DIGITIZER'S HANDBOOK
Version 3.3

I. Introduction

I.1. Trustworthy facsimile-quality digital reproduction

A trustworthy digital reproduction is a digital image, which has been created with the aim of precisely conveying the original document in both its fullness of content (including autographs, remarks and other special characteristics) as well as its outlook (tone and coloring) and the sequence of pages. It is possible to create printed facsimiles from trustworthy digital reproductions, if they are created in the same size as the original (1:1). Trustworthy digital reproduction must provide long-term availability (eternal accessibility) and be compatible with various computing platforms and software.

Practically, there are several versions of digital reproductions made during digitalization:

Archive images (files): optimized for long-term storage and creation of copies of differing quality. Those are uncompressed and unedited images.

It is not recommended to try to improve archive files by post-processing (see Addendum B). The users will be able to make their own enhanced copies, which will be designated by the access rights given for each digital object. Archive images must be created in two copies, which are stored on separate media in two physically remote sites.

Access files: these are archive files post-processed and enhanced (see. Chapter III) to satisfy the following needs:

- Publishing on the Internet (WWW) and viewing on computer screen;

- Generation of thumbnails for preview in Internet browsers.

II. Archive files (master files)

A major part of digitalization project consists of scanning a high-quality archive master image, which provides the source for creation of different format or smaller images for specific needs. Digital archive file is created keeping in mind preservation, access and economic criteria – it must provide for informationally rich, unedited, research-quality facsimile and there must be an assurance that there will be no need to rescan the original in future. As the needs of users and technologies are

subject to change in time, the digital archive file must be accessible and of high enough quality to satisfy even future needs and be compatible with future applications. It should not be edited, post-processed or compressed (for a list of acceptable post-processing enhancements see table in Appandix B).

Rigid quality control must be maintained during creation of an archive file by carefully comparing it with the original.

> Archive files must be of as high quality as possible, considering economic constraints. It should be noted that storage costs are decreasing every year.

The same as with images, archive files of texts are to be saved as TIFF images, from which access files are created later on. Access files are usually created by a single operation from a larger number of archive files.

II.1. Mandatory requirements

- It must be possible to create facsimile copies of the original document from all archive images – precise both in its fullness as well as outlook (tone and coloring) and the sequence of pages, volumes etc.;

- The archive files must show the relations between various parts of the original (pages, volumes, etc.);

- Empty pages (eg. image reverses) must be included in the total count and sequence of pages;

- Descriptive metadata of the object (eg. book) must be attached to the parts of the object (eg. pages) insofar as neccessary, for example, for citations and references.

II.2. Requirements dependant on the original

There are requirements, such as numbering of pages, which can only be maintaned, if they exist in the original. Another example is missing pages, which can only be replaced, if it is established without doubt that original has had them.

- Wherever possible archive images must ensure navigation to and between logical parts (eg. chapters, volumes, sections, etc.) as well as show important characteristics of the document (eg. tables, illustrations, empty pages). References to these characteristics must be maintained;

- Page numbering must be demonstrated, if the original has such numbering. Various types of numeration must be preserved (eg. Roman numerals in introductions);

- Missing pages must be demonstrated by using appropriately numbered blank pages. To create a digital facsimile, archive files must provide information on missing components in object-level metadata. If missing pages can be obtained from other organizations, information about this fact must be entered in descriptive metadata;

- Covers must be scanned as well, preferrably in original binding. If those are missing, it must be indicated in descriptive metadata;

- Monography structure must include title pages, content pages, illustration lists, indices, chapters, etc.;

- Serial publication structure must include information on volume, section, issue number, etc.

Object-level metadata will include references to page numbers, logical structure and other clearly visible characteristics.

> **DPI** – dots per inch. The number of pixels or points that are used to represent one inch (2,54 cm) of original in a digital image; used in relation to printer output and scanner settings. It is important to note that in case of printers it actually means ink dots.
>
> **Bit depth** – (also bpp, bits per pixel) number of bits used to define a single pixel's color in a digital image. 1-bit images provide B&W images, current technology usually allows for colour images to have from 8 to 24 bpp to describe each pixel's color.

III. Access files

All access files are created from archive files, meeting the following general conditions:

1. Best quality access files are created first.

- Format: JPEG (photographs) or PNG (graphics and drawings); JPEG image resolution is reduced to 96DPI, color depth should remain the same as archive file.

- Access files for extremely large materials or such as require specific browsing and navigational capabilities (zooming, area marking, playback, paging, etc.) can be created with specialized software. For example, maps – GeoExpress, newspapers, books – DjVu or Adobe Acrobat, etc.

> Note!
>
> Adobe Acrobat PDF documents for long-term storage must be created according to PDF/A-1 standard (ISO 19005-1). Its distinctive characteristics are embedded fonts, XMP metadata support and device independant color models.

2. Thumbnails are created by several passes (3-6) of archive image size reduction, after each reduction the image is sharpened to help preserve details. After size reduction to optimal size (for maps and other large materials it is currently estimated at around 640 pixels for widest side, for smaller materials – around 150; thumbnail image size is dependant on the resulting webpage design), image is preserved in JPEG 96DPI same color depth as archive file, 85% compression.

IV. Scanning and post-processing

The most important thing is evaluation of the original, since the best results are obtained by taking into account both the original and the chosen scanner (drum, flatbed, craddle, etc.) and it software. The person responsible for scanning shall enter his name in the appropriate metadata field to indicate that he has evaluated the original. See Appendix D for examples of scanning of different types of documents.

> Note: When using a scanner without a cover, it is vital to have the exact same lighting conditions while working on a single object.

It is not possible to describe all the cases one may encounter in scanning process, and there are always exceptions, however all the formats (color, b&w, half-tones) have some common principles that should be observed:

> To ensure the coverage of all details, scanning should use the smallest or worst-readable element as base;

The following materials should be used in conjunction with a ruler and greyscale separation guide, color materials – also color separation guide[1]:

- Rarities;

- Cartographic materials;

- All imagery (photographs, postcards, posters, graphics, etc.)

It is advisable to check scanner calibration at least once per month with specific tools such as a template that allows to check for precise scanning of thin parallel lines or delicate gradients

---

[1] Such as provided by, for example, ColourConfidence.com or other colour management providers.

We recommend using a greyscale (for b&w newspapers) or colour (for colour) separation guides and a ruler for the first page of the first issue of a newspaper file, saving it as a separate archive file.

IV.1. Stage 1 – creation of archive files.

1. Material is placed on clean scanner surface. Since old documents may shred or sap, it may be neccessary to clean the surface after each image;

2. If possible, do a preview;

3. Crop the image, leaving enough of a border (≥ 2 mm);

4. Set the DPI with the scanner software;

5. Scan the material;

6. Save a high-resolution unedited TIFF on the post-processing workstation;

7. Crop the image in post-processing image editing software.

Note: Do not crop too much! Archive files must have at least 2 mm border to assure the future user that all of the page has been digitised.

8. Adjust the histogram, if needed.

9. When enough of scanned material is accumulated, backup on safe media (eg. magnetic tapes).

IV.2. Stage 2 – creation of access files.

10. Setup gray mid-point, if used;

11. Adjust image size, if neccessary;

12. Make other adjustments (tone, sharpness, noise, etc.), to provide for as clear a picture as possible;

13. Adjust the resolution depending on the access file target (print, display, etc);

14. Check, if access file visual characteristics represent the original well enough;

15. Save the file as JPEG (or PNG for graphics). This will be the access file that can be used for eg., Internet;

16. Create the thumbnail as JPEG or GIF;

17. Upload the access files to the web page.

IV.3. Some practical guidelines to determine optimal scanning methods

- All manuscripts must be scanned 1:1 (100%) in relation to original.

- Film negatives and microfilms must be scanned at least 500% (5x bigger) than the original.

- DPI choice is dependant on the amount of minor details in the original, but should not be less than 300dpi (facsimile quality) since this is also the resolution for submitting materials for typographic processing.

- Higher resolution may be required in order to fully reproduce smaller texts, delicate lines and dots as well as specific alphabets and fonts (such as *italics*, Arabian letters etc.),

- The smaller the original image, the higher the resolution needed.

- Even if the original is purportedly b&w (eg. old newspapers), the yellowed paper may be viewed as a colour material by an artist. Use half-tone (8-bit grayscale) for materials, which contain gray tones and gradual changes: b&w photographs, half-tone photographs, typewritten text, b&w archive materials which contain changes in ink coloring and in which it is important to show the tone of the paper (lighter/darker).

- Use 24-bit or better color depth in all cases where the original color is of importance. Object may be scanned in color as decided by its artistic value, if its physical attributes such as ruddish spots are viewed as bearers of important information or meaning.

- In cases where only the cover or title page is coloured, it is scanned as coloured apart from the rest of the document (this is characteristic of such documents as music scores, programmes, etc.).

- In cases of planned text recognition (OCR) and the original is barely readable, scanner Brightness or Threshold settings may help:

- From larger to smaller (eg. from 128 to 85), if the background is dark or spotty and the letters are dark, too.

- From smaller to larger (eg 128 to 180), if the background is light and the letters faded.

- If the paper is so thin the elements on the opposite side of the page are shown when scanning a page, it is advisable to put a black cardboard sheet between the pages.

- Illustrations of the books are usually scanned at higher resolution that its text.

IV.4. Image processing

IV.4.1. Orientation.

- All the images in the storage (DVD, server or other) must be saved in portrait orientation, except cases when the original itself is landscape.

- Pages of bound editions such as newspapers are scanned each page to a separate file.

- It is permissible to scan spread when scanning rarities.

- Top of the page of the original must be oriented towards the top of the display in the access file .

- In order to fit the page, an illustration or table may be printed in landscape orientation, in which case the top side of the image will match the top side of the page rather than that of illustration.

IV.4.2. Cropping

- Under no circumstances should the scanned archive file be cropped, since it is needed as a whole for research purposes. Whenever possible even cropped access files should leave some small (2 mm) whitespace around the edges of the image.

IV.4.3. Skewing

Image scanned from unbound originals should have no cropping. For bound originals, it may be impossible to completely avoid skewing if the binding is tight. However, the person responsible for scanning should make an effort to avoid skewing as much as possible. Post-processing (deskewing in this case) of archive files is only allowable in exceptional circumstances (see Appendix B).

IV.4.4. Merging

Documents, which are too large to be scanned in one pass, can be scanned in two or four parts, however, there must be an overlap of at least several centimeters, so that the parts can be merged together later with special software such as PhotoMerge. It should be remembered that all parts of the document must be scanned with the same settings and under the exact same lighting conditions.

If the document is scanned in two parts, the merging process is oriented by the centers of both parts, which allows for smaller displacement.
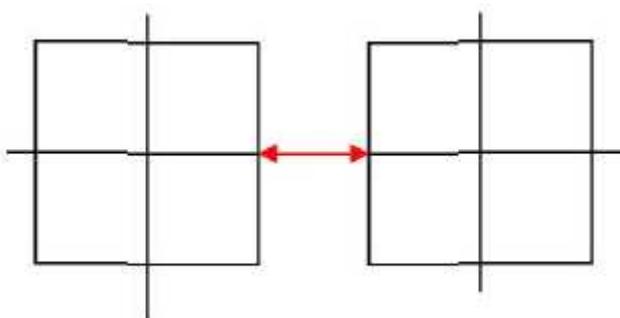
Whenever possible scanning should be completed in one pass.

II DATA ORGANIZATION

File names and the structure of directories

The system of file naming shall be defined before scanning. It must be considered whether the files will be indexed manually, automatically, or both ways. In the last case several file identifiers would be necessary. The file names may be either descriptive (presenting the relationship between the file and the original document), or non-descriptive (e.g., a string of digits). The descriptive file names contain self-referencing metadata, whereas the non-descriptive file titles are linked with the metadata that are stored separately but serve to identify the respective file.

Generally for smaller projects the files can be given descriptive names thus improving browsing; for larger projects the system assigns file names, and the search and retrieval of corresponding metadata is performed via databases.

1. The structure of directories

Files usually are arranged in the system of directories irrespective the file names. Often master files are stored separately from access files. The directories may have their own arrangement that is not linked with the image in the file (e.g., it may be the arrangement according to dates or the group numbers of records). They may as well replicate the original physical or logical arrangement of the original documents (e.g., newspapers/ /maps/, etc.)

The files themselves also may be arranged only according to the structure of the directories without assigning them descriptive names. This approach works well if the objects contain many pages. Identification numbers are assigned to images within the framework of the logical object (e.g., books, chapters, issues, etc.); they demand a descriptive name for the directory. The names of individual images are unique only within this directory, not in general.

For example, the book 0001 contains image files 001.tif, 002.tif, 003.tif, etc.

The book 0002 contains image files 001.tif, 002.tif, 003.tif. This solution occurs dangerous if the images are taken away from their initial context or directory. In such a case it is impossible to distinguish them from the images of another directory.

2. Versions

For various reasons one scanned object can have several different versions (the images have been scanned by different scanners; additional editing; layers, alpha channels, etc.)

Ideally the descriptions and the aims of production of various versions should be reflected in metadata. If the file naming principles are strongly followed the file names allow fast identifying of images. Usually it means adding a denotation to the file name. The aim of this action is to place all files of one object into system side by side under one identifier. It is important to consider whether numerical identifiers (001, 002, 003) are not used for some other purpose, for example, for page numbers.

3. File naming

When developing the system of file naming it must be regarded that access files will be created from the master files. In most cases the access files inherit their names from their master files acquiring additional name differing them according to their use (e.g., ''p'' - for printing, ''th'' - for thumbnail, etc.). As long as it is possible to link them with the respective master files the names of access files have no need to be descriptive.

Users can employ the access files searchable in the World Wide Web when searching for the master files of better quality. In this case the file names must be descriptive enough to provide for easy access to the master file or another digital version.

It is to be regarded that some operational systems have a range of characters used for specific purposes; the use of these characters for file naming is not permitted.

Example: . / | ?*

Recommendations for file naming:

- ❖ Each file name of the object **is unique;**
- ❖ The maximal length of the file name does not surpass 18 characters;
- ❖ Only Latin small letters and Arabic numerals 0-9 are used;
- ❖ All characters are exploited as the code for the object;
- ❖ The assigned file name **remains unchanged** over time;
- ❖ The names remain unchanged when rewriting the files on another data carrier;
- ❖ File names survive when changing the system of data storage.

In practice the following file names are also used: 5 digit ID _ the year of scanning _ the number of page or file _ the type of the image (archive image, access image, thumbnail, tif)

Example: 0001_2003_001_aa_tif

Appendix E *The file naming system for the National Library of Latvia* presents the samples of the file names of the objects stored in the file system of the library. Their complexity is caused by the lack of a unifying element for all the objects stored in the library, e.g., inventory number, code, system ID, etc.

Appendix F *Description of the structure of directories* shows the structure of directories for the server of backup copies of the National Library of Latvia.

METADATA

INTRODUCTION

Metadata is the description of the qualities of an object serving for the discovery of this object - like bibliographic description or file naming (see chapter II *File Naming*). Usually it is a structured description. The metadata of digital objects in accordance to their tasks describe: the content, origin and context (descriptive metadata); the structure and relationship of the object (structural metadata); technical characteristic (technical metadata); living cycle (storage metadata); access and use conditions (rights' metadata). The last three types of metadata are called also administrative metadata.

Example:

Descriptive	Structural	Administrative		
		Technical	**Rights'**	**Preservation**
Title	Type of structure	Size of the file	Copyright status	Date of changes
Creator	Map of structure	Dimensions	Copyright owner	History of changes
Publisher	Sequence	Resolution	Right to copy	Version of the file
Subject	Structural links	Compression algorithm	Right to download	

In practice the frontier between the descriptive metadata and the structural metadata often disappears, and the functionality of elements overlaps. For example, the element of descriptive metadata ''Relation'' used to show the relationship of an object with other objects is usable also for reflecting the structure of the object.

Metadata are added to all kinds of digital objects – text documents, audio files, images, video records, etc., although the content of metadata (the fields and their number) of various types of objects may differ. The metadata of digital objects can be created manually (characteristic for the descriptive metadata), or they may be generated automatically (characteristic for the technical metadata).

The task of metadata is to provide for arranging and finding information. The metadata of objects or collections consist of individual elements (called fields) reflected in the metadata scheme. The schemes define the names of elements and their meaning. The values assigned to elements are their content. The terms used in metadata are shown in application profile.

Normally internationally acknowledged standards are used for metadata creation (e.g., *Dublin Core Metadata Element Set*). Standards provide for the interoperability of different systems.

When tailoring the metadata scheme for the digital library of the National Library of Latvia the international standards of the realm and the standards approved by the states of the European Union were considered.

COLLECTIONS

Collection is the aggregate of digitalized objects framed by one project or by the criteria its compilers have defined (common origin of the objects; belonging to one and the same genre, theme, etc.)

Example:
- ❖ Post-card collection of the Baltic Central Library
- ❖ Misinsh' collection of the Academic Library
- ❖ Jazeps Vitols

Browsing collections is one of the ways to get acquainted with the content of the digital library. When the number of collections increases, and it becomes difficult to overview them it is time to introduce the hierarchy of collections.

It is not recommended to form more than three levels of the collection hierarchy to avoid navigation complications.

Example:
1. Latvian periodicals
   a. Magazines, journals
   b. Newspapers
      i. Newspapers in Latvian
      ii. Newspapers in German
2. Images
   a. Placards in Latvia
   b. Post-cards (picture-cards, greeting-cards)
   c. The post-card collection of the Baltic Central Library
   d. The history of Latvian culture in pictures: Portraits
   e. Black-and white pictures: Ex-libris and prints
3. Latvian music
   a. Jazeps Vitols

According to the opinion of the National Digital Library of Latvia collection is a digital object, and like other objects metadata shall be added to it (see Appendix G *Metadata of collection*). Each of the collections is assigned also a unique URN.

The portal of the European Library (TEL) forms the digital collections on similar principles. The codes of the Universal Decimal Classification (UDC) provide for thematic browsing of collections (see Appendix H *UDC scheme for the indexing of documents of the digital library*).

Example from TEL:

Philosophy and Psychology
   100 – Philosophy and Psychology
   110 – Metaphysics

120 – Epistemology, causation, humankind
130 – Paranormal Phenomena
140 – Specific philosophical schools
150 – Psychology
160 – Logic
170 – Ethics
180 – Ancient, medieval, Oriental philosophy
190 – Modern Western philosophy

A shortened UDC version is prepared for the needs of the digital library, it includes all the basic classes of UDC, but the subclasses and the specification level is adjusted for the content of the library. To respect unitary indexing principles references are added to those indexes that might be differently interpreted.

Example:

3 Social sciences
31 Statistics. Demography. Sociology
311 Statistics (including concrete realms – medicine, elections, press, libraries, etc.)
314 Demography
316 Sociology

For users not familiar with UDC structure alphabetical index of wordy formulations of the notations is available. For more convenient use inverse forms of wordings are supplied.

Example:
Social psychology   316.6
Social work   36
Social problems   36
Social theories   1
Sociology   316

The alphabetic index and wordings of the UDC scheme are available also in English (see Appendix J), it provides for the browsing of our collections in TEL and other international portals.

One or more indexes may be selected from the UDC scheme to reflect the content of the collection.

For example, the collection of the history of the Songs' and Dances' Fest is assigned 3 indexes:

78.087.68 Choirs. Choral music
793 Dancing. Folk dances
39 Ethnology. Ethnography. Life-style and manner

One multiform collection may be assigned several UDC codes. (It is not recommended to assign more than 5). At the same time one digital object may belong to several collections.

If the thematic coverage of the collection is wide, and many codes should be used the criterion is the theme of the dominating objects of the collection.

For example, the post-card collection is assigned index 656.8 – Postal connections. Philately.

The UDC Authority File can be used to select the code. Although only the indexes of the UDC scheme with methodical references are used for the collection indexing. It is possible to suggest complements if necessary.

DIGITAL OBJECTS

Descriptive metadata

Descriptive metadata is an important part of the digital library. These metadata usually are added manually, and they characterize the content, origin and context of the digital object.

Descriptive metadata are added to all kinds if digital objects (texts, photographs, maps, video, audio recordings, etc.) employing internationally acknowledged metadata scheme. The metadata scheme includes metadata elements (called also fields), for example: Title, Creator, Publisher, Date, Size, Identifier, etc. The metadata scheme (aggregation of elements) may differ for various types of objects.

The content of metadata elements is filled in accordance with internationally acknowledged standards and methods. For example, the element ''Date'' is filled according to ISO 8601 and *Anglo-American Cataloguing Rules,* 2nd edition. The element ''Subject'' is filled selecting values from the Authority database of the National Library of Latvia where authority records are adjusted according to the subject headings and the methods of their derivation of the Library of Congress of the USA. At present The National Library develops the methods for subject heading derivation for the digital library.

The descriptive metadata provide for convenient and qualitative search (hitting) of digital objects based on concrete metadata elements.

The scheme of descriptive metadata is applicable for all the memory institutions involved in the digital library (museums, archives, libraries…)

The scheme of descriptive metadata applied in the National Library of Latvia is reflected in **Appendix K**. Short characteristic of the scheme:
  ❖ The standardized metadata of Dublin Core Metadata Set provide for the minimal description level;
  ❖ Specific metadata are added to individual types of digital resources (scale and projection to maps, text-first-row to printed music, etc.);
  ❖ The scheme follows *TEL Application Profile for objects* (version 1.5) - for international cooperation and interoperability purposes;
  ❖ The scheme contains general references, guidelines and methods for filling each of the metadata elements.

Appendix A.   Common minimal requirements for archive images.

Bit depth	Grayscale – 8bpp, Colour – 24bpp
Archive file format	Uncompressed TIFF v6 images
Recommended color space	RGB or Photo YCC, especially for cases when only a single copy of archive file is made. Bess suited for viewing on screen.

Appendix B Archive file characteristics

Table 1 - Image file characteristics

Document type	Resolution (dpi)		Bit depth	Permitted post-processing	Remarks
	Minimal	Recommended			
Printed text	300	300	8GS[1]	Minimal tone adjustments	300dpi is optimal for OCR
Newspapers, magazines, catalogues (if the page contains both text and images)	300	300-400	8GS / 24C[2]	Sharpening, cropping, deskewing	400dpi for very old and/or damaged newspapers and magazines. Even if the original is black on white, scanning is done in 8-bit grayscale (see Example 4 of Appendix D)
Small printed Latin alphabet text ≤ 6dpi	400	400-600	8GS	Sharpening, cropping, deskewing	
Rarities or blurred printed text	300	400-600	24C	Contrast and minimal tone adjustments	400dpi, if OCR is planned, especially for Gothic fonts

---

[1] 8GS – 8-bit grayscale
[2] 24C – 24-bit colour

Book illustrations	300	400-600	8GS / 24C	Contrast and minimal tone and colour adjustments	600dpi for illustrations with thin lines, such as etchings
Handwritings, lettters, manuscripts	300	300	24C	Contrast and minimal tone and colour adjustments	
Maps and other large-scale documents	200-300	300-400	24C	Contrast and minimal tone and colour adjustments, sharpening	200dpi can be quite enough for Soviet-time maps
Sheet music, scores	300	400	8GS / 24C	Sharpening, cropping, deskewing	PhotoScore MIDI (OCR software that comes with Sibelius software) is best at dealing with 200-400dpi 8-bit grayscale. If the original contains eg. conductor's notes such as made with coloured pencil, scan 24-bit colour
Art drawings (graphics)	300-400	400-600	8GS / 24C	Sharpening, cropping, deskewing	Eg. posters, postcards, old concert programs, handwritten texts Pre-war colour postcards and similar documents have

					large raster size, therefore high resolution scanning can create noise (see Appendix D, Example 6)
		1200	1-bit B&W / 8GS / 24C	Contrast adjustment, minimal colour adjustment, sharpening	For drawings with very thin lines such as etchings or ink drawings
Paper-based artworks	300	400	24C	Contrast normalizatrion, minimal colour adjustment	Eg. sketches, drawings, etc.
Developed photographs	300	400-600	8GS	Contrast normalization, minimal colour adjustment, sharpening	600dpi is needed for old photographs containing lots of fine detail (see Appandiex D, Example 3)
	300	400	24C		Duotone Sepia and B&W photographs with added colour information (such as seal) are scanned 24-bit colour
35mm negatives or slides	2100	3300	8GS / 24C	Contrast normalization, minimal tone and colour adjustment, sharpening	DPI calculated depending on film grain

| Microfilms | 300-400 | 600 | 1-bit B&W | Sharpening and cropping | |
| | 200 | 300-400 | 8GS | Minimal tone adjustments | |

Table 2 - Audio and video file formats

Object type	Sampling rate		Bit depth		File format	Remarks
	Minimal	Recommended	Minimal	Recommended		
Audio recordings	44,1 kHz	44,1 kHz - 96 kHz	16-bit	24-bit	BWF (broadcast wave format, the European Broadcasting Union variant of WAV) for audio	A human ear can not perceive a sound with a 20kHz or higher pitch. 96 kHz is a professional standard
Video (see Appendix C)	Uncompressed or original format	Uncompressed or original format			MXF	

Appendix C Access files for viewing on computer screen

> Access files may be enhanced (despeckled, adjusted for better readability, etc.).

GIF files are suited for: illustrations, clipi art, images with large areas of single colour.

JPEG files are suited for: photographs and images with tone and colour gradients.

Larger access files may be created in different formats for the same archive file – DjVu, PDF, MrSID, Sibelius, etc. and in different size and resolution. It depends on how we want to offer the digitized object to the end user. Therefore common principles can be established for thumbnails only.

Table 1. Thumbnail characteristics

Document type	Resolution, dpi	Bit depth	Format	Dimensions
Printed text (eg. books)		4-bit grayscale/ 8-bit colour	GIF	Reduced to 150-200 pixels longest side
Letters, manuscripts			GIF / JPEG	
Maps, drawings				
B&W photographs		4-bit grayscale		
Colour photographs	96	8-bit colour		
Paper-based artwork		4-bit grayscale / 8-bit colour		
Microfilms, transparencies		4-bit grayscale / 8-bit colour	GIF	
Sheet music		Specific icon created to denote all sheet music files	GIF	Eg. ¼-note image 150-200 pixels on longest side

Audiofiles	44,1 kHz	16-bit	MP3	Digitally born audio files remain in original format
Videofiles	1,8 kb/s − 500 kb/s (typical)			

Appendix G.
Metadata elements for collections

The creation of metadata for collections is based on the European digital library guidelines based on Dublin Core standard elements (*TEL Application Profile for collection descriptions (version 1.5)*)
The following minimum of metadata should be used:

No	Field	Notes	Example
1.	**Title**	The title of the collection. If the collection has an additional explanatory title it is added and separated by space-colon-space.	*Collection of pictures : Portraits*
	AlternativeTitle	It is very desirable to add the name of the collection in English and/or another language, parallel to the original	
2.	**Creator**	Memory institution – creator of the content of the collection	
3.	**Contributor**	Memory institution contributing to the creation or enriching of the content of the collection, e.g., by adding scientific comments	
4.	**Publisher**	Institution responsible for the availability of the collection. In case the institution publishes the collection of another institution it shall be specified as Publisher	
5.	**Subject - UDC**	The thematic deciphering of the objects included into the collection	

		using UDC indexes. UDC index list (appendix) is used for this purpose. It is recommended to conform the selected indexes with the leading specialist: Elita Eglite (elita.eglite@lnb.lv phone: 67224763)	
6.	**Description**	A short description of the collection. It is desirable to include the numerical, temporal and spatial coverage and other relevant information	*The project comprises a wide range of objects:* **Places:** *creeks, sites of ancient settlements, streets that once existed, etc.* **Buildings:** *palaces, manors, houses, fishermen villages destroyed during war, etc.*
7.	**Type**	List the types of the resources included into the collection	*Photographs Postcards Audio records*
8.	**Language**	List all the languages used in the digital objects. Start with those mentioned more often	
9.	**Relation**	List various kinds of relationship (e.g., with similar collections; with the collections whose digital objects have been used, etc.)	
10.	**Audience**		
11.	**Thumbnail URL**	It is possible to add an application icon to the metadata (a logo or a particular image characterizing the collection)	

J. UNIVERSAL DECIMAL CLASSIFICATION APPLICATION FOR A DIGITAL COLLECTION

0 Generalities

001 Science and knowledge in general. Inventions
002 Documentations. Books
004 Computer science
005 Management theory
008 Civilization. Culture. Progress
01 Bibliography. Catalogues
02 Librarianship. Libraries
030 General reference works
050 General serial publications, periodicals, yearbooks, calendars etc.
06 Organizations of a general nature. International organizations (EU, UN etc.)
070 The press. Journalism
08 Polygraphies. Publications for particular classes of reader (popular writing, children's literature etc.)
09 Rare books. Manuscripts

1 Philosophy. Psychology. Logic. Ethics
1 Philosophical and social systems. Feminism. Logic. Aesthetics
133 Occultism. Astrology

159.9 Psychology
159.92 Developmental psychology. Sex psychology. Genetic psychology. Ethnopsychology.
159.93 Personality
159.94 Sensory perception
159.94 Emotions. Motor functions
159.95/159.97 Mental processes
159.98 Applied psychology

16 Logic

17 Ethics

2 Religion

257 Ancient Religions of Europe. Paganism. *Dievturība*
26 Judaism
27 Christianity
271 Orthodox Church
272/273 Catholic Church
274 Protestant churches. Lutheran churches

3 Social sciences

31 Statistics. Demography
311 Statistics (incl. branch industry statistics)
314 Demography
316 Sociology
316.6 Social psychology
316.7 Sociology of culture. Communication. Mass media. Media

32 Political science. Politics
323 Home affairs. Internal policy. Ethnic relations. Relation between people and state
324 Elections. Referendums
327 Foreign policy. International relations
328 Representation of the people. Governments. Parliaments
329 Political parties and movements

33 Economics
331 Labor. Labor economics
332 Regional economics. Land economics. Landed property. Housing economics
336 Finance. Public finance. Public budget. Taxes. Money. Banking
338 Economic situation. Economic policy. Economic panning. Economic sectors
    incl., tourism. Prices
339 Trade. Commerce. International economic relations.
339.7 International finances
339.9 International economic relations

34 Law
341 International law. Law of international organizations. International treaties.
    Diplomatic law. Law of war. War crimes. Human rights. International courts
342 Public law. Constitutional law. Constitutions. Constitutional courts. Civil
    rights. Administrative law
343 Criminal law. Penal offences. Criminology. Forensic science. Penitentiary
    institutions
346 Law of government control of the economy
347 Civil law. Private law.
347.9 Courts. Legal procedure
348 Ecclesiastical law.
349 Special branches of law. Labor law

351/354 Public administration
        High level (cabinet), regional level of administration. Local government

355/359 Military affairs. Army service

36 Social welfare
    Financial benefits. Pensions Basic human needs. Housing problems. Social
    work Consumerism
368 Insurance

37/378 Education
37 Educational theories. Educational psychology. Educational sociology
371 Organization of educational system. School organization
373. Preschool education. Primary school. Secondary school
374 Further education
377 Vocational education
378 Higher education

378.9 Leisure
Collecting (as hobby). Home handicraft hobbies. Pet-keeping as hobby. Outdoor excursions

39 Ethnography. Ethnology. Customs. Traditions. Way of life. Manners
391 National dress. National adornment

398 Folklore
398.2 Folk tales, legends, anecdotes
398.3 Beliefs. Beliefs and customs concerning particular times and festivals
398.6 Riddles
398.8 Folksongs
398.9 Proverbs. Folk sayings

5 Mathematics and natural sciences
50 Environment. Damage to the environment. Environment Protection
51 Mathematics
52 Astronomy. Astrophysics. Space research. Geodesy. Cartography
53 Physics
54 Chemistry.
55 Earth sciences. Geological sciences. Meteorology
56 Paleontology
57 Biology. Anthropology
58 Botany
59 Animals. Zoology

6 Applied sciences. Medicine. Technology

61 Medical sciences
613 Personal hygiene
614 Public health and hygiene
614.8 Accidents. Accident prevention. First aid
615 Therapeutics. Pharmacology. Alternative medicine. Homeopathy
616 Pathology. Pediatrics. Psychiatry
617 Surgery
618 Gynecology.

62 Engineering. Mechanical engineering
620 Materials science
621 Energy industries. Nuclear technology. Electrical engineering. Machinery

621.3 Electrical engineering. Radio. Television. Video. Telephone
623 Military engineering
624 Structural engineering
625 Railway engineering. Highway engineering
626/627 Hydraulic engineering. Hydraulic structures
628 Sanitary engineering
629 Transport vehicle engineering

63 Agriculture. Forestry. Hunting. Fishing
630 Forestry
631 Agricultural industries
632/635 Field crops and their production. Garden plants. Gardening.
Vegetable gardening
636/637 Domestic animals. Animal husbandry. Veterinary science
638 Beekeeping
639 Hunting. Fishing

64 Home economics. Housekeeping. Hospitality industry
641 Food. Cookery
643/645 The dwelling. Furniture and household fittings
646 Clothing. Fashion. Fancy work. Body care. Beauty culture

654 Telecommunication
Radio broadcasting. Television broadcasting. Telephone systems

655 Printing. Publishing. Book trade

656 Transport and postal services
656.8 Postal services. Stamp collecting

657 Accountancy

658 Business management

659 Public relations. Advertising

66/68 Industries

69 Building. Building materials

7 Art. Entertainment. Sport
7.0 Theory and history of art. Design
71 Regional, town and country planning. Planning of landscape. Parks, gardens
etc.

72/76 The Arts
72 Architecture
73 Sculpture
73.071.1 Sculptors

738 Ceramics. Pottery
739 Metal arts. Jewelry making
74 Drawing. Caricatures
745/749 Decorative arts. Craftsmanship
75 Painting
75.071.1 Painters
76 Graphic art. Posters

77 Photography. Cinematography

78 Music
78.071.1 Composers
78. 071.2 Performing musicians
78.087.68 Choirs
781 Musicology
781.24 Notation
782/785 Kinds of music
782 Dramatic music
783 Sacred music
784 Vocal music
785 Instrumental music
786/789 Music for individual instruments

79 Entertainment. Performances. Sport
791 Cinema. Films
792 Stagecraft. Dramatic performances. Theatre. Opera. Ballet. Choreography
792.071.2.027 Stage directors
792.071.2.028 Actors
793 Dance. Folk dancing
794 Table games
796/799 Sport
796.3 Ball games
796.4 Gymnastics
796.5 Mountaineering. Orienteering. Tourism
796.6 Cycling
796.8 Motorcycle racing
796.9 Winter sports
797 Water sports.
798 Horse sports
799 Fishing. Hunting

8 Languages. Linguistics. Literature
80 Philology

81 Linguistics
81'255 Translating
811 Languages
811. 174 Latvian language
811.174' 34 Latvian phonetics

811.174'35 Latvian orthography
811.174'36 Latvian grammar
811.174'37 Latvian semantics
811.174'38 Latvian stylistics

82 Literature
82.0 Literary theory. Literary criticism. History of literature
821 Literature of individual languages
821.174 Latvian literature
821.174-1 Latvian poetry
821.174-2 Latvian drama
821.174-3 Latvian prose
821.174-84 Aphorisms

9 Geography. Biography. History

90 Archaeology. Prehistory. Antiquities
902 Archaeology
903 Prehistory. Antiquities
908 Local history

91 Geography
91(474.3) Geography of Latvia
910 Discoveries. Travels
911 Physical geography. Population geography
912 Maps
913 Regional geography

929 Biography. Genealogy

93/94 History. Civilization
930 Historiography. Archival science. Chronology
930.85 Civilization. History of civilization
94(100) History of the world
94(4) History of Europe
94(474) History of Baltic States
94(474.3) History of Latvia
94(47+57) History of Soviet Union
94(479) History of Russia

UDC index

Sanitary engineering	628
School organization	371
Science	001
Sculptors	73.071.1
Sculpture	73
Secondary school	373
Sensory perception	159.94
Serial publications	050
Sex psychology	159.92
Social psychology	316.6
Social sciences	3
Social systems	1
Social welfare	36
Social work	36
Sociology	31
Sociology	316
Sociology of culture	316.7
Space research	52
Sport	796/799
Stage directors	782.071.2.027
Stagecraft	792
Stamp collecting	656.8
State and people, relation between	323
Statistics	31
Statistics	311
Structural engineering	624
Surgery	617
Table games	794
Taxes	336
Technology	60
Telecommunication	654
Telephone	621.3
Telephone systems	654
Television	621.3
Television broadcasting	654
Theatre	792
Therapeutics	615
Tourism (economic sector)	338
Tourism	796.5
Town planning	71
Trade	339
Traditions	39
Translating	81'255
Transport services	656
Transport vehicle engineering	629
Travels	910
Vegetable gardening	632/635
Veterinary science	636/637
Video	621.3

Appendix K

DESCRIPTIVE METADATA SCEME

Author
Contributor
Publisher
Date
Subject
Notes
Edition
FormatExtent
FormatMedium
Identifiers
Location organisation
Language
Coverage
Rights

GUIDELINES

1. If no specification is given the elements should be filled the same way as for MARC21 according to AACR2 respecting the elements' separation signs of AACR2 (*AACR2: Anglo-American cataloguing rules, 2ⁿᵈ ed., 2002 revision*).
2. The repeatable elements are used where appropriate. An element should be repeated also in case one metadata element relates to several fields of MARC – in such a situation the content of each field is given as a separate (repeated) element.
3. Personal names, corporate names, geographic names and subjects are added consulting Authority database of the National Library of Latvia for the form of the needed value. If the respective value cannot be found in the database it should be described according to AACR2, and the subject is to be defined following the methods of subject derivation for the digital library.
4. The personal names adjusted according to the Authority file of the National Library of Latvia or AACR2 shall be given in the following elements: Creator, Contributors, Subject, Inscription. In other fields (Title, Abstract, Notes, etc., personal names should follow the sequence: first name, surname (GGGG-GGGG).
5. No abbreviations should be used in metadata fields.

No.	**Field (element)** Qualifier	Explanation	Example
1.	Title MARC21 related field: 245	Usually the title mentioned in the publication is used. If the title is	*Dailes teātra kolektīvs*

| | | followed by an explanatory title it is separated by space-colon-space.

If there is no title in the publication a short descriptive title should be assigned. Square brackets are not used.

The metadata creator assigns a short descriptive title for images, photographs and the like if the title is not understandable or does not reveal what is seen at the image (The title of the original may be used in the field Alternative title) | |
|---|---|---|---|
| | **Alternative title** Repeatable field MARC21 related fields: 130, 210, 240, 242, 246, 730, 740 | Alternative title is the title in any possible form used as substitute or alternative of the official title of the resource. It may include: abbreviation (210), the translation of the title (242), unified title (130), other forms of the title (246) | *Collective of the Daile theatre* |
| 2. | **Type** | The type of the digital object | *Postcard* *Photograph* |
| 3. | **Member of collection** | The collection to which the specific digital object belongs | *Zudusī Latvija* |
| 4. | **Creator** Repeatable field MARC21 related | The person or institution responsible for the | *Skalberga Elfrīda* *Bārda Fricis* |

	fields: 100, 110, 111, 700, 710, 711, 720	creation of the content of the resource The names of creators or institutions shall be justified according to the Authority file of the National Library of Latvia, subdivisions ''Personal Names'', ''Corporate Names'' and ''Meeting Names''. Address: www.lnb.lv, El.katalogs-> Datu bāzes ->Autoritatīvā datu bāze **Personal name** is used in the form: Surname, First name GGGG-GGGG **Corporate name** is used in the form: The name of the institution Or: The name of the institution. The subordinate name of the institution	*1880-1919* *Rīgas vēstures un kuģniecības muzejs* *Latvijas Universitāte. Bibliotēka*
5.	**Contributor** Repeatable field MARC21 related fields: 700, 710, 711, 720, 245$c	Person and/or institution who has contributed to the creation of the resource. It is defined following the same rules as for ''Creator''	*Vītols, Jāzeps, 1863-1948* *Loskiel, Georg Heinrih, 1740-1814* *Bērziņš, Roberts, 1868-1935*
	ContributorRole Repeatable field		*Translator* *The author of the*

		MARC21related field: 700$e		*text*
6.		**Publisher** MARC21 related field: 260ab	A person, institution or service (responsible for access to the resource) may be the publisher. Use the form: Place : Publisher (The place of production : Producer) Give the printing house, producer, printer after the publisher if it differs from the publisher. Enclose it in parenthesis. If no publication place and/or no publisher is given but it is available from another source it is permitted to include these data into description. Use square brackets: [Place : Publisher] Place : [Publisher] If the place of publication is not found anywhere leave this element empty – do not write anything. **The data found on photographs concerning the photo studio is given in general notes (do not**	*Riga : Verlag Lenz & Rudolff (Hamburg : Knackstedt & Nather)*

		include them into Publisher's field). Include them in quotation marks citing as given in the source.	
7.	**Date Created** The date when the original has been created MARC21 related fields: 260cg (printing, printing-house) (date), 533$d (reproduction date)	Give the date when the resource you describe has been created. Do not use the date of the creation of previous versions **The permitted forms of date:** GGGG-MM-DD GGGG-MM GGGG [GGGG or GGGG] – one of these years [Between GGGG and GGGG] [GGGG?] – presumable date [GGG-] –the decade is known exactly [GG--] –the century is known exactly [GG--?] – presumable century	*1925-02-05* *1926-10* *[1925 or 1926]* *[amid 1890 and 1909]* *[192-?]* *[191-]*
	Date Issued **The date of publication** MARC21 related fields: 260$c 008/07-10	The data of the official publishing of the resource Use this specification in case the official date differs from the other date related to this resource **The permitted forms of date – the same as for Date Created**	
	Subject	Subjects shall be	*Children*

	Repeatable field MARC21 related fields: 600, 610, 611, 630, 650, 653	selected from the Authority file of the National Library of Latvia, subdivisions Personal names, Corporate Names, Meeting Names, Topical Names and Uniform Titles Address of the Authority file: www.lnb.lv ,El.Katalogs->->Datu bāzes -> Autoritatīvā datu bāze	*Photographs Portraits* *Drupas, Fridrich, b.1890 Men Latvians-France Photographs Portraits*
8.	**Abstract** Repeatable field MARC21 related field: 520	Resume of **the content** of the resource. This field is searchable. Write simple short sentences. Start with the name of the object, personal names, etc. as they are seen on the digital object, continue with a short description of them. Do not start with the phrase ''On this picture…'', ''You can see on the photograph…'' The abstract may explain the values of the field Subject, e.g., the subject is ''Interior'', the abstract may include the list of its components (table, chair, mirror, lamp	*The vitriolic of the Small Guild is made in Hanover in 1888 after the drawing of A. Freystadt. The Small Guild was erected in 1864-1866, the project of the architect Johann Daniel Felsko (1813-1902). The group portrait*

		TableOfContent MARC21 related field: 505	Separate the parts of content by -- (2 minuses). Separate the parts of content of sound recordings by ; (semicolon).	*Introduction / Howard H.Brinton -- William I.Hull : a biographical sketch / Janet Whitney* *Quatrain II (16:35) --Water Ways (1:57) -- Waves (10:49)*
		Origin Repeatable field MARC21 related field: 500	The date of the creation / acquirement of the resource. Include the information about the previous owner, grantor	*The picture acquired on March 18, 1952 from the person whose surname Fick*
		TextFirstRow MARC21 related field (For NLL locally defined field: 909)	The first row of the text and gamut. Used only for the metadata of printed music	*Kas skaistāks mūža dienās*
		ScientificComment Repeatable field	A comment by a specialist, a professional or a competent working group	*Iļģuciems was situated in the area of the Riga city and was strongly demarcated...*
		Description **General notes** Repeatable field MARC21 related fields: 500-599, except 505, 506, 520, 530, 540, 546	Repeat the fields and give the other significant information. Give the data about photo studious and the persons included into the names of the studious in cases when: 1. the person is the only photographer of the studio; 2. it is possible to identify the	*On the reverse of the photograph ths studio: ''Fotografija F.Knoke Galgauskas ''Božās''* *(For the element ''Creator'': Knoks, Ferdinands, (1870-1953)*

		data about this person. In this case the data should be included also into the element Creator	
9.	**Edition** MARC21 related field: 250$a	Edition or version	*3rd ed.* *Version 2.0*
	DataCaptured The date of digitalisation	The date of digitalisation, harvesting, inclusion into metadata	*2007-12-20*
	FormatExtent Repeatable field MARC21related fields: 300$a, 533$e (the physical characteristics of the reproduction) .	Write each distinct piece of information as the repeated element of Extent. Separate the numeral(s) from the word(s) it denotes by a space irrespective its location (before or after the name).	*1 postcard*
	Extent2		
	Extent3		
	FormatMedium MARC21 related field: 340$a		*Paper* *Audiocassette* *Slide* *Photon* *Photo paper* *Disc*
10.	**IdentifierISBN** The unique identifier ISBN		
	IdentifierISSN The unique identifier ISSN		
	Identifier	Another unique identifier	
11.	**Location Organisation**	This element is used if the resource	*The National Library of Latvia*

		is digitalised from non-digitalised original.	
	Location department The department holding the original		*Department of Rare Books and Handwritings*
	Location Code The code of the disposition of the original	Shelf cipher	*NLL RXA27, 11, p.117*
12.	**Language** Repeatable field	The language of the intellectual content of the resource	
13.	**Coverage Temporal** Coverage of the content Repeatable field MARC21 related fields: 513$b, 033$a, 651$y	Coverage Temporal means the time when the intellectual content of the resource was created, e.g., the resource describes an aspect of the 19th century whereas the resource itself has been created in 2003. The 19th century is the Coverage Temporal, but the date when the resource has been created is 2003. Usually this information comes together with the chronological subdivision of the subject, and this field remains empty. If it is impossible to reveal the period of time by subject include it in this element using the	

		rules of date description mentioned above.	
	Coverage Spatial Repeatable field Marc21 related fields: 522, 651, 255, 650$z, 752	The names of geographic places are selected from the Authority file of the National Library of Latvia, subdivision Geographic Names. Address: www.lnb.lv, El.Katalogs->datu bāzes->Autoritatīvā datu bāze	*Riga (Latvia)*
	Address Repeatable field	This element is used mainly for pictures to show the recent address of the buildings seen at the picture	*Krišjāņa Valdemāra Street 10a, Riga*
	Scale		*Scale: 1:500 000*
	Projection		
	Coordinates		*(A 21˚00—A 28˚00'/Z 58˚00'—Z 56˚00')*
	Datum		*Sea level, equator, zero meridian*
14.	**Rights** MARC21 related fields: 506, 540	Information about the rights concerning the resource and its content	*Free access Access from project organizations only*
	Audience Repeatable field	The audience to which the resource is addressed or for what it might be useful.	*For general public For research*