST be prepared to interoperate with another implementation which does not include the option (except, of course, for the feature the option provides.)
optional	This word means that an item is truly optional.

The table below (Table 3) presents the terms used in the record and their definitions:

Table 3. Concepts

CONCEPT	DESCRIPTION
Analogue object	An analogue version of an analogue record entity to be digitised.

¹ <https://www.ietf.org/rfc/rfc2119.txt> Retrieved 11 January 2021.



19 January 2021

KA/15780/07.01.02.04.02/2020

CONCEPT	DESCRIPTION
Analogue record	Archived record created, stored, and used in paper format or some other physical format.
Archiving	Archiving is done after the end of the storage period (Act on Information Management in Public Administration 906/2019, section 21). Archiving also means permanent preservation pursuant to the Archives Act (section 8, subsection 3).
Archiving order	Archiving order determines the order of analogue records within a single series.
Decision on appraisal and disposal	Decision of the National Archives of Finland on the records to be permanently preserved (archived) and their preservation format under sections 8, 11 and 14a of the Archives Act.
Derivative file²	A file that is offered for use in various user interfaces. Compare with production master file.
Destruction	Physical (including technical) destruction of a temporary record as well as the destruction of an archived record's analogue format.
Digital object	A digital version of an analogue record entity to be digitised.
Digitisation	Converting records stored or archived in an analogue format to digital format. A larger group of functions where one phase is scanning.
Digitisation device	A device used to convert an analogue record into digital format.
Digitisation process	A group of functions with which an analogue object is converted into digital format.
Digitised record	A digital version of an analogue record produced using a digitisation process that can consist of <i>n</i> number of files.
Field of view	Information on an image file shown by means of a visual inspection.
File	Forms a digitised record either on its own or with other files. In these instructions, a file is either a bitmap image or an XML file.
Image file	A bitmap image produced in a digitisation process. Specifies the format of a digital file as a bitmap image.
Main orientation	The main orientation allows the information content of a record to be interpreted without turning an image file. If a record includes information content in several orientations, the main orientation is the orientation in which most of the information content can be read.
Permanent preservation	<p>In the processing of personal data, permanent preservation means that data is stored for its original purpose in such a way that it is not actually destroyed or transferred to the archive, and there is a permanent justification to preserve the data as they determine benefits, rights, or obligations. (Act on Information Management in Public Administration, government proposal 284/2018, p. 98).</p> <p>The term “permanent preservation” used in the Archives Act means archiving in accordance with the General Data Protection Regulation (EU) 2016/679 (GDPR) and the Act on Information Management in Public Administration (906/2019).</p>

² FADGI, derivative file: <http://www.guidelines.gov/term.php?term=derivativefile> Retrieved 11 January 2021.



19 January 2021

KA/15780/07.01.02.04.02/2020

CONCEPT	DESCRIPTION
Production day	A day during which digital files are produced using equipment.
Production master file³	A file produced in the digitisation process (bitmap image), which is an image file of the highest quality in terms of its technical properties produced in the digitisation process. A production master file is generally preserved in the long-term preservation systems.
Record	<p>Record is information created, received, and maintained as evidence and as an asset by an organization or person, in pursuit of legal obligations or in the functions, activities, transactions and work processes. A record is an entity consisting of content, structure/layout and context. Metadata are included in or related to the record's information. (ISO15489-1: Information and documentation. Records management. Part 1. Concepts and principles).</p> <p>In Finnish legislation, a record is defined as a written or visual presentation, and also as a message relating to a given topic or subject-matter and consisting of signs which, by virtue of the use to which they are put, are meant to be taken as a whole, but are decipherable only by means of a computer, an audio or video recorder or some other technical device. (Act on the Openness of Government Activities 621/1999, section 5).</p>
Scanning	In these instructions, scanning is a part of the digitisation process. Scanning can be performed either using a camera or a scanner. For the sake of readability, the term "scanning" is used in these instructions.
Special archival material and collections	Material entities/collections consisting of archived maps, drawings, photographs, negatives, microfilms, or microfiche. Also bound volumes where the information content is presented in double-page layout are categorised as special archival material.
The cultural-historical value of an analogue format	The archived analogue record's feature, cultural-historical value, that justifies preserving the record also in its analogue format despite the digitisation.
Transfer package	An entity generated in the digitisation process that is transferred to the National Archives' data systems.

2 Introduction

Digitisation contributes to digitalisation and the preservation and usability of cultural heritage resources. The definitions described in these instructions apply to the digitisation process of special archival material and collections. These specifications **must** be followed when public sector agencies digitise special analogue archival material. These instructions **must** also be followed when digitising special archival materials that have cultural-historical value as analogue objects.

³ FADGI, production master file: <http://www.guidelines.gov/term.php?term=productionmasterfile> Retrieved 11 January 2021.

19 January 2021

KA/15780/07.01.02.04.02/2020

Records subject to digitisation **must** have the National Archives' decision on appraisal and disposal already before digitisation begins. The decision on appraisal and disposal determines the cultural-historical value of the analogue records and the preservation format after digitisation. It determines whether the analogue object has cultural-historical value that prevents its destruction after the digitisation. If there is no cultural-historical value, and if the quality, preservation, and usability of the digital objects has been ensured, the original analogue records may be disposed of. The analogue records **must not** be disposed of before they have a decision on appraisal and disposal.

Archiving the information content of analogue records solely in digital form requires that the digital conversion has been carried out with methods that do not weaken record's value as evidence or its integrity and authenticity. Compliance with these instructions, for its part, ensures the preservation of the digitised analogue records' information content solely in a digital format.

Adding digital objects in the National Archives' data systems requires that they fulfil the specifications presented in these instructions. Standards used generally in the archives sector and the quality requirements set by other national archives for digitisation have been taken into account in the preparation of the instructions. In addition, chapters 6 Accepted formats and 7 Packages generated in the digitisation process have taken into account the specifications for preserved material set by the national long-term preservation services.⁴

3 General recommendations and good practices for the digitisation process

This chapter describes the general good practices. The terms listed in Table 2 are not used in this chapter.

Digitisation is generally understood as a process that includes the steps shown in Figure 1 below.

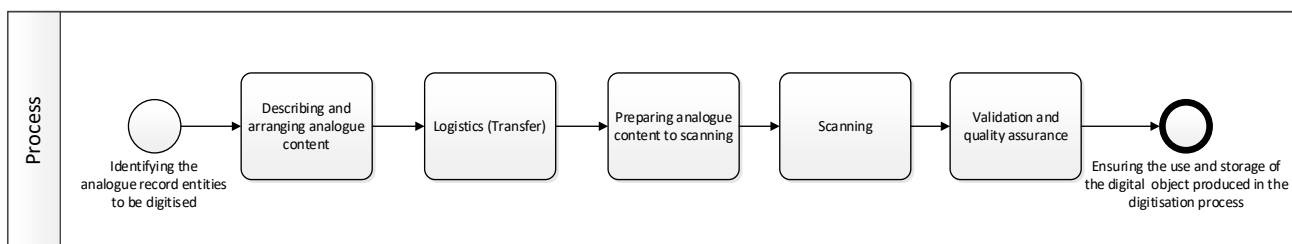


Figure 1. Digital format conversion process - general.

Quality assurance cannot be allocated to a specific stage in the digitisation process; instead, quality assurance must be carried out at different stages of the process. These instructions focus particularly on the quality assurance of scanning.

Records should be described in the system that is used for the management of metadata before digitisation. This ensures that the processing chain of the analogue object and, for example, information concerning the material's preparation can be documented more extensively. After scanning, metadata may be enriched either manually or automatically.

⁴ <http://digitalpreservation.fi/files/PAS-tiedostomuodot-1.8.0.pdf> Retrieved 11 January 2021.

19 January 2021

KA/15780/07.01.02.04.02/2020

Scanning quality assurance can be roughly divided into actions taken before scanning and quality assurance after scanning, i.e. validation, in which the material is inspected visually to detect any quality deviations.

The performance of the scanning infrastructure is optimized before scanning to ensure that the digital object it produces is of the best quality that can be produced with the technical equipment in question. After optimization, the performance of the infrastructure should be monitored as planned so that the quality of digital objects produced during the process remains the same. For monitoring purposes, a test target, reference values for the test target and analysis software are needed. In addition to image quality, the equipment infrastructure should convert analogue records into digital format so that they retain their full information content. For example, the scanning process must prevent two analogue records passing through the device together one on top of the other (feed-through scanners, open-track scanners and other scanning solutions with which records are scanned in large numbers). Furthermore, the image file produced by the infrastructure must not be distorted or otherwise changed in proportions, which would compromise construing the information content.

Validation after scanning can be carried out using samples. The sample size depends on the reliability of the scanning process. The goal of validation is to ensure that the specifications presented in these instructions are fulfilled.

If image files are processed after scanning, the processing history should be saved in detail at least in the image files' metadata and, when possible, also in the XML (MIX) data describing the image file's creation.

It should also be noted that every migration and file conversion (change of a file's saving format) risks data integrity. For this reason, it is important to digitise analogue data using the formats presented in these instructions. A lower-quality saving format cannot be reprocessed into a higher quality version at a later point. The specifications of these instructions also prohibit it.

4 General specifications of digitisation processes

The stages of converting analogue objects into a digital format (digitisation process) **must** be documented in the manner and accuracy described in these instructions. Documentation of a process means saving metadata concerning the scanning and possible image processing. Other actions during the process **may** also be saved as metadata.

The digitisation process **must** ensure that the material to be digitised is digitised as a whole, according to archiving order, and with its full content. This means that all analogue records that belong to the digitised entity **must** be digitised in such a way that no information is left outside due to technical or functional errors.



19 January 2021

KA/15780/07.01.02.04.02/2020

Every single image file related to specific material must contain the same information as its analogue counterpart. This **must** be verifiable through visual inspection. The image file **must not** include any elements that are not included in the analogue object. As an exception, the image file may contain scanned or pictured technical targets for colours, greyscale, dimensions, and resolution of the image file. These targets **must** be positioned so that they do not cover the digitised object.

Pages containing markings **must not** be removed in the digitisation process. Image files produced during a digitisation process **must** be rotated in the main orientation. Files produced during the digitisation process **may** only be rotated by 90-degree angles after scanning.

5 Material type-specific specifications

This chapter describes the material type-specific specifications for digitisation processes and production master files produced in the process. The general definitions of digital files are presented in chapter 6. The material types are described in Table 1. The types of data have been merged in chapter 5 if the specifications for the data types are similar.

5.1 Map and drawing collections

Historical maps and drawings **must** be digitised into image files so that the generated image file is physically ($n \text{ cm} \times n \text{ cm}$) on a scale of 1:1 with the analogue record at a resolution of 300 ppi. To ensure the integrity of the digitised analogue record's information content, the record **must** be digitised so that the field of view exceeds the edges of the pages. The record must lie flat when scanned. If a glass or other tool is used to achieve this, it **must** be taken into account in colour management. Maps and drawings are digitised into colour images (RGB, 24 bit).

Before scanning, the performance of the infrastructure used for digitisation **must** be optimized using technical targets and software suitable for this purpose. After optimization, the quality of the image files produced by the infrastructure **should** be verified once during the production day by using the technical targets intended for that purpose.

The table below (Table 4) presents the specifications for the digitisation of map and drawing collections.

Table 4. Map and drawing

Element	Required value
Archival master file's format	TIFF 6.0
Compression of production master file	Uncompressed or LZW
Resolution	300 ppi
Colour model	2 = RGB
Bit depth	8 bits per sample, 3 samples (24-bit)
ICC profiling of the production master file	eciRGB v2, ProPhoto RGB, AdobeRGB (1998)

19 January 2021

KA/15780/07.01.02.04.02/2020

5.2 Bound volumes where the information content is presented in double-page layout

If the information content of a bound volume is presented in double-page layout, the requirements set out in this section **must** be applied to its digitisation. Such content may be, for example, tables exceeding the page or other information content originally created in the analogue record so that individual pages of the analogue record must be viewed side by side (Figure 2) for the sake of comprehensibility. To ensure the information content's integrity, the material **must** be digitised in full so that the field of view extends over the edges of the pages.

Spread

Page					Page				
<i>Tuomien ja Poisien</i>					<i>kuukausi vuosi 1887</i>				
					8				
Perustustilaja	Kirkkokuva	Pöytäkirja	Kytkä	Tilaa	Tilaa nimeltä	Perustustilustajan nimi	Perustustilustajien		
No. ja päivämäärä				Ennen, Tähden, Muuttanut			Koko perustustilustajien	Koko perustustilustajien	% osuudesta
							luku	luku	luku
<i>1887</i>									
<i>1. Kallio, Anttoni</i>	<i>Siedon</i>	<i>Ingvisten</i>	<i>2. 1. 18</i>		<i>Ingvisten</i>	<i>Talokas Kaarle Anttoni</i>	<i>1887</i>		
						<i>Ingvisten</i>			
<i>2. Kallio, Anttoni</i>	<i>Seuranta</i>	<i>Talokas</i>	<i>3. 31. 87</i>		<i>Ingvisten</i>	<i>Ingvisten</i>			
						<i>Ingvisten</i>			

Figure 2. Information content is presented on a double-page spread

⁵ The image file is not a qualitative reference, but an example of data content spread out on a double page.



19 January 2021

KA/15780/07.01.02.04.02/2020

The table below (Table 5) presents the specifications for the double-paged digitisation of bound volumes.

Table 5. Bound volume specifications

Element	Required value
Production master file's format	TIFF 6.0 or JPEG part 1 version
Compression of production master file	TIFF = LZW or JPEG = JPEG quality level 90%
Resolution	300 ppi
Colour model	2 = RGB
Bit depth	8 bits per sample, 3 samples (24-bit)
ICC profiling of the production master file	eciRGB v2, ProPhoto RGB, AdobeRGB (1998), sRGB

5.3 Photograph collections

In order to ensure the integrity of the information content of photographs, analogue records **must** be fully digitised so that the field of view exceeds the edges of the pages. Photographs **must** be digitised in colour (RGB, 24 bit). The table below (Table 6) presents specifications for the digitisation of photo collections.

Table 6. Photo collection specifications

Element	Required value
Production master file's format	TIFF 6.0
Compression of production master file	LZW
Resolution	According to the analogue object, see Table 7 Resolution table
Colour model	2 = RGB
Bit depth	RGB = 8 bits per sample, 3 samples (24-bit)
ICC profiling of the production master file	RGB: eciRGB v2, ProPhoto RGB, AdobeRGB (1998)

When the analogue record to be digitised is small, the resolution **must** be raised in accordance with the table below (Table 7). If the analogue record's physical dimensions are smaller than those shown in Table 7, the resolution used in its digitisation **must** be increased by applying the same calculation formula. The resolution of the production master file **must not** be changed when processing the image. This is to ensure that the physical dimensions of the analogue record can be assessed based on the digital production master file. The target size of a production master file is approximately A3 (cm x cm) with a resolution of 300 ppi.

Table 7. Resolution table

Analogue record size	Resolution
6 cm × 8 cm	1,600 ppi

19 January 2021

KA/15780/07.01.02.04.02/2020

9 cm × 12 cm	1,200 ppi
13 cm × 18 cm	900 ppi
18 cm × 24 cm	600 ppi
A3 (or larger)	300 ppi

5.4 Glass negatives, other negatives, and slides

To ensure the information content integrity of glass negatives, other negatives and slides, analogue records **must** be fully digitised so that the field of view exceeds the edges of the analogue negative or slide being scanned. If the negative or slide has been set in frames, extreme care **must** be observed to prevent cropping any information content from the image file. Slides and negatives **must be** digitised in colour (RGB, 24 bit). The table below (Table 8) presents definitions of the digitisation of glass negatives, other negatives, and slides.

Table 8. Specifications for glass negatives, other negatives, and slides

Element	Required value
Production master file's format	TIFF 6.0
Compression of production master file	LZW
Resolution	According to the analogue object, see Table 7 Resolution table
Colour model	2 = RGB
steep Bit depth	RGB = 8 bits per sample, 3 samples (24-bit)
ICC profiling of the production master file	RGB: eciRGB v2, ProPhoto RGB, AdobeRGB (1998)

When the analogue record to be digitised is small, the resolution **must** be raised in accordance with the resolution table (Table 7). If the analogue record's physical dimensions are smaller than those shown in Table 7, the resolution used in its digitisation must be increased by applying the same calculation formula. The resolution of the production master file **must not** be changed when processing the image. This is to ensure that the physical dimensions of the analogue record can be assessed based on the digital production master file. The target size of a production master file is approximately A3 (cm x cm) with a resolution of 300 ppi.

5.5 Microfilms

The reduction factor of the microfilm **must** be checked at the start of the roll of film. If no reduction factor has been stated, the magnification factor **must** be set sufficiently large so that the digitised record's information content can be read. Using magnification factors, material on microfilms can be reproduced to be sufficiently readable at a resolution of 300 ppi. The resolution used in scanning is the magnification factor × the desired resolution (e.g., 12 × 300 = 3,600). Microfilm is a steep format. In order to comply with the specifications presented in chapter 4, you must be cautious not to increase the contrast when digitising.

During microfilming other methods (for example, a measuring tape) may have also been used to verify the size of the original material. These methods **must** be examined on a case-by-case basis.



19 January 2021

KA/15780/07.01.02.04.02/2020

Even if none of the above methods were applied, the physical dimensions of the original analogue record must be assessed.

The table below (Table 9) presents the digitisation specifications for microfilms.

Table 9. Microfilm specifications

Element	Required value
Production master file's format	TIFF 6.0 or JPEG part 1 version 1.00 or 1.01 tai 1.02
Compression of production master file	TIFF = LZW or JPEG = JPEG quality level 90%
Resolution	300 ppi
Colour model	Greyscale
Bit depth	8 bits per sample, 1 sample (8-bit)
ICC profiling of the production master file	Gray Gamma 2.2

The requirements set out in this chapter **must** be applied also to colour microfilms. The colour model of the production master file for them **must** be RGB, the bit depth **must** be 24-bit and the ICC profile **must** be eciRGB v2, ProPhoto RGB, AdobeRGB (1998) or sRGB.

6 Accepted formats

This section on formats has been divided into two subsections:

1. Image file
2. Metadata and structure describing an image file and its processing

6.1 Image file

The quality of the image file produced in the digitisation process **must not** be lower at any processing stage than the requirements set in this chapter. The production master file **must** be saved either in the uncompressed/lossless compressed TIFF format or in the lossy compressed JPEG format (not in both).⁶ The storage format of the production master file depends on the material type of the analogue manifestation. Material type-specific determinations are presented in chapter 5 and its subsections. If the first image file (production master file) produced in the process is compressed with the lossy method, it **must not** be processed after the first save and then re-saved as a production master file.

The tables below (Table 10 and Table 11) present the required format-specific data that an image file **must** contain in a machine-readable format. If the "Element" column in the table is not specified, there is no required field for data in this connection. However, the data **must** be indicated. In addition to the data presented in the table, the image file **may** include other metadata fields. Metadata defined as personal data **must not** be saved in image files.

⁶ JPEG: <https://jpeg.org/jpeg/index.html> Retrieved 11 January 2021

TIFF: <https://www.itu.int/itudoc/itu-t/com16/tiff-fx/docs/tiff6.pdf> Retrieved 11 January 2021

19 January 2021

KA/15780/07.01.02.04.02/2020

Tables 10 and 11 show all possible values that the storage formats of the image files specified in these instructions **must** contain. The values in tables 10 and 11 **must** be interpreted together with the material-specific requirements presented in the chapter 5 and its subchapters. The above-mentioned sections are indicated in the column “Required value” with the value “Material type dependent”.

If the “Required value” column reads “Image file specific”, each image file has unique value. If the column reads “Production process dependent”, the value depends on the process that created the digital image files.

Table 10. Mandatory metadata in an image file (TIFF)

Element	Specifier	Required value	Metadata schema	Metadata field
Format	TIFF	image/tiff		MIME Type
Version	Required version of the TIFF file	6.0		
Image name	The image name is expressed as four numbers 0001.tif, 0002.tif, 0003.tif, etc. See chapter 7.			
Image file size				
Colorspace	Image file colorspace	Material type dependent. 0 = WhitesZero or 1 = BlacksZero or 2 = RGB	Exif.Image	PhotometricInterpretation (262)
ICC profile	Colour profile embedded in an image file’s metadata	Material type dependent. sRGB, eciRGB v2, ProPhoto RGB, AdobeRGB (1998) or Gray gamma 2.2	ICC	profileDescription
Bit depth	Number of bits per component.	8	Exif.Image	BitsPerSample (258)
	The number of components per pixel	Material type dependent. 3 or 1	Exif.Image	SamplesPerPixel (277)
File compression		Material type dependent: 1 = Uncompressed or 5 LZW	Exif.Image	Compression (259)
Image width	The number of columns in the image, i.e., the number of pixels per row	Image file specific	Exif.Image	ImageWidth (256)
Image length	The number of rows of pixels in the image.	Image file specific	Exif.Image	ImageLength (257)
Image file creator	Organization that created the image file from the analogue object	Production process dependent	Exif.Image	Artist (315)
Digitisation device (scanning or photography)	The device manufacturer. The manufacturer of the	Production process dependent	Exif.Image	Make (271)



19 January 2021

KA/15780/07.01.02.04.02/2020

Element	Specifier	Required value	Metadata schema	Metadata field
	equipment used to convert an analogue object into digital format (name of manufacturer)			
Digitisation device model (scanning or photography)	Specifying information about digitisation device by indicating the name of the model	Production process dependent	Exif.Image	Model (272)
Serial number of digitisation device	Identifies the device used (unique number)	Production process dependent	Exif.Image	CameraSerial-Number (50735)
Program used to generate image file	Software and version used to generate digital file	Production process dependent	Exif.Image	Software (305)
Date and time when the image file was generated (scanning date)	In format: YYYY:MM:DDTHH:MM:SS	Image file specific	Exif.Image	DateTimeOriginal (36867)
Orientation	File orientation (horizontal or vertical). This orientation does not concern the image's information content's orientation. Instead, this indicates the orientation of the file.	Production process dependent/ Image file specific	Exif.Image	Orientation (274)
Resolution unit	Measuring unit used to interpret the X and Y resolution	2 = inch	Exif.Image	ResolutionUnit (296)
XResolution	The number of pixels per resolution unit in the width direction	Material type dependent	Exif.Image	XResolution (282)
YResolution	The number of pixels per resolution unit in the length direction	Material type dependent	Exif.Image	YResolution (283)
Byte order		Big-endian or little-endian		ByteOrder
Image file processing software	If an uncompressed file is first generated in the digitisation process and the file is edited after scanning, the name and version of the editing program must be saved	Production process dependent	Exif.Image	Image.Processing-Software (11)

Table 11. Mandatory metadata in an image file (JPEG)

Element	Specifier	Required value/specification	Metadata schema	Metadata field
Format	JPEG	image/jpeg		MIME Type
Version	JPEG part 1 version	1.00 or 1.01 or 1.02		JFIF Version
Image name	The image name is expressed as four numbers 0001.jpg, 0002.jpg,			



19 January 2021

KA/15780/07.01.02.04.02/2020

Element	Specifier	Required value/specification	Metadata schema	Metadata field
	0003.jpg, etc. See chapter 7.			
Image file size				
Colour model	Image file colour model	Material type dependent. 0 = WhiteIsZero or 1 = BlackIsZero or 2 = RGB	Exif.Image	PhotometricInterpretation (262)
ICC profile	Colour profile embedded in an image file's metadata.	Material type dependent. sRGB, eciRGB v2, ProPhoto RGB, AdobeRGB (1998) or Gray gamma 2.2	ICC	profileDescription
Bit depth	Number of bits per component.	8	Exif.Image	BitsPerSample (258)
	The number of components per pixel	Material type dependent. 3 or 1	Exif.Image	SamplesPerPixel (277)
File compression	JPEG	6 = JPEG (old-style) or 7 = JPEG	Exif.Image.	Compression (259)
JPEG quality	The JPEG compression's quality on a scale of 0–100%	90%		
Creator of the image file	Organization that created the image file from the analogue object	Production process dependent	Exif.Image	Artist (315)
Image length	The number of rows of pixels in the image	Image file specific	Exif.Image	ImageLength(257)
Image width	The number of columns in the image, i.e., the number of pixels per row	Image file specific	Exif.Image	ImageWidth(256)
Digitisation device (scanning or photography)	The device manufacturer. The manufacturer of the equipment used to convert an analogue object into digital format (name of manufacturer)	Production process dependent	Exif.Image	Make (271)
Digitisation device model (scanning or photography)	Specifying information about digitisation device by indicating the name of the model	Production process dependent	Exif.Image	Model (272)
Serial number of digitisation device	Identifies the device used (unique number)	Production process dependent	Exif.Image	CameraSerial-Number (50735)
Program used to generate the image file	Software and version used to generate digital file	Production process dependent	Exif.Image	Software (305)
Date and time when the image file was generated	In format: YYYY:MM:DD HH:MM:SS	Image file specific	Exif.Image	DateTimeOriginal (36867)
Orientation	File orientation (horizontal or vertical). This ori-	Production process dependent/ Image file specific	Exif.Image	Orientation (274)

19 January 2021

KA/15780/07.01.02.04.02/2020

Element	Specifier	Required value/specification	Metadata schema	Metadata field
	entation does not concern the image's information content's orientation. Instead, this indicates the orientation of the file.			
Resolution unit	Measuring unit used to interpret the X and Y resolution	2 = inch	Exif.Image	Image.ResolutionUnit (296)
XResolution	The number of pixels per resolution unit in the image width direction	Material type dependent/ Image file specific	Exif.Image	Image.XResolution (282)
YResolution	The number of pixels per resolution unit in the image length direction	Material type dependent/ image file specific	Exif.Image	Image.YResolution (283)
Image file processing software	If an uncompressed file is first generated in the digitisation process and the file is edited after scanning, the name and version of the editing program must be saved	Production process dependent	Exif.Image	Image.Processing-Software (11)

6.2 Metadata and structure describing an image file and its processing

The metadata described in this section represents the creation history of image files. This history also verifies the authenticity of the digital object generated in the process. The technical metadata required for image files **must** be presented in accordance with the MIX metadata schema, version 2.0.⁷

The table (Table 12) below presents, from left to right, the name of the MIX field, the purpose of the field and the obligation. The obligation field indicates whether or not the specific field and data in accordance with its schema are obligatory as follows:

- R = required – this data **must** be described
- O = optional – this data **should** be described but it is not required

The MIX metadata schema contains two types of fields: containers and data elements. Data elements contain a specific value, while containers contain one or more data elements, and they can also contain other containers containing data elements. Table 12 only presents fields that contain a specific value, i.e., data elements.

⁷ The Library of Congress » Standards » MIX. Website of the Library of Congress <http://www.loc.gov/standards/mix/> and ANSI/NISO Z39.87-2006 (R2017) Data Dictionary - Technical Metadata for Digital Still Images <https://www.niso.org/publications/ansiniso-z3987-2006-r2017-data-dictionary-technical-metadata-digital-still-images>
Retrieved 11 January 1

19 January 2021

KA/15780/07.01.02.04.02/2020

Table 12. Metadata describing an image file and its processing (the table only presents fields that contain data that MUST be presented in a structure in accordance with the MIX metadata schema, version 2.0)

MIX field name	Field purpose	Obligation
objectIdentifierType	A data element that designates the system or domain in which the identifier of the object is unique.	R
objectIdentifierValue	A data element that provides the value of the objects identifier.	R
fileSize	A data element that designates the size in bytes of the image, e.g., 72839.	R
formatName	A data element that designates the format name or description of the file format file. Required value image/jpeg or image/tiff	R
formatVersion	A data element that designates the version of the format, such as 1.01 OR 6.0.	O
byteOrder	A data element that defines the order of bytes. The required value is either big endian or little endian.	R
compressionScheme	A data element that designates the compression scheme used to store the image data. Required value JPEG (when the production master file is in jpeg format) or LZW (when the production master file is in TIFF format).	R
compressionRatio	A data element that designates the ratio of compressed file size to original file size. Expressed using the numerator "90" and denominator "100".	R (JPEG only)
messageDigestAlgorithm	A data element that identifies the specific algorithm used to construct the value in the messageDigest-field. The field value is one of the following: MD5, SHA-1, SHA256, SHA384, SHA512.	R
messageDigest	A data element that specifies the output of messageDigestAlgorithm, e.g., e8064dc0.	R
imageWidth	A data element that specifies the width of the digital image, i.e. horizontal or X dimension, in pixels, e.g., 1,330	R
imageHeight	A data element that specifies the height of the digital image, i.e. vertical or Y dimension, in pixels, e.g., 1,600	R
colorSpace	A data element that designates the color model of the decompressed image data. The required value RGB or WhiteIsZero or BlackIsZero.	R
iccProfileName	A data element that designates the well-defined name of the ICC profile used. Required value sRGB, eciRGB v2, ProPhoto RGB, AdobeRGB (1998) or Gray gamma 2.2.	R
iccProfileVersion	A data element that designates the version of the ICC profile used, e.g., v4 [sRGB v4].	R
iccProfileURL	A data element that designates the URL/URN where the ICC profile is located, if the ICC profile is not well documented.	O
dateTimeCreated	A data element that designates the Date or DateTime the image was created. In format: YYYY-MM-DDTHH:MM:SS.	R
imageProducer	A data element that identifies the organization-level producer(s) of the image.	R
scannerManufacturer	A data element that identifies the manufacturer of the scanner used to create the image.	R
scannerModelName	A data element that identifies the model name of the scanner used to create the image.	R

19 January 2021

KA/15780/07.01.02.04.02/2020

MIX field name	Field purpose	Obligation
scannerModelNumber	A data element that identifies the model number of the scanner used to create the image.	R
scannerModelSerialNo	A data element that identifies the serial number of the scanner used to create the image.	R
scanningSoftwareName	A data element that identifies the name of the capture software used to create the image.	R
scanningSoftwareVersionNo	A data element that identifies the version number of the capture software used to create the image.	R
orientation	A data element that designates the orientation of the image.	R
samplingFrequencyUnit	a data element that specifies the unit of measurement for xSamplingFrequency and ySamplingFrequency. Required value "in."	R
xSamplingFrequency	A data element that specifies the number of pixels per samplingFrequencyUnit in the image width. The required value is material type dependent.	R
ySamplingFrequency	A data element that specifies the number of pixels per samplingFrequencyUnit in the image length. The required value is material type dependent.	R
bitsPerSampleValue	A data element that designates the number of bits per component for each pixel. The required value is 8.	R
bitsPerSampleUnit	A data element that specifies the interpretation of the bits as integer values or floating point values.	R
samplesPerPixel	A data element that designates the number of color components per pixel. The required value is either 3 or 1.	R
targetType	A data element that identifies the targets as either internal or external.	O
targetManufacturer	A data element that identifies the manufacturer or organization that created the target.	O
targetName	A data element that identifies the name of the target.	O
targetNo	A data element that identifies the version or number of the target.	O
externalTarget	A data element that identifies the path where the digital image of the reference target identified in TargetID-container is located.	O
performanceData	A data element that identifies the path of the file that contains the image performance data relative to the target identified in TargetID.	O

7 Transfer package generated in the digitisation process

Files produced in digitisation processes described in chapter 6 and its subchapters **must** be saved in accordance with the directory structure presented in the figure below (Figure 3) before they can be accepted to the National Archives. The digital object **must** be produced in the directory structure regardless of when it is transferred to the National Archives. When transferring material, the transfer package **must not** contain anything other than the directories shown in Figure 3.

19 January 2021

KA/15780/07.01.02.04.02/2020

In addition to the directory structure defined here, organizations **may**, for example, save access files in their own data systems using the data structure required by each specific system. Therefore, the structure defined in this record does not exclude the use of other storage structures.

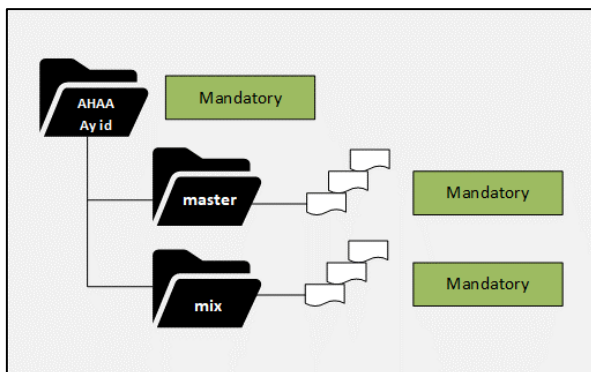


Figure 3. The required transfer package structure of the digitisation process

The table below (Table 13) describes how digital files **must** be named within the presented directory structure. The digital files produced in the process **must** match each other, in other words, the 0001.xml file in accordance with the MIX metadata schema **must** contain metadata describing the image file 0001.jpg or 0001.tif.

Table 13. Content of transfer package directories

Directory	Description
Identifier	An identifier of a digital object, with which it must be possible to identify the record entity in question (e.g., archival item). ⁸ The directory includes “digital file directories”.
Master	The image files presented in Table 10. Mandatory metadata in an image file (TIFF) or Table 11. Mandatory metadata in an image file (JPEG) must be saved in the directory as individual files. The file names must consist of four digits, starting from 0001.tif or 0001.jpg.
Mix	The required data presented in Table 12 concerning all image files located in the master directory must be saved in the directory. Other data in accordance with the MIX metadata schema MAY also be saved in files following a structure in accordance with the schema. The files must be named containing four digits, starting from 0001.xml.

If the data is delivered to the National Archives, each transfer package **must** be packaged into a TAR or ZIP package. The content of TAR or ZIP packages **must not** be compressed at this stage. A checksum **must** be calculated for TAR or ZIP packages in MD5 format, and it **must** be delivered in conjunction with the transfer. When delivering material to the National Archives, the identifier directory refers to the identifier Ai01 of the AHAA system’s material (the role of the identifier is “AHAA technical”).

Example packages

1. EsimerkkipakettiJPEG2020.tar

⁸ The analogue object to be digitised should be described (descriptive metadata produced) before its digitisation. Using the identifier, it **MUST** be possible to connect digital objects created in the digitisation process to the aforementioned descriptive metadata.

19 January 2021

KA/15780/07.01.02.04.02/2020

2. EsimerkkipakettiTIF2020.tar

The image files contain the metadata defined as requirement in this document. If not stated otherwise, the MIX-directories' files are examples of master files in the package. Please note that the image files in these master directories cannot be used as references for image quality.

Signatures

Director General

Jussi Nuorteva

Director of Research

Päivi Happonen